

PATENT APPLICATION



09524618

JC598 U.S. PTO

09/524618



03/14/00

INITIALS

MAR 29 2004

CONTENTS

	Date Received (Incl. C. of M.) or Date Mailed		Date Received (Incl. C. of M.) or Date Mailed
1. Application <u>15</u> papers.		42.	
2. <u>Small Entity</u>	<u>5-11-00</u>	43.	
3. <u>131</u>	<u>1-22-01</u>	44.	
4. <u>Notice Allowed</u>	<u>12-18-01</u>	45.	
5. <u>Amend A</u>	<u>3-18-02</u>	46.	
6. <u>PTOL-271</u>	<u>5-3-02</u>	47.	
7. <u>Priority Papers</u>	<u>3-15-02</u>	48.	
8. <u>Drawings</u>	<u>3-18-02</u>	49.	
9. <u>Drawings</u>	<u>3-18-02</u>	50.	
10. <u>Formal Drawings (15 sheets) set</u>	<u>3-18-02</u>	51.	
11. <u>PTO 948 & PTO 455</u>	<u>7/8/02</u>	52.	
12. <u>Formal Drawings (15 sheets) set</u>	<u>8-19-02</u>	53.	
13.		54.	
14.		55.	
15.		56.	
16.		57.	
17.		58.	
18.		59.	
19.		60.	
20.		61.	
21.		62.	
22.		63.	
23.		64.	
24.		65.	
25.		66.	
26.		67.	
27.		68.	
28.		69.	
29.		70.	
30.		71.	
31.		72.	
32.		73.	
33.		74.	
34.		75.	
35.		76.	
36.		77.	
37.		78.	
38.		79.	
39.		80.	
40.		81.	
41.		82.	

(LEFT OUTSIDE)

SEARCHED			
Class	Sub.	Date	Exmr.
382	232	12/11/01	107
	244		
	254		
	260		
	263		
	264		
	270		
	274		
	162		
	240		
348	608		
	620		
375	240		
	240.29		

SEARCH NOTES (INCLUDING SEARCH STRATEGY)		
	Date	Exmr.

INTERFERENCE SEARCHED			
Class	Sub.	Date	Exmr.
382	244	12/12/01	PT
	260		
	270		
	263		
	264		
	274		

Page 3 of 297

ISSUE SLIP STAPLE AREA (for additional cross references)

POSITION	INITIALS	ID NO.	DATE
FEE DETERMINATION		67834	2/22
O.I.P.E. CLASSIFIER		109.16	5-15-00
FORMALITY REVIEW	RUBS		
RESPONSE FORMALITY REVIEW			

INDEX OF CLAIMS

✓ Rejected N Non-elected
 = Allowed I Interference
 - (Through numeral) ... Canceled A Appeal
 ÷ Restricted O Objected

Claim	Date
Final Original	
1	12/2/01
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	

Claim	Date
Final Original	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99	
100	

Claim	Date
Final Original	
101	
102	
103	
104	
105	
106	
107	
108	
109	
110	
111	
112	
113	
114	
115	
116	
117	
118	
119	
120	
121	
122	
123	
124	
125	
126	
127	
128	
129	
130	
131	
132	
133	
134	
135	
136	
137	
138	
139	
140	
141	
142	
143	
144	
145	
146	
147	
148	
149	
150	

If more than 150 claims or 10 actions
staple additional sheet here

(LEFT INSIDE)

(12) **United States Patent**
Seraizin et al.

(10) Patent No.: **US 6,473,532 B1**
(45) Date of Patent: **Oct. 29, 2002**

(54) **METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE SYNTACTIC ENCODING**

(75) Inventors: **Semion Seraizin; Vitaly Seraizin**, both of Mazkeret Batya (IL)

(73) Assignee: **VLS COM Ltd.,** Rechovot (IL)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/524,618**

(22) Filed: **Mar. 14, 2000**

(30) **Foreign Application Priority Data**

Jan. 23, 2000 (IL) 134182

(51) Int. CL.⁷ **G06K 9/36; G06K 9/46**

(52) U.S. Cl. **382/244; 382/260; 382/263; 382/264; 382/270; 382/274**

(58) Field of Search **382/232, 244, 382/254, 260, 263, 264, 270, 274, 162, 240; 348/608, 620; 375/240, 240.29**

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,341,442 A 8/1994 Barrett
5,491,519 A 2/1996 Kim
5,586,200 A 12/1996 Devaney et al.
5,774,593 A 6/1998 Zick et al.
5,787,203 A * 7/1998 Lee et al. 382/232
5,796,864 A 8/1998 Callahan
5,845,012 A 12/1998 Jung
5,847,766 A 12/1998 Peak
5,870,501 A 2/1999 Kim
5,901,178 A * 5/1999 Lee et al. 375/240

OTHER PUBLICATIONS

Raj Malluri, et al., "A Robust, Scalable, Object-Based Video Compression Technique for Very Low Bit-Rate Coding", IEEE Transaction of Circuit and Systems for Video Technology, vol. 7, No. 1, Feb. 1997.

AwadKh. Al-Asmari, "An Adaptive Hybrid Coding Scheme for HDTV and Digital Sequences," IEEE Transactions on Consumer Electronics, vol. 42, No. 3, pp. 926-936, Aug. 1995.

Kwok-tung Lo & Jian Feng, "Predictive Mean Search Algorithms for Fast VQ Encoding of Images," IEEE Transactions On Consumer Electronics, vol. 41, No. 2, pp. 327-331, May 1995.

James Goel et al., "Pre-processing for MPEG Compression Using Adaptive Spatial Filtering", IEEE Transactions On Consumer electronics, vol. 41, No. 3, pp. 687-698, Aug. 1995.

Jian Feng, et al., "Motion Adaptive Classified Vector Quantization for ATM Video Coding", IEEE Transactions on Consumer Electronics, vol. 41, No. 2, pp. 322-326, May 1995.

Austin Y. Lian, et al., "Scene-Context Dependent Reference—Frame Placement for MPEG Video Coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 9, No. 3, pp. 478-489, Apr. 1999.

(List continued on next page.)

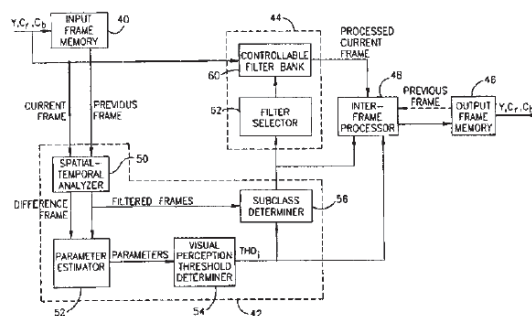
Primary Examiner—Phuoc Tran

(74) Attorney, Agent, or Firm—Eitan, Pearl, Latzer & Cohen-Zedek

(57) **ABSTRACT**

A visual lossless encoder for processing a video frame prior to compression by a video encoder includes a threshold unit, a filter unit, an association unit and an altering unit. The threshold unit identifies a plurality of visual perception threshold levels to be associated with the pixels of the video frame, wherein the threshold levels define contrast levels above which a human eye can distinguish a pixel from among its neighboring pixels of the frame. The filter unit divides the video frame into portions having different detail dimensions. The association unit utilizes the threshold levels and the detail dimensions to associate the pixels of the video frame into subclasses. Each subclass includes pixels related to the same detail and which generally cannot be distinguished from each other. The altering unit alters the intensity of each pixel of the video frame according to its subclass.

17 Claims, 15 Drawing Sheets



OTHER PUBLICATIONS

Kuo-Chin Fan, Kou-Sou Kati, "An Active Scene Analysis-Based approach for Pseudoconstant Bit-Rate Video Coding", IEEE Transactions on Circuits and Systems for Video Technology, vol. 8, No. 2, pp. 159-170, Apr. 1998.

Takashi Ida and Yoko Sambansugi, "Image Segmentation and Contour Detection Using Fractal Coding", IEEE Transactions on Circuits and Systems for Video Technology, vol. 8, No. 8, pp. 968-975, Dec. 1998.

Liang Shen & RangarajM. Rangayyan, "A Segmentation-Based Lossless Image Coding Method for

High-Resolution Medical Image Compression", IEEE Transactions on Medical Imaging, vol. 16, No. 3, pp. 301-316, Jun. 1997.

Adrian Munteanu et al., "Wavelet-Based Lossless Compression of Coronary Angiographic Images", IEEE Transactions on Medical Imaging, vol. 18, No. 3, pp. 272-281, Mar. 1999.

Akira Okumura, et al., "Signal Analysis and Compression performance Evaluation of Pathological Microscopic Images", IEEE Transactions on Medical Imaging, vol. 16, No. 6, pp. 701-710, Dec. 1997.

* cited by examiner



Fig. 1

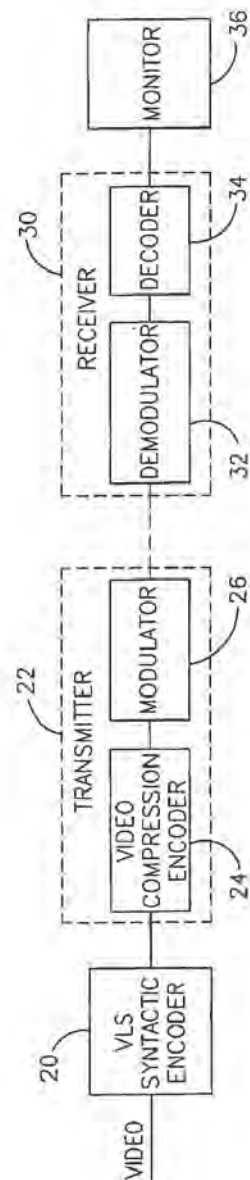


FIG. 2

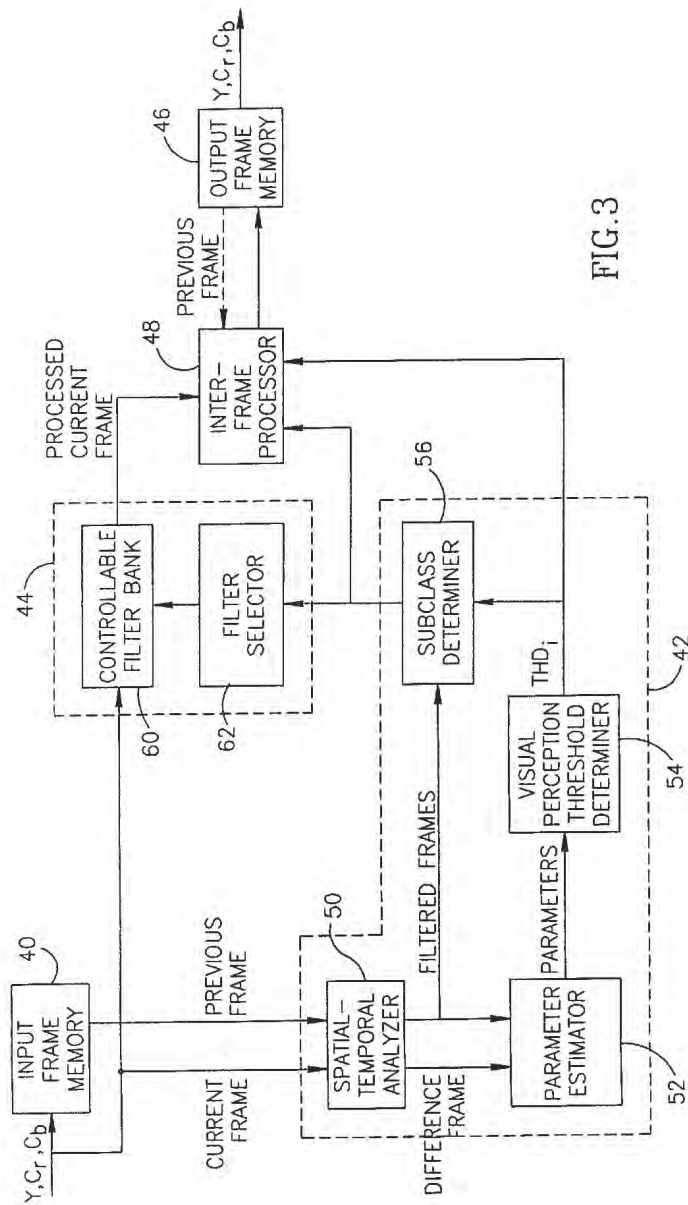


FIG. 3

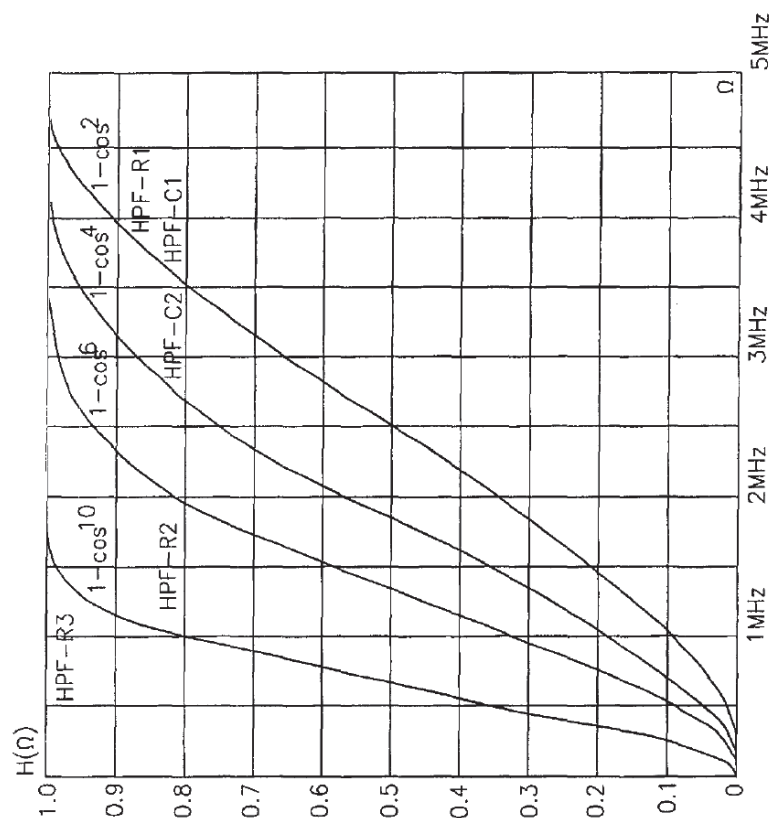


FIG. 4

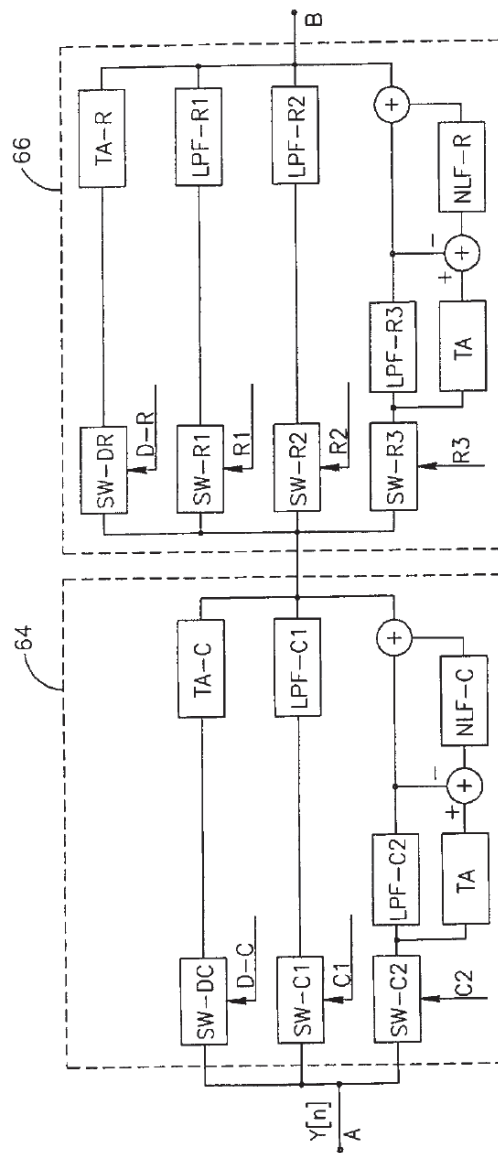


FIG. 5A

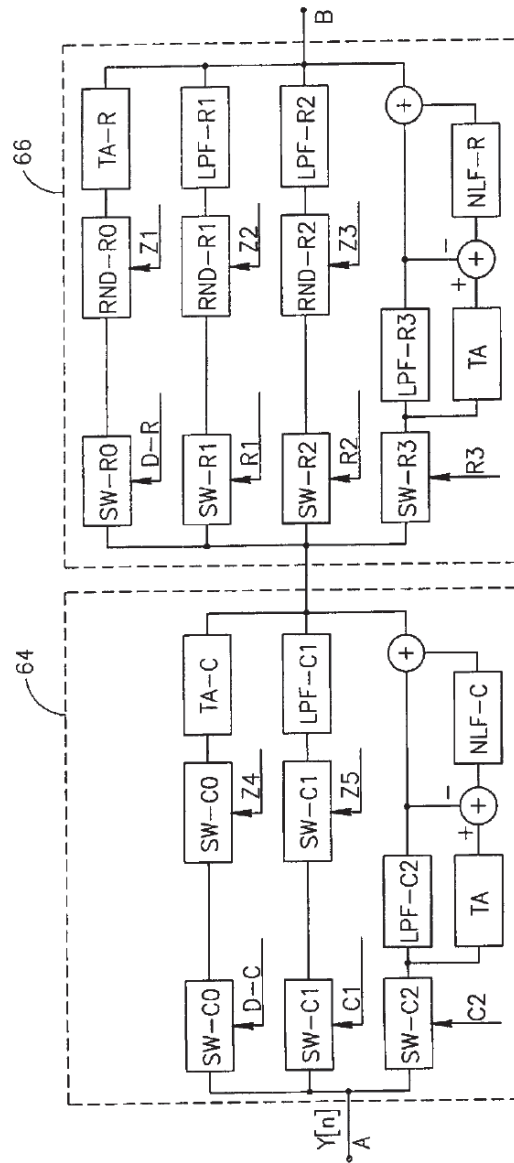


FIG. 5B

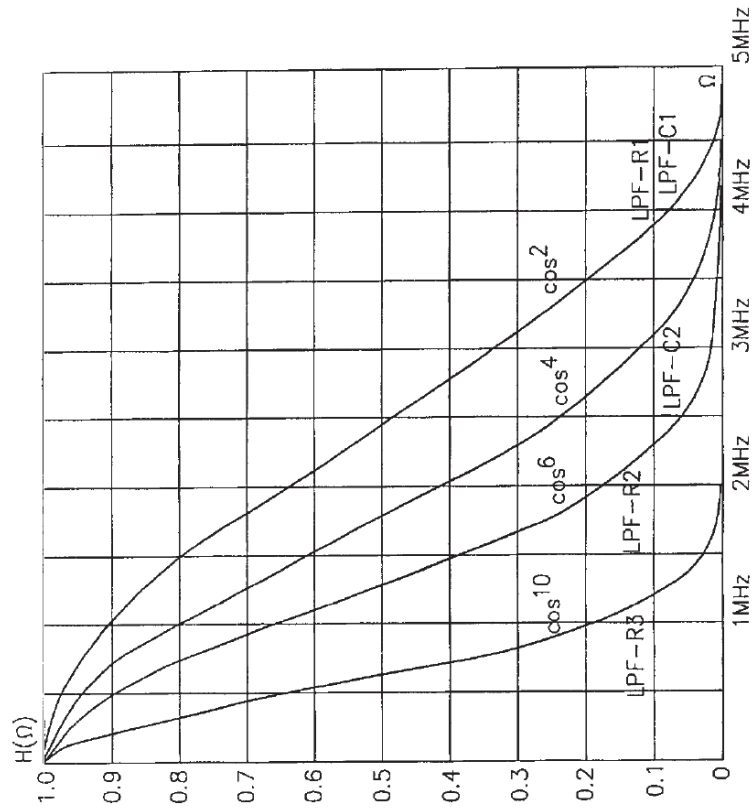


FIG. 6

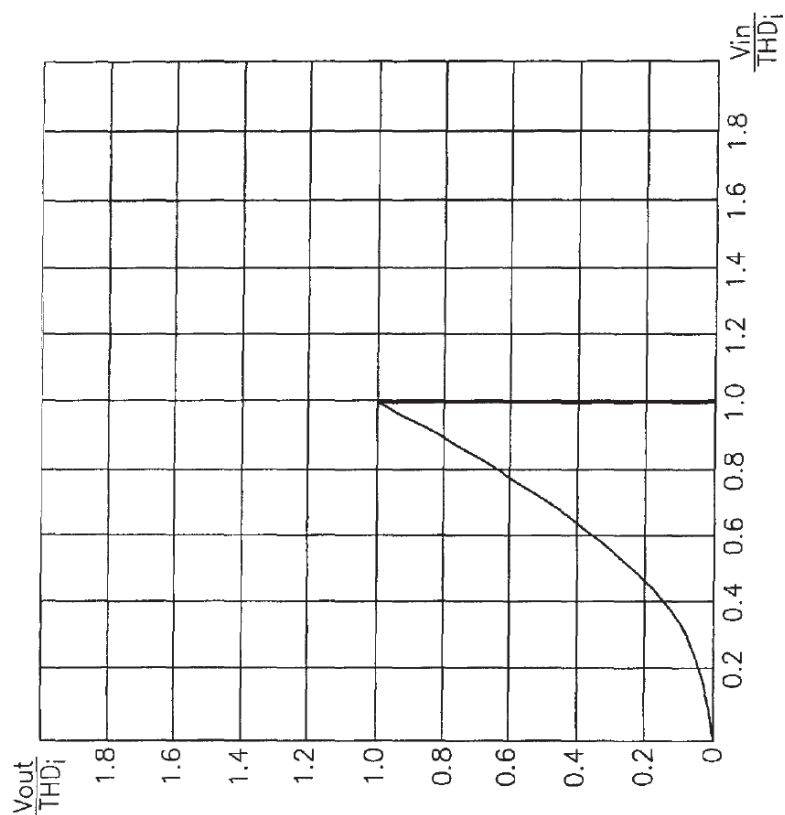


FIG. 7

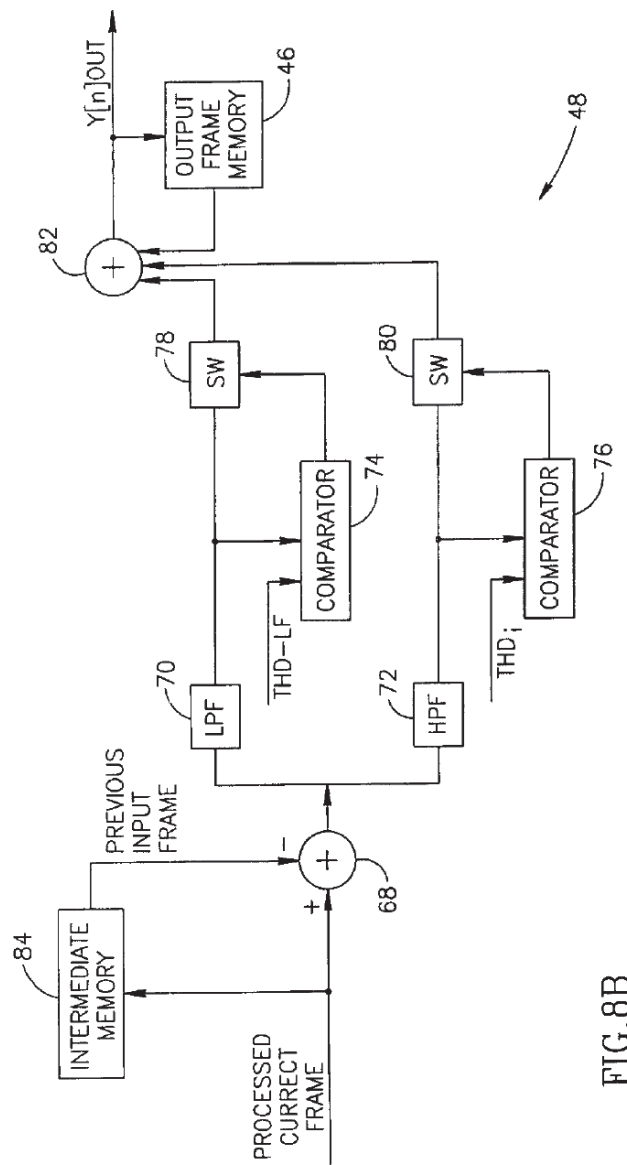


FIG. 8B

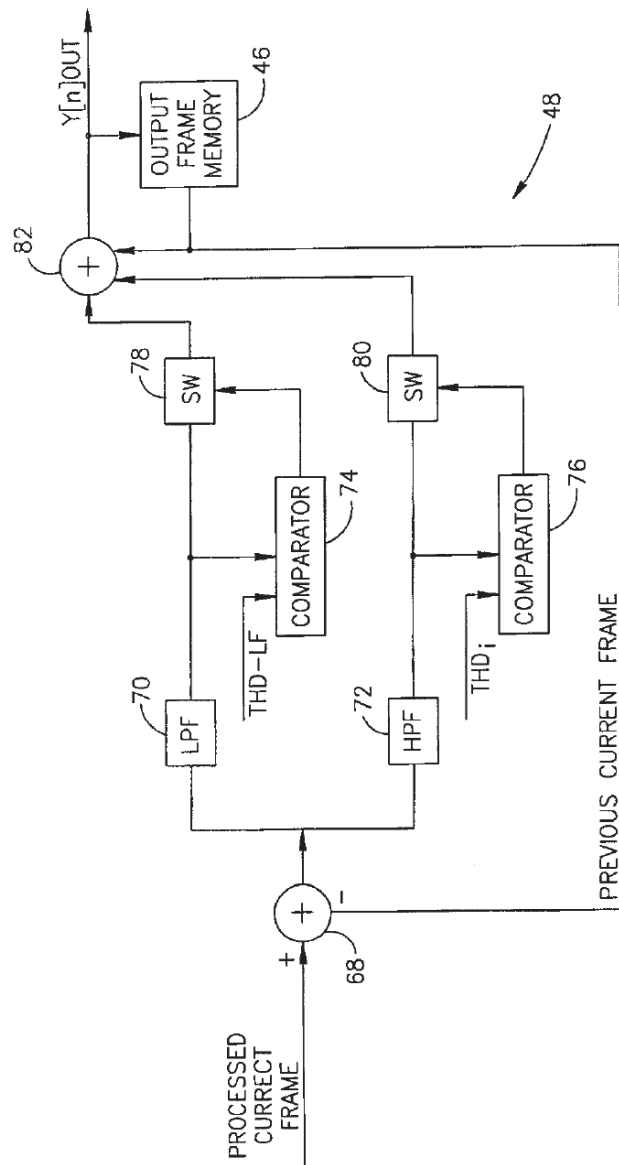


FIG. 8C

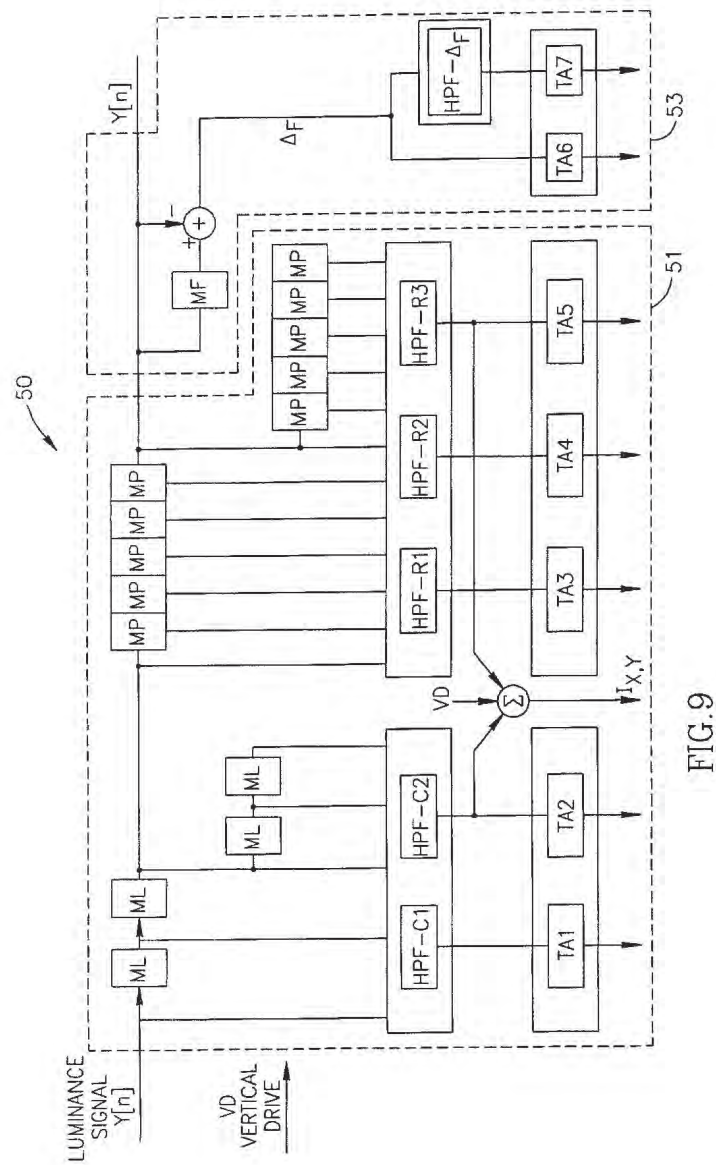


FIG. 9

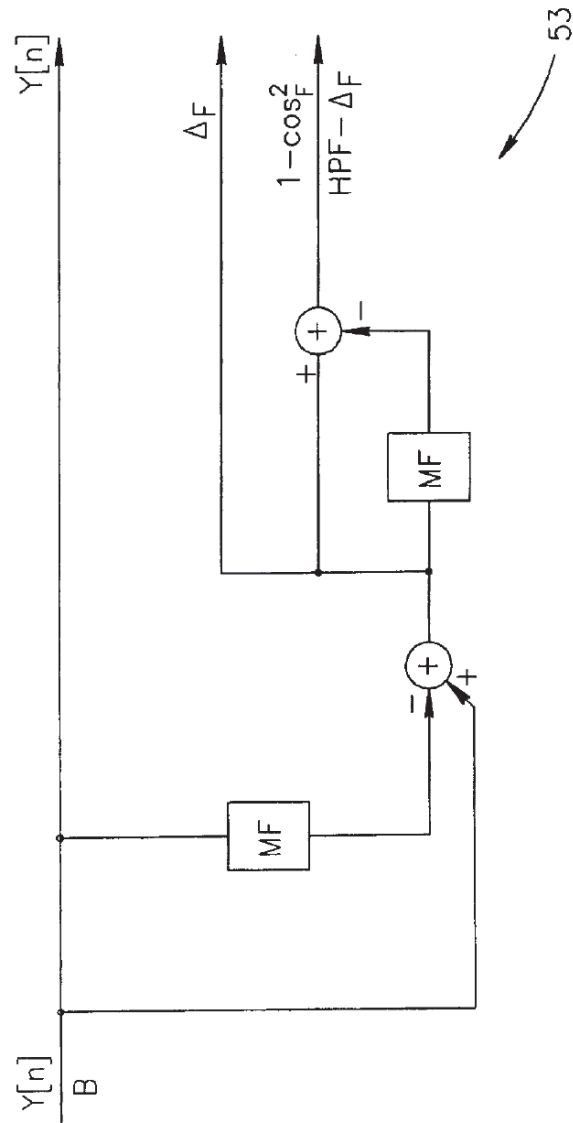
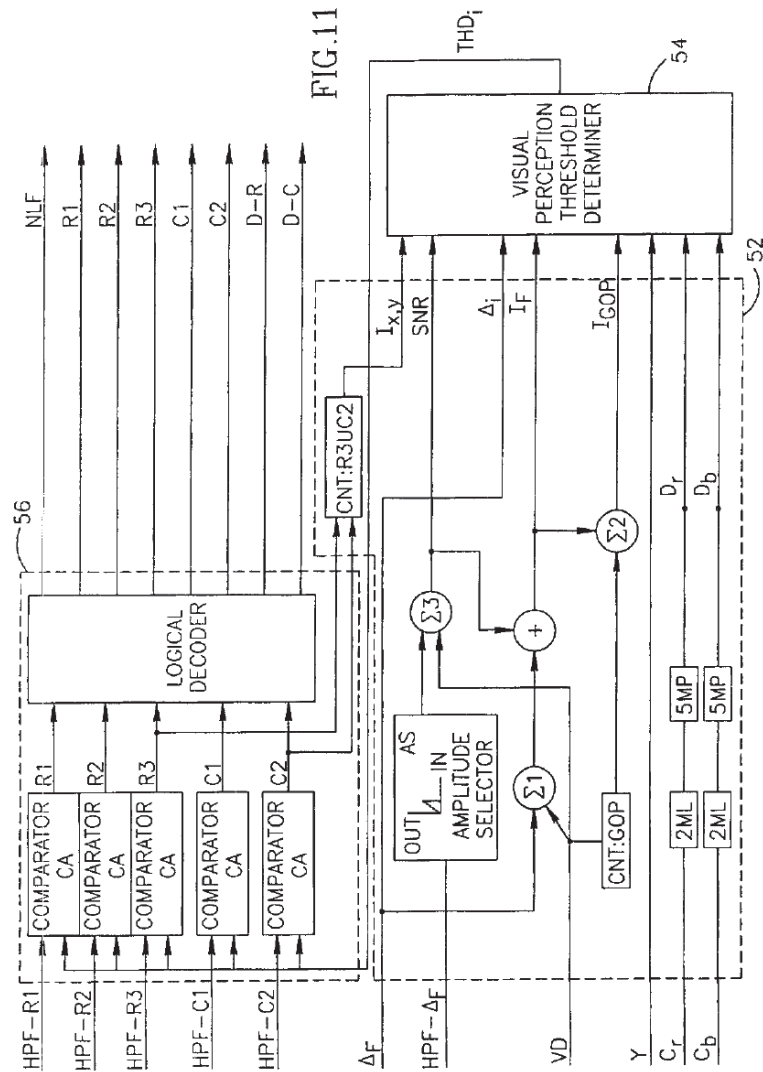


FIG.10B



METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE SYNTACTIC ENCODING

FIELD OF THE INVENTION

The present invention relates generally to processing of video images and, in particular, to syntactic encoding of images for later compression by standard compression techniques.

BACKGROUND OF THE INVENTION

There are many types of video signals, such as digital broadcast television (TV), video conferencing, interactive TV, etc. All of these signals, in their digital form, are divided into frames, each of which consists of many pixels (image elements), each of which requires 8-24 bits to describe them. The result is megabits of data per frame.

Before storing and/or transmitting these signals, they typically are compressed, using one of many standard video compression techniques, such as JPEG, MPEG, H-compression, etc. These compression standards use video signal transforms and intra- and inter-frame coding which exploit spatial and temporal correlations among pixels of a frame and across frames.

However, these compression techniques create a number of well-known, undesirable and unacceptable artifacts, such as blockiness, low resolution and wiggles, among others. These are particularly problematic for broadcast TV (satellite TV, cable TV, etc.) or for systems with very low bit rates (video conferencing, videophone).

Much research has been performed to try and improve the standard compression techniques. The following patents and articles discuss various prior art methods to do so:

U.S. Pat. Nos. 5,870,501, 5,847,766, 5,845,012, 5,796,884, 5,774,593, 5,586,200, 5,491,519, 5,341,442;

Raj Malluri et al., "A Robust, Scalable, Object-Based Video Compression Technique for Very Low Bit-Rate Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, No. 1, February 1997;

AwadKh. Al-Asmari, "An Adaptive Hybrid Coding Scheme for HDTV and Digital Sequences," *IEEE Transactions on Consumer Electronics*, vol. 42, No. 3, pp. 926-936, August 1995;

Kwok-tung Lo and Jian Feng, "Predictive Mean Search Algorithms for Fast VQ Encoding of Images," *IEEE Transactions On Consumer Electronics*, vol. 41, No. 2, pp. 327-331, May 1995;

James Goel et al., "Pre-processing for MPEG Compression Using Adaptive Spatial Filtering," *IEEE Transactions On Consumer Electronics*, "vol. 41, No. 3, pp. 687-698, August 1995;

Jian Feng et al., "Motion Adaptive Classified Vector Quantization for ATM Video Coding," *IEEE Transactions on Consumer Electronics*, vol. 41, No. 2, p. 322-326, May 1995;

Austin Y. Lan et al., "Scene-Context Dependent Reference-Frame Placement for MPEG Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, No. 3, pp. 478-489, April 1999;

Kuo-Chin Fan, Kou-Sou Kan, "An Active Scene Analysis-Based approach for Pseudocostant Bit-Rate Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8 No. 2, pp. 159-170, April 1998;

Takashi Ida and Yoko Sambansugi, "Image Segmentation and Contour Detection Using Fractal Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, No. 8, pp. 968-975, December 1998;

Liang Shen and Rangaraj M. Rangayyan, "A Segmentation-Based Lossless Image Coding Method for High-Resolution Medical Image Compression," *IEEE Transactions on Medical Imaging*, vol. 16, No. 3, pp. 301-316, June 1997;

Adrian Munteanu et al., "Wavelet-Based Lossless Compression of Coronary Angiographic Images," *IEEE Transactions on Medical Imaging*, vol. 18, No. 3, p. 272-281, March 1999; and

Akira Okumura et al., "Signal Analysis and Compression Performance Evaluation of Pathological Microscopic Images," *IEEE Transactions on Medical Imaging*, vol. 16, No. 6, pp. 701-710, December 1997.

SUMMARY OF THE INVENTION

An object of the present invention is to provide a method and apparatus for video compression which is generally lossless vis-a-vis what the human eye perceives.

There is therefore provided, in accordance with a preferred embodiment of the present invention, a visual lossless encoder for processing a video frame prior to compression by a video encoder. The encoder includes a threshold determination unit, a filter unit, an association unit and an altering unit. The threshold determination unit identifies a plurality of visual perception threshold levels to be associated with the pixels of the video frame, wherein the threshold levels define contrast levels above which a human eye can distinguish a pixel from among its neighboring pixels of the frame. The filter unit divides the video frame into portions having different detail dimensions. The association unit utilizes the threshold levels and the detail dimensions to associate the pixels of the video frame into subclasses. Each subclass includes pixels related to the same detail and which generally cannot be distinguished from each other. The altering unit alters the intensity of each pixel of the video frame according to its subclass.

Moreover, in accordance with a preferred embodiment of the present invention, the altering unit includes an inter-frame processor and an intra-frame processor.

Furthermore, in accordance with a preferred embodiment of the present invention, the intra-frame processor includes a controllable filter bank having a plurality of different filters and a filter selector which selects one of the filters for each pixel according to its subclass.

Further, in accordance with a preferred embodiment of the present invention, the inter-frame processor includes a low pass filter and a high pass filter operative on a difference frame between a current frame and a previous frame, large and small detail threshold elements for thresholding the filtered difference frame with a large detail threshold level and a small detail threshold level, respectively, and a summer which sums the outputs of the two filters as amended by the threshold elements.

Still further, in accordance with a preferred embodiment of the present invention, the threshold unit includes a unit for generating a plurality of parameters describing at least one of the following parameters: the volume of information in the frame, the per pixel color and the cross-frame change of intensity, and unit for generating the visual perception threshold from at least one of the parameters.

3

There is also provided, in accordance with a preferred embodiment of the present invention, a method of visual lossless encoding of frames of a video signal. The method includes steps of spatially and temporally separating and analyzing details of the frames, estimating parameters of the details, defining a visual perception threshold for each of the details in accordance with the estimated detail parameters, classifying the frame picture details into subclasses in accordance with the visual perception thresholds and the detail parameters and transforming each the frame detail in accordance with its associated subclass.

Additionally, in accordance with a preferred embodiment of the present invention, the step of separating and analyzing also includes the step of spatial high pass filtering of small dimension details and temporal filtering for detail motion analysis.

Moreover, in accordance with a preferred embodiment of the present invention, the step of estimating comprises at least one of the following steps:

determining ND_n , a per-pixel signal intensity change between a current frame and a previous frame, normalized by a maximum intensity;

determining a NI_{eq} , a normalized volume of intraframe change by high frequency filtering of the frame, summing the intensities of the filtered frame and normalizing the sum by the maximum possible amount of information within a frame;

generating NI_p , a volume of inter-frame changes between a current frame and its previous frame normalized by a maximum possible amount of information volume within a frame;

generating NI_{GOF} , a normalized volume of inter-frame changes for a group of pictures from the output of the previous step of generating;

evaluating a signal-to-noise ratio SNR by high pass filtering a difference frame between the current frame and the previous frame by selecting those intensities of the difference frame lower than a threshold defined as three times a noise level under which noise intensities are not perceptible to the human eye, summing the intensities of the pixels in the filtered frame and normalizing the sum by the maximum intensity and the total number of pixels in the frame;

generating NY_p , a normalized intensity value per-pixel;

generating a per-pixel color saturation level p_s ;

generating a per-pixel hue value h_i ; and

determining a per-pixel response $R(h_i)$ to the hue value.

Further, in accordance with a preferred embodiment of the present invention, the step of defining includes the step of producing the visual perception thresholds, per-pixel, from a minimum threshold value and at least one of the parameters.

Still further, in accordance with a preferred embodiment of the present invention, the step of defining includes the step of producing the visual perception thresholds, per-pixel, according to the following equation,

$$THD_i = THD_{min}(1 + ND_i + NI_{eq} +$$

$$NI_p + NI_{GOF} + NY_p + p_s + (1 - R(h_i)) + \frac{200}{SNR})$$

wherein THD_{min} is a minimum threshold level.

Moreover, in accordance with a preferred embodiment of the present invention, the step of classifying includes the

4

steps of comparing multiple spatial high frequency levels of a pixel against its associated visual perception threshold and processing the comparison results to associate the pixel with one of the subclasses.

Further, in accordance with a preferred embodiment of the present invention, the step of transforming includes the step of filtering each subclass with an associated two-dimensional low pass filter.

Still further, in accordance with a preferred embodiment of the present invention, the step of transforming includes the steps of generating a difference frame between the current frame and a previous transformed frame, low and high pass filtering of the difference frame, comparing the filtered frames with a large detail threshold and a small detail threshold and summing those portions of the filtered frames which are greater than the thresholds.

Additionally, in accordance with a preferred embodiment of the present invention, the large detail threshold is 2 to 5 percent.

Moreover, in accordance with a preferred embodiment of the present invention, the method includes the step of rounding the output of the step of transforming.

Finally, the intensity can be a luminance value or a chrominance value.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be understood and appreciated more fully from the following detailed description taken in conjunction with the appended drawings in which:

FIG. 1 is an example of a video frame;

FIG. 2 is a block diagram illustration of a video compression system having a visual lossless syntactic encoder, constructed and operative in accordance with a preferred embodiment of the present invention;

FIG. 3 is a block diagram illustration of the details of the visual lossless syntactic encoder of FIG. 2;

FIG. 4 is a graphical illustration of the transfer functions for a number of high pass filters useful in the syntactic encoder of FIG. 3;

FIGS. 5A and 5B are block diagram illustrations of alternative embodiments of a controllable filter bank forming part of the syntactic encoder of FIG. 3;

FIG. 6 is a graphical illustration of the transfer functions for a number of low pass filters useful in the controllable filter bank of FIGS. 5A and 5B;

FIG. 7 is a graphical illustration of the transfer function for a non-linear filter useful in the controllable filter bank of FIGS. 5A and 5B;

FIGS. 8A, 8B and 8C are block diagram illustrations of alternative embodiments of an inter-frame processor forming a controlled filter portion of the syntactic encoder of FIG. 3;

FIG. 9 is a block diagram illustration of a spatial-temporal analyzer forming part of the syntactic encoder of FIG. 3;

FIGS. 10A and 10B are detail illustrations of the analyzer of FIG. 9; and

FIG. 11 is a detail illustration of a frame analyzer forming part of the syntactic encoder of FIG. 3.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

Applicants have realized that there are different levels of image detail in an image and that the human eye perceives these details in different ways. In particular, Applicants have realized the following:

5

1. Picture details whose detection mainly depends on the level of noise in the image occupy approximately 50-80% of an image.
2. A visual perception detection threshold for image details does not depend on the shape of the details in the image.
3. A visual perception threshold THD depends on a number of picture parameters, including the general brightness of the image. It does not depend on the noise spectrum.

The present invention is a method for describing, and then encoding, images based on which details in the image can be distinguished by the human eye and which ones, can only be detected by it.

Reference is now made to FIG. 1, which is a grey-scale image of a plurality of shapes of a bird in flight, ranging from a photograph of one (labeled 10) to a very stylized version of one (labeled 12). The background of the image is very dark at the top of the image and very light at the bottom of the image.

The human eye can distinguish most of the birds of the image. However, there is at least one bird, labeled 14, which the eye can detect but cannot determine all of its relative contrast details. Furthermore, there are large swaths of the image (in the background) which have no details in them.

The present invention is a method and system for syntactic encoding of video frames before they are sent to a standard video compression unit. The present invention separates the details of a frame into two different types, those that can only be detected (for which only one bit will suffice to describe each of their pixels) and those which can be distinguished (for which at least three bits are needed to describe the intensity of each of their pixels).

Reference is now made to FIG. 2, which illustrates the present invention within an image transmission system. Thus, FIG. 2 shows a visual lossless (VLS) syntactic encoder 20 connected to a standard video transmitter 22 which includes a video compression encoder 24, such as a standard MPEG encoder, and a modulator 26. VLS syntactic encoder 20 transforms an incoming video signal such that video compression encoder 24 can compress the video signal two to five times more than video compression encoder 24 can do on its own, resulting in a significantly reduced volume bit stream to be transmitted.

Modulator 26 modulates the reduced volume bit stream and transmits it to a receiver 30, which, as in the prior art, includes a demodulator 32 and a decoder 34. Demodulator 32 demodulates the transmitted signal and decoder 34 decodes and decompresses the demodulated signal. The result is provided to a monitor 36 for display.

It will be appreciated that, although the compression ratios are high in the present invention, the resultant video displayed on monitor 36 is not visually degraded. This is because encoder 20 attempts to quantify each frame of the video signal according to which sections of the frame are more or less distinguished by the human eye. For the less-distinguished sections, encoder either provides pixels of a minimum bit volume, thus reducing the overall bit volume of the frame or smoothest the data of the sections such that video compression encoder 24 will later significantly compress these sections, thus resulting in a smaller bit volume in the compressed frame. Since the human eye does not distinguish these sections, the reproduced frame is not perceived significantly differently than the original frame, despite its smaller bit volume.

Reference is now made to FIG. 3, which details the elements of VLS encoder 20. Encoder 20 comprises an input

6

frame memory 40, a frame analyzer 42, an intra-frame processor 44, an output frame memory 46 and an inter-frame processor 48. Analyzer 42 analyzes each frame to separate it into subclasses, where subclasses define areas whose pixels cannot be distinguished from each other. Intra-frame processor 44 spatially filters each pixel of the frame according to its subclass and, optionally, also provides each pixel of the frame with the appropriate number of bits. Inter-frame processor 48 provides temporal filtering (i.e. inter-frame filtering) and updates output frame memory 46 with the elements of the current frame which are different than those of the previous frame.

It is noted that frames are composed of pixels, each having luminance Y and two chrominance C₁ and C₂ components, each of which is typically defined by eight bits. VLS encoder 20 generally separately processes the three components. However, the bandwidth of the chrominance signals is half as wide as that of the luminance signal. Thus, the filters (in the x direction of the frame) for chrominance have a narrower bandwidth. The following discussion shows the filters for the luminance signal Y.

Frame analyzer 42 comprises a spatial-temporal analyzer 50, a parameter estimator 52, a visual perception threshold determiner 54 and a subclass determiner 56. Details of these elements are provided in FIGS. 9-11, discussed hereinbelow.

As discussed hereinabove, details which the human eye distinguishes are ones of high contrast and ones whose details have small dimensions. Areas of high contrast are areas with a lot of high frequency content. Thus, spatial-temporal analyzer 50 generates a plurality of filtered frames from the current frame, each filtered through a different high pass filter (HPF), where each high pass filter retains a different range of frequencies therein.

FIG. 4, to which reference is now briefly made, is an amplitude vs. frequency graph illustrating the transfer functions of an exemplary set of high pass filters for frames in a non-interlaced scan format. Four graphs are shown. It can be seen that the curve labeled HPF-R3 has a cutoff frequency of 1 MHz and thus, retains portions of the frame with information above 1 MHz. Similarly, curve HPF-R2 has a cutoff frequency of 2 MHz, HPF-C2 has a cutoff frequency of 3 MHz and HPF-R1 and HPF-C1 have a cutoff frequency of 4 MHz. As will be discussed hereinbelow, the terminology "Rx" refers to operations on a row of pixels while the terminology "Cx" refers to operations on a column of pixels.

In particular, the filters of FIG. 4 implement the following finite impulse response (FIR) filters on either a row of pixels (the x direction of the frame) or a column of pixels (the y direction of the frame), where the number of pixels used in the filter defines the power of the cosine. For example, a filter implementing $\cos^{10}x$ takes 10 pixels around the pixel of interest, five to one side and five to the other side of the pixel of interest.

$$\text{HPF-R3: } 1 - \cos^{10}x$$

$$\text{HPF-R2: } 1 - \cos^6x$$

$$\text{HPF-R1: } 1 - \cos^2x$$

$$\text{HPF-C2: } 1 - \cos^6y$$

$$\text{HPF-C1: } 1 - \cos^2y$$

The high pass filters can also be considered as digital equivalents of optical apertures. The higher the cutoff frequency, the smaller the aperture. Thus, filters HPF-R1 and HPF-C1 retain only very small details in the frame (of 1-4

7

pixels in size) while filter HPF-R3 retains much larger details (of up to 11 pixels).

In the following, the filtered frames will be labeled by the type of filter (HPF-X) used to create them.

Returning to FIG. 3, analyzer 50 also generates difference frames between the current frame and another, earlier frame. The previous frame is typically at most 15 frames earlier. A "group" of pictures or frames (GOP) is a series of frames for which difference frames are generated.

Parameter estimator 52 takes the current frame and the filtered and difference frames and generates a set of parameters that describe the information content of the current frame. The parameters are determined on a pixel-by-pixel basis or on a per frame basis, as relevant. It is noted that the parameters do not have to be calculated to great accuracy as they are used in combination to determine a per pixel, visual perception threshold THD.

At least some of the following parameters are determined:

Signal to noise ratio (SNR): this parameter can be determined by generating a difference frame between the current frame and the frame before it, high pass filtering of the difference frame, summing the intensities of the pixels in the filtered frame, normalized by both the number of pixels N in a frame and the maximum intensity I_{MAX} possible for the pixel. If the frame is a television frame the maximum intensity is 255 quanta (8 bits). The high frequency filter selects only those intensities lower than 3σ , where σ indicates a level less than which the human eye cannot perceive noise. For example, σ can be 46 dB, equivalent to a reduction in signal strength of a factor of 200.

Normalized ΔI_i : this measures the change ΔI_i per pixel i , from the current frame to its previous frame. This value is then normalized by the maximum intensity I_{MAX} possible for the pixel.

Normalized volume of intraframe change NI_{IT} : this measures the volume of change in a frame I_{IT} (or how much detail there is in a frame), normalized by the maximum possible amount of information MAX_{INFO} within a frame (i.e. 8 bits per pixel \times N pixels per frame). Since the highest frequency range indicates the amount of change in a frame, the volume of change I_{IT} is a sum of the intensities in the filtered frame having the highest frequency range, such as filtered frame HPF-R1.

Normalized volume of interframe changes NI_F : this measures the volume of changes I_F between the current frame and its previous frame, normalized by the maximum possible amount of information MAX_{INFO} within a frame. The volume of interframe changes I_F is the sum of the intensities in the difference frame.

Normalized volume of change within a group of frames NI_{GOP} : this measures the volume of changes I_{GOP} over a group of frames, where the group is from 2 to 15 frames, as selected by the user. It is normalized by the maximum possible amount of information MAX_{INFO} within a frame and by the number of frames in the group.

Normalized luminance level NY_i : Y_i is the luminance level of a pixel in the current frame. It is normalized by the maximum intensity I_{MAX} possible for the pixel.

Color saturation p_i : this is the color saturation level of the i th pixel and it is determined by:

$$\left[0.78 \left(\frac{C_{ri} - 128}{160} \right)^2 + 0.24 \left(\frac{C_{bi} - 128}{126} \right)^2 \right]^{\frac{1}{2}}$$

where C_{ri} and C_{bi} are the chrominance levels of the i th pixel.

8

Hue h_i : this is the general hue of the i th pixel and is determined by:

$$\arctan \left(1.4 \frac{C_{ri} - 128}{C_{bi} - 128} \right)$$

Alternatively, hue h_i can be determined by interpolating Table 1, below.

Response to hue $R(h_i)$: this is the human vision response to a given hue and is given by Table 1, below. Interpolation is typically used to produce a specific value of the response $R(h)$ for a specific value of hue h .

TABLE 1

Color	Y	C_r	C_b	h (nm)	R (h)
White	235	128	128	—	—
Yellow	210	16	146	575	0.92
Cyan	170	166	16	490	0.21
Green	145	54	34	510	0.59
Magenta	106	202	222	—	0.2
Red	81	90	240	630	0.3
Blue	41	240	110	475	0.11
Black	16	128	128	—	—

Visual perception threshold determiner 54 determines the visual perception threshold THD, per pixel as follows:

$$THD_i = THD_{min}(1 + NI_{IT} + NI_F + NI_{GOP} + p_i + (1 - R_i(h_i)) + \frac{200}{SNR})$$

Subclass determiner 56 compares each pixel i of each high pass filtered frame HPF-X to its associated threshold THD _{i} to determine whether or not that pixel is significantly present in each filtered frame, where "significantly present" is defined by the threshold level and by the "detail dimension" (i.e. the size of the object or detail in the image of which the pixel forms a part). Subclass determiner 56 then defines the subclass to which the pixel belongs.

For the example provided above, if the pixel is not present in any of the filtered frames, the pixel must belong to an object of large size or the detail is only detected but not distinguished. If the pixel is only found in the filtered frame of HPF-C2 or in both frames HPF-C1 and HPF-C2, it must be a horizontal edge (an edge in the Y direction of the frame). If it is found in filtered frames HPF-R3 and HPF-C2, it is a single small detail. If the pixel is found only in filtered frames HPF-R1, HPF-R2 and HPF-R3, it is a very small vertical edge. If, in addition, it is also found in filtered frame HPF-C2, then the pixel is a very small, single detail.

The above logic is summarized and expanded in Table 2.

TABLE 2

High Pass Filters						
Subclass	R1	R2	R3	C1	C2	Remarks
1	0	0	0	0	0	Large detail or detected detail only
2	0	0	0	0	1	Horizontal edge
3	0	0	0	1	1	Horizontal edge
4	0	0	1	0	0	Vertical edge
5	0	0	1	0	1	Single small detail
6	0	0	1	1	1	Single small detail
7	0	1	1	0	0	Vertical edge
8	0	1	1	0	1	Single small detail

TABLE 2-continued

High Pass Filters					
Subclass	R1	R2	R3	C1	C2
9	0	1	1	1	1
10	1	1	1	0	0
11	1	1	1	0	1
12	1	1	1	1	1

The output of subclass determiner 56 is an indication of the subclass to which each pixel of the current frame belongs. Intra-frame processor 44 performs spatial filtering of the frame, where the type of filter utilized varies in accordance with the subclass to which the pixel belongs.

In accordance with a preferred embodiment of the present invention, intra-frame processor 44 filters each subclass of the frame differently and according to the information content of the subclass. The filtering limits the bandwidth of each subclass which is equivalent to sampling the data at different frequencies. Subclasses with a lot of content are sampled at a high frequency while subclasses with little content, such as a plain background area, are sampled at a low frequency.

Another way to consider the operation of the filters is that they smooth the data of the subclass, removing "noisiness" in the picture that the human eye does not perceive. Thus, intra-frame processor 44 changes the intensity of the pixel by an amount less than the visual distinguishing threshold for that pixel. Pixels whose contrast is lower than the threshold (i.e. details which were detected only) are transformed with non-linear filters. If desired, the data size of the detected only pixels can be reduced from 8 bits to 1 or 2 bits, depending on the visual threshold level and the detail dimension for the pixel. For the other pixels (i.e. the distinguished ones), 3 or 4 bits is sufficient.

Intra-frame processor 44 comprises a controllable filter bank 60 and a filter selector 62. Controllable filter bank 60 comprises a set of low pass and non-linear filters, shown in FIGS. 5A and 5B to which reference is now made, which filter selector 62 activates, based on the subclass to which the pixel belongs. Selector 62 can activate more than one filter, as necessary.

FIGS. 5A and 5B are two, alternative embodiments of controllable filter bank 60. Both comprise two sections 64 and 66 which operate on columns (i.e. line to line) and on rows (i.e. within a line), respectively. In each section 64 and 66, there is a choice of filters, each controlled by an appropriate switch, labeled SW-X, where X is one of C1, C2, R1, R2, R3 (selecting one of the low pass filters (LPF)), D-C, D-R (selecting to pass the relevant pixel directly). Filter selector 62 switches the relevant switch, thereby activating the relevant filter. It is noted that the non-linear filters NLF-R and NLF-C are activated by switches R3 and C2, respectively. Thus, the outputs of non-linear filters NLF-R and NLF-C are added to the outputs of low pass filters LPF-R3 and LPF-C2, respectively.

Controllable filter bank 60 also includes time aligners (TA) which add any necessary delays to ensure that the pixel currently being processed remains at its appropriate location within the frame.

The low pass filters (LPF) are associated with the high pass filters used in analyzer 50. This, the cutoff frequencies of the low pass filters are close to those of the high pass filters. The low pass filters thus pass that which their associated high pass filters ignore.

FIG. 6, to which reference is now briefly made, illustrates exemplary low pass filters for the example provided here-

inabove. Low pass filter LPF-R3 has a cutoff frequency of 0.5 MHz and thus, generally does not retain anything which its associated high pass filter HPF-R3 (with a cutoff frequency of 1 MHz) retains. Filter LPF-R2 has a cutoff frequency of 1 MHz, filter LPF-C2 has a cutoff frequency of 1.25 MHz and filters LPF-C1 and LPF-R1 have a cutoff frequency of about 2 MHz. As for the high frequency filters, filters LPF-Cx operate on the columns of the frame and filters LPF-Rx operate on the rows of the frame.

FIG. 7, to which reference is now briefly made, illustrates an exemplary transfer function for the non-linear filters (NLF) which models the response of the eye when detecting a detail. The transfer function defines an output value V_{out} normalized by the threshold level THD, as a function of an input value V_{in} also normalized by the threshold level THD. As can be seen in the figure, the input-output relationship is described by a polynomial of high order. A typical order might be six, though lower orders, of power two or three, are also feasible.

Table 3 lists the type of filters activated per subclass, where the header for the column indicates both the type of filter and the label of the switch SW-X of FIGS. 5A and 5B.

TABLE 3

Low Pass Filters								
Subclass	R1	R2	R3	C1	C2	D-R	D-C	
1	0	0	1	0	1	0	0	
2	0	0	1	1	0	0	0	
3	0	0	1	0	0	0	1	
4	0	1	0	0	1	0	0	
5	0	1	0	1	0	0	0	
6	0	1	0	0	0	0	1	
7	1	0	0	0	1	0	0	
8	1	0	0	1	0	0	0	
9	1	0	0	0	0	0	1	
10	0	0	0	0	1	1	0	
11	0	0	0	1	0	1	0	
12	0	0	0	0	0	1	1	

FIG. 5B includes rounding elements RND which reduce the number of bits of a pixel from eight to three or four bits, depending on the subclass to which the pixel belongs. Table 4 illustrates the logic for the example presented hereinabove, where the items which are not active for the subclass are indicated by "N/A".

TABLE 4

subclass	RND-R0 (Z1)	RND-R1 (Z2)	RND-R2 (Z3)	RND-C0 (Z4)	RND-C1 (Z5)
1	N/A	N/A	N/A	N/A	N/A
2	N/A	N/A	N/A	N/A	4 bit
3	N/A	N/A	N/A	4 bit	N/A
4	N/A	N/A	4 bit	N/A	N/A
5	N/A	N/A	4 bit	N/A	4 bit
6	N/A	N/A	4 bit	4 bit	N/A
7	N/A	4 bit	N/A	N/A	N/A
8	N/A	3 bit	N/A	N/A	3 bit
9	N/A	3 bit	N/A	3 bit	N/A
10	4 bit	N/A	N/A	N/A	N/A
11	3 bit	N/A	N/A	N/A	3 bit
12	3 bit	N/A	N/A	3 bit	N/A

The output of intra-frame processor 44 is a processed version of the current frame which uses fewer bits to describe the frame than the original version.

Reference is now made to FIGS. 8A, 8B and 8C, which illustrate three alternative embodiments for inter-frame processor 48 which provides temporal filtering (i.e. inter-frame

11

filtering) to further process the current frame. Since the present invention provides a full frame as output, inter-frame processor 48 determines which pixels have changed significantly from the previous frame and amends those only, storing the new version in the appropriate location in output frame memory 46.

The embodiments of FIGS. 8A and 8B are open loop versions (i.e. the previous frame is the frame previously input into inter-frame processor 48) while the embodiment of FIG. 8C is a closed loop version (i.e. the previous frame is the frame previously produced by inter-frame processor 48). All of the embodiments comprise a summer 68, a low pass filter (LPF) 70, a high pass filter (HPF) 72, two comparators 74 and 76, two switches 78 and 80, controlled by the results of comparators 74 and 76, respectively, and a summer 82. FIGS. 8A and 8B additionally include an intermediate memory 84 for storing the output of intra-frame processor 44.

Summer 68 takes the difference of the processed current frame, produced by processor 44, and the previous frame, stored in either in intermediate memory 84 (FIGS. 8A and 8B) or in frame memory 46 (FIG. 8C). The difference frame is then processed in two parallel tracks.

In the first track, the low pass filter is used. Each pixel of the filtered frame is compared to a general, large detail, threshold THD-LF which is typically set to 5% of the maximum expected intensity for the frame. Thus, the pixels which are kept are only those which changed by more than 5% (i.e. those whose changes can be "seen" by the human eye).

In the second track, the difference frame is high pass filtered. Since high pass filtering retains the small details, each pixel of the high pass filtered frame is compared to the particular threshold THD, for that pixel, as produced by threshold determiner 54. If the difference pixel has an intensity above the threshold THD, (i.e. the change in the pixel is significant for detailed visual perception), it is allowed through (i.e. switch 80 is set to pass the pixel).

Summer 82 adds the filtered difference pixels passed by switches 78 and/or 80 with the pixel of the previous frame to "produce the new pixel". If switches 78 and 80 did not pass anything, the new pixel is the same as the previous pixel. Otherwise, the new pixel is the sum of the previous pixel and the low and high frequency components of the difference pixel.

Reference is now briefly made to FIGS. 9, 10A, 10B and 11 which detail elements of frame analyzer 42. In these figures, the term "ML" indicates a memory line of the current frame, "MP" indicates a memory pixel of the current frame, "MF" indicates a memory frame, "VD" indicates the vertical drive signal, "TA" indicates a time alignment, e.g. a delay, and CNT indicates a counter.

FIG. 9 generally illustrates the operation of spatial-temporal analyzer 50 and FIGS. 10A and 10B provide one detailed embodiment for the spatial analysis and temporal analysis portions 51 and 53, respectively. FIG. 11 details parameter estimator 52, threshold determiner 54 and subclass determiner 56. As these figures are deemed to be self-explanatory, no further explanation will be included here.

It is noted that the present invention can be implemented with a field programmable gate array (FPGA) and the frame memory can be implemented with SRAM or SDRAM.

The methods and apparatus disclosed herein have been described without reference to specific hardware or software. Rather, the methods and apparatus have been described in a manner sufficient to enable persons of ordi-

12

nary skill in this art to readily adapt commercially available hardware and software as may be needed to reduce any of the embodiments of the present invention to practice without undue experimentation and using conventional techniques.

It will be appreciated by persons skilled in the art that the present invention is not limited by what has been particularly shown and described herein above. Rather the scope of the invention is defined by the claims that follow:

What is claimed is:

1. A visual lossless encoder for processing a video frame prior to compression by a video encoder, the encoder comprising:

threshold means for identifying a plurality of visual perception threshold levels to be associated with the pixels of said video frame, wherein said threshold levels define contrast levels above which a human eye can distinguish a pixel from among its neighboring pixels of said frame;

filter means for dividing said video frame into portions having different detail dimensions;

means, utilizing said threshold levels and said detail dimensions, for associating said pixels of said video frame into subclasses, wherein each subclass includes pixels related to the same detail and which generally cannot be distinguished from each other; and

means for altering the intensity of each pixel of said video frame according to its subclass.

2. An encoder according to claim 1 wherein said means for altering comprises an inter-frame processor and an intra-frame processor.

3. An encoder according to claim 2 and wherein said intra-frame processor comprises a controllable filter bank having a plurality of different filters and a filter selector which selects one of said filters for each pixel according to its subclass.

4. An encoder according to claim 2 and wherein said inter-frame processor comprises a low pass filter and a high pass filter operative on a difference frame between a current frame and a previous frame, large and small detail threshold elements for thresholding the filtered difference frame with a large detail threshold level and a small detail threshold level, respectively, and a summer which sums the outputs of said two filters as amended by said threshold elements.

5. An encoder according to claim 1 and wherein said threshold means comprises means for generating a plurality of picture parameters describing at least one of the following parameters: the volume of information in said frame, the per pixel color and the cross-frame change of intensity and means for generating said virtual perception threshold from said at least one parameter.

6. A method of visual lossless encoding of frames of a video signal, the method comprising steps of:

spatially and temporally separating and analyzing details of said frames;

estimating parameters of said details;

defining a visual perception threshold for each of said details in accordance with said estimated detail parameters;

classifying said frame picture details into subclasses in accordance with said visual perception thresholds and said detail parameters; and

transforming each said frame detail in accordance with its associated subclass.

7. The method according to claim 6 and wherein said step of separating and analyzing also includes the step of spatial high pass filtering of small dimension details and temporal filtering for detail motion analysis.

13

8. The method according to claim 6 and wherein the step of estimating comprises at least one of the following steps:

- determining ΔI_i , a per-pixel signal intensity change between a current frame and a previous frame, normalized by a maximum intensity;
- determining a NI_{IFP} , a normalized volume of intraframe change by high frequency filtering of said frame, summing the intensities of said filtered frame and normalizing said sum by the maximum possible amount of information within a frame;
- generating NI_{IF} , a volume of inter-frame changes between a current frame and its previous frame normalized by a maximum possible amount of information volume within a frame;
- generating NI_{GOP} , a normalized volume of inter-frame changes for a group of pictures from the output of said previous step of generating;
- evaluating a signal-to-noise ratio SNR by high pass filtering a difference frame between said current frame and said previous frame by selecting those intensities of said difference frame lower than a threshold defined as three times a noise level under which noise intensities are not perceptible to the human eye, summing the intensities of the pixels in the filtered frame and normalizing said sum by said maximum intensity and by the total number of pixels in a frame;
- generating NY_i , a normalized intensity value per-pixel;
- generating a per-pixel color saturation level p_i ;
- generating a per-pixel hue value h_i ; and
- determining a per-pixel response $R(h_i)$ to said hue value.

9. The method according to claim 8, wherein said step of defining includes the step of producing said visual perception thresholds, per-pixel, from a minimum threshold value and at least one of said parameters.

14

10. The method according to claim 8, wherein said step of defining includes the step of producing said visual perception thresholds, per-pixel, according to the following equation,

$$THD_i = THD_{min}(1) + \Delta I_i + NI_{IF} +$$

$$NI_{IF} + NI_{GOP} + NY_i + p_i + (1 - R_i(h_i)) + \frac{200}{SNR}$$

wherein THD_{min} is a minimum threshold level.

11. The method according to claim 6, wherein said step of classifying includes the steps of comparing multiple spatial high frequency levels of a pixel against its associated visual perception threshold and processing said comparison results to associate said pixel with one of said subclasses.

12. The method according to claim 6, wherein said step of transforming includes the step of filtering each subclass with an associated two dimensional low pass filter.

13. The method according to claim 12, wherein said step of transforming includes the steps of generating a difference frame between said current frame and a previous transformed frame, low and high pass filtering of said difference frame, comparing said filtered frames with a large detail threshold and a small detail threshold and summing those portions of said filtered frames which are greater than said thresholds.

14. The method according to claim 13 and wherein said large detail threshold is 2 to 5 percent.

15. The method according to claim 6 and also including the step of rounding the output of said step of transforming.

16. The method according to claim 6, and wherein said intensity is a luminance value.

17. The method according to claim 6, and wherein said intensity is a chrominance value.

* * * *

PATENT APPLICATION SERIAL NO. _____

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

03/23/2000 TWILLIAM 00000018 050649 09524618
01 FC:101 690.00 CH

PTO-1556
(5/87)

*U.S. GPO: 1999-452-082/19144

52461P

ABSTRACT

A visual lossless encoder for processing a video frame prior to compression by a video encoder includes a threshold unit, a filter unit, an association unit and an altering unit. The threshold unit identifies a plurality of visual perception threshold levels to be associated with the pixels of the video frame, wherein the threshold levels define contrast levels above which a human eye can distinguish a pixel from among its neighboring pixels of the frame. The filter unit divides the video frame into portions having different detail dimensions. The association unit utilizes the threshold levels and the detail dimensions to associate the pixels of the video frame into subclasses. Each subclass includes pixels related to the same detail and which generally cannot be distinguished from each other. The altering unit alters the intensity of each pixel of the video frame according to its subclass.

004450.034400

15

30

**METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE
SYNTACTIC ENCODING**

Semion SHERAIZIN and Vitaly SHERAIZIN

FIELD OF THE INVENTION

The present invention relates generally to processing of video images and, in particular, to syntactic encoding of images for later compression by standard compression techniques.

BACKGROUND OF THE INVENTION

10 There are many types of video signals, such as digital broadcast television (TV), video conferencing, interactive TV, etc. All of these signals, in their digital form, are divided into frames, each of which consists of many pixels (image elements), each of which requires 8 – 24 bits to describe them. The result is megabits of data per frame.

15 Before storing and / or transmitting these signals, they typically are compressed, using one of many standard video compression techniques, such as JPEG, MPEG, H-compression, etc. These compression standards use video signal transforms and intra- and inter-frame coding which exploit spatial and temporal correlations among pixels of a frame and across frames.

20 However, these compression techniques create a number of well-known, undesirable and unacceptable artifacts, such as blockiness, low resolution and wiggles, among others. These are particularly problematic for broadcast TV (satellite TV, cable TV, etc.) or for systems with very low bit rates (video conferencing, videophone).

Much research has been performed to try and improve the standard compression techniques. The following patents and articles discuss various prior art methods to do so:

US Patents 5,870,501, 5,847,766, 5,845,012, 5,796,864, 5,774,593,
5 5,586,200, 5,491,519, 5,341,442;

Raj Malluri et al, "A Robust, Scalable, Object-Based Video Compression Technique for Very Low Bit-Rate Coding," IEEE Transactions of Circuit and Systems for Video Technology, vol. 7, No. 1, Feb. 1997;

AwadKh. Al-Asmari, "An Adaptive Hybrid Coding Scheme for HDTV
10 and Digital Sequences," IEEE Transactions on Consumer Electronics, vol. 42, No. 3, pp. 926-936, Aug. 1995;

Kwok-tung Lo and Jian Feng, "Predictive Mean Search Algorithms for Fast VQ Encoding of Images," IEEE Transactions On Consumer Electronics, vol. 41, No. 2, pp. 327-331, May 1995;

15 James Goel et al. "Pre-processing for MPEG Compression Using Adaptive Spatial Filtering", IEEE Transactions On Consumer Electronics, " vol. 41, No. 3, pp. 687-698, Aug. 1995;

Jian Feng et al. "Motion Adaptive Classified Vector Quantization for ATM Video Coding", IEEE Transactions on Consumer Electronics, vol. 41, No.
20 2, p. 322-328, May 1995;

Austin Y. Lan et al., "Scene-Context Dependent Reference - Frame Placement for MPEG Video Coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 9, No. 3, pp. 478-489, Apr. 1999;

Kuo-Chin Fan, Kou-Sou Kan, "An Active Scene Analysis-Based
25 approach for Pseudoconstant Bit-Rate Video Coding", IEEE Transactions on

Circuits and Systems for Video Technology, vol. 8 No. 2, pp. 159-170, Apr. 1998;

Takashi Ida and Yoko Sambansugi, "Image Segmentation and Contour Detection Using Fractal Coding", IEEE Transactions on Circuits and Systems for Video Technology, vol. 8, No. 8, pp. 968-975, Dec. 1998;

Liang Shen and Rangaraj M. Rangayyan, "A Segmentation-Based Lossless Image Coding Method for High-Resolution Medical Image Compression," IEEE Transactions on Medical Imaging, vol. 16, No. 3, pp. 301-316, June 1997;

10 Adrian Munteanu et al., "Wavelet-Based Lossless Compression of Coronary Angiographic Images", IEEE Transactions on Medical Imaging, vol. 18, No. 3, p. 272-281, March 1999; and

Akira Okumura et al., "Signal Analysis and Compression Performance Evaluation of Pathological Microscopic Images," IEEE Transactions on Medical
15 Imaging, vol. 16, No. 6, pp. 701-710, Dec. 1997.

SUMMARY OF THE INVENTION

An object of the present invention is to provide a method and apparatus for video compression which is generally lossless vis-a-vis what the human eye perceives.

5 There is therefore provided, in accordance with a preferred embodiment of the present invention, a visual lossless encoder for processing a video frame prior to compression by a video encoder. The encoder includes a threshold determination unit, a filter unit, an association unit and an altering unit. The threshold determination unit identifies a plurality of visual perception
10 threshold levels to be associated with the pixels of the video frame, wherein the threshold levels define contrast levels above which a human eye can distinguish a pixel from among its neighboring pixels of the frame. The filter unit divides the video frame into portions having different detail dimensions. The association unit utilizes the threshold levels and the detail dimensions to
15 associate the pixels of the video frame into subclasses. Each subclass includes pixels related to the same detail and which generally cannot be distinguished from each other. The altering unit alters the intensity of each pixel of the video frame according to its subclass.

Moreover, in accordance with a preferred embodiment of the present
20 invention, the altering unit includes an inter-frame processor and an intra-frame processor.

Furthermore, in accordance with a preferred embodiment of the present invention, the intra-frame processor includes a controllable filter bank having a plurality of different filters and a filter selector which selects one of the
25 filters for each pixel according to its subclass.

US 2013/034400

Further, in accordance with a preferred embodiment of the present invention, the inter-frame processor includes a low pass filter and a high pass filter operative on a difference frame between a current frame and a previous frame, large and small detail threshold elements for thresholding the filtered
5 difference frame with a large detail threshold level and a small detail threshold level, respectively, and a summer which sums the outputs of the two filters as amended by the threshold elements.

Still further, in accordance with a preferred embodiment of the present invention, the threshold unit includes a unit for generating a plurality of
10 parameters describing at least one of the following parameters: the volume of information in the frame, the per pixel color and the cross-frame change of intensity, and unit for generating the visual perception threshold from at least one of the parameters.

There is also provided, in accordance with a preferred embodiment of
15 the present invention, a method of visual lossless encoding of frames of a video signal. The method includes steps of spatially and temporally separating and analyzing details of the frames, estimating parameters of the details, defining a visual perception threshold for each of the details in accordance with the estimated detail parameters, classifying the frame picture details into
20 subclasses in accordance with the visual perception thresholds and the detail parameters and transforming each the frame detail in accordance with its associated subclass.

Additionally, in accordance with a preferred embodiment of the present invention, the step of separating and analyzing also includes the step of spatial

high pass filtering of small dimension details and temporal filtering for detail motion analysis.

Moreover, in accordance with a preferred embodiment of the present invention, the step of estimating comprises at least one of the following steps:

5 determining ΔI_i , a per-pixel signal intensity change between a current frame and a previous frame, normalized by a maximum intensity;

determining a NI_{XY} , a normalized volume of intraframe change by high frequency filtering of the frame, summing the intensities of the filtered frame and normalizing the sum by the maximum possible amount of information within
10 a frame;

generating NI_F , a volume of inter-frame changes between a current frame and its previous frame normalized by a maximum possible amount of information volume within a frame;

generating NI_{GOP} , a normalized volume of inter-frame changes for a
15 group of pictures from the output of the previous step of generating;

evaluating a signal-to-noise ratio SNR by high pass filtering a difference frame between the current frame and the previous frame by selecting those intensities of the difference frame lower than a threshold defined as three times a noise level under which noise intensities are not perceptible to the human
20 eye, summing the intensities of the pixels in the filtered frame and normalizing the sum by the maximum intensity and the total number of pixels in the frame;

generating NY_i , a normalized intensity value per-pixel;

generating a per-pixel color saturation level p_i ;

generating a per-pixel hue value h_i ; and

25 determining a per-pixel response $R_i(h_i)$ to the hue value.

Further, in accordance with a preferred embodiment of the present invention, the step of defining includes the step of producing the visual perception thresholds, per-pixel, from a minimum threshold value and at least one of the parameters.

5 Still further, in accordance with a preferred embodiment of the present invention, the step of defining includes the step of producing the visual perception thresholds, per-pixel, according to the following equation,

$$THD_i = THD_{min} \left(1 + N\Delta_i + NI_{XT} + NI_F + NI_{GOF} + NY_i + p_i + (1 - R_i(h_i)) + \frac{200}{SNR} \right)$$

10 wherein THD_{min} is a minimum threshold level.

Moreover, in accordance with a preferred embodiment of the present invention, the step of classifying includes the steps of comparing multiple spatial high frequency levels of a pixel against its associated visual perception threshold and processing the comparison results to associate the pixel with one
15 of the subclasses.

Further, in accordance with a preferred embodiment of the present invention, the step of transforming includes the step of filtering each subclass with an associated two-dimensional low pass filter.

Still further, in accordance with a preferred embodiment of the present
20 invention, the step of transforming includes the steps of generating a difference frame between the current frame and a previous transformed frame, low and high pass filtering of the difference frame, comparing the filtered frames with a large detail threshold and a small detail threshold and summing those portions of the filtered frames which are greater than the thresholds.

Additionally, in accordance with a preferred embodiment of the present invention, the large detail threshold is 2 to 5 percent.

Moreover, in accordance with a preferred embodiment of the present invention, the method includes the step of rounding the output of the step of transforming.

Finally, the intensity can be a luminance value or a chrominance value.

09524618.031400
DATED: 01/29/2000

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be understood and appreciated more fully from the following detailed description taken in conjunction with the appended drawings in which:

5 Fig. 1 is an example of a video frame;

Fig. 2 is a block diagram illustration of a video compression system having a visual lossless syntactic encoder, constructed and operative in accordance with a preferred embodiment of the present invention;

10 Fig. 3 is a block diagram illustration of the details of the visual lossless syntactic encoder of Fig. 2;

Fig. 4 is a graphical illustration of the transfer functions for a number of high pass filters useful in the syntactic encoder of Fig. 3;

15 Figs. 5A and 5B are block diagram illustrations of alternative embodiments of a controllable filter bank forming part of the syntactic encoder of Fig. 3;

Fig. 6 is a graphical illustration of the transfer functions for a number of low pass filters useful in the controllable filter bank of Figs. 5A and 5B;

Fig. 7 is a graphical illustration of the transfer function for a non-linear filter useful in the controllable filter bank of Figs. 5A and 5B;

20 Figs. 8A, 8B and 8C are block diagram illustrations of alternative embodiments of an inter-frame processor forming a controlled filter portion of the syntactic encoder of Fig. 3;

Fig. 9 is a block diagram illustration of a spatial-temporal analyzer forming part of the syntactic encoder of Fig. 3;

25 Figs. 10A and 10B are detail illustrations of the analyzer of Fig. 9; and

Fig. 1A is a detail illustration of a frame analyzer forming part of the syntactic encoder of Fig. 3.

004150-81942560

DETAILED DESCRIPTION OF THE PRESENT INVENTION

Applicants have realized that there are different levels of image detail in an image and that the human eye perceives these details in different ways. In particular, Applicants have realized the following:

- 5 1. Picture details whose detection mainly depends on the level of noise in the image occupy approximately 50 – 80% of an image.
2. A visual perception detection threshold for image details does not depend on the shape of the details in the image.
3. A visual perception threshold THD depends on a number of picture
10 parameters, including the general brightness of the image. It does not depend on the noise spectrum.

The present invention is a method for describing, and then encoding, images based on which details in the image can be distinguished by the human eye and which ones can only be detected by it.

- 15 Reference is now made to Fig. 1, which is a grey-scale image of a plurality of shapes of a bird in flight, ranging from a photograph of one (labeled 10) to a very stylized version of one (labeled 12). The background of the image is very dark at the top of the image and very light at the bottom of the image.

- The human eye can distinguish most of the birds of the image.
20 However, there is at least one bird, labeled 14, which the eye can detect but cannot determine all of its relative contrast details. Furthermore, there are large swaths of the image (in the background) which have no details in them.

- The present invention is a method and system for syntactic encoding of video frames before they are sent to a standard video compression unit. The
25 present invention separates the details of a frame into two different types, those

that can only be detected (for which only one bit will suffice to describe each of their pixels) and those which can be distinguished (for which at least three bits are needed to describe the intensity of each of their pixels).

~~Reference is now made to Fig. 2, which illustrates the present invention~~
5 within an image transmission system. Thus, Fig. 2 shows a visual lossless (VLS) encoder 20 connected to a standard video transmitter 22 which includes a video compression encoder 24, such as a standard MPEG encoder, and a modulator 26. VLS encoder 20 transforms an incoming video signal such that
10 more than encoder 24 can do on its own, resulting in a significantly reduced ~~volume bit stream to be transmitted.~~

Modulator 26 modulates the reduced volume bit stream and transmits it to a receiver 30, which, as in the prior art, includes a demodulator 32 and a decoder 34. Demodulator 32 demodulates the transmitted signal and decoder
15 34 decodes and decompresses the demodulated signal. The result is provided to a monitor 36 for display.

It will be appreciated that, although the compression ratios are high in the present invention, the resultant video displayed on monitor 36 is not visually degraded. This is because encoder 20 attempts to quantify each frame of the
20 video signal according to which sections of the frame are more or less distinguished by the human eye. For the less-distinguished sections, encoder 20 either provides pixels of a minimum bit volume, thus reducing the overall bit volume of the frame or smoothes the data of the sections such that video compression encoder 24 will later significantly compress these sections, thus
25 resulting in a smaller bit volume in the compressed frame. Since the human eye

does not distinguish these sections, the reproduced frame is not perceived significantly differently than the original frame, despite its smaller bit volume.

Reference is now made to Fig. 3, which details the elements of VLS encoder 20. Encoder 20 comprises an input frame memory 40, a frame analyzer 42, an Intra-frame processor 44, an output frame memory 46 and an inter-frame processor 48. Analyzer 42 analyzes each frame to separate it into subclasses, where subclasses define areas whose pixels cannot be distinguished from each other. Intra-frame processor 44 spatially filters each pixel of the frame according to its subclass and, optionally, also provides each pixel of the frame with the appropriate number of bits. Inter-frame processor 48 provides temporal filtering (i.e. inter-frame filtering) and updates output frame memory 46 with the elements of the current frame which are different than those of the previous frame.

It is noted that frames are composed of pixels, each having luminance Y and two chrominance C_r and C_b components, each of which is typically defined by eight bits. VLS encoder 20 generally separately processes the three components. However, the bandwidth of the chrominance signals is half as wide as that of the luminance signal. Thus, the filters (in the x direction of the frame) for chrominance have a narrower bandwidth. The following discussion shows the filters for the luminance signal Y.

Frame analyzer 42 comprises a spatial-temporal analyzer 50, a parameter estimator 52, a visual perception threshold determiner 54 and a subclass determiner 56. Details of these elements are provided in Figs. 9 – 11, discussed hereinbelow.

As discussed hereinabove, details which the human eye distinguishes are ones of high contrast and ones whose details have small dimensions. Areas of high contrast are areas with a lot of high frequency content. Thus, spatial-temporal analyzer 50 generates a plurality of filtered frames from the current frame, each filtered through a different high pass filter (HPF), where each high pass filter retains a different range of frequencies therein.

Fig. 4, to which reference is now briefly made, is an amplitude vs. frequency graph illustrating the transfer functions of an exemplary set of high pass filters for frames in a non-interlacing scan format. Four graphs are shown. It can be seen that the curve labeled HPF-R3 has a cutoff frequency of 1MHz and thus, retains portions of the frame with information above 1MHz. Similarly, curve HPF-R2 has a cutoff frequency of 2 MHz, HPF-C2 has a cutoff frequency of 3MHz and HPF-R1 and HPF-C1 have a cutoff frequency of 4MHz. As will be discussed hereinbelow, the terminology "Rx" refers to operations on a row of pixels while the terminology "Cx" refers to operations on a column of pixels.

In particular, the filters of Fig. 4 implement the following finite impulse response (FIR) filters on either a row of pixels (the x direction of the frame) or a column of pixels (the y direction of the frame), where the number of pixels used in the filter defines the power of the cosine. For example, a filter implementing $\cos^{10}x$ takes 10 pixels around the pixel of interest, five to one side and five to the other side of the pixel of interest.

$$\text{HPF-R3: } 1 - \cos^{10}x$$

$$\text{HPF-R2: } 1 - \cos^6x$$

$$\text{HPF-R1: } 1 - \cos^2x$$

$$\text{HPF-C2: } 1 - \cos^4y$$

HPF-C1: $1 - \cos^2 y$

The high pass filters can also be considered as digital equivalents of optical apertures. The higher the cut-off frequency, the smaller the aperture. Thus, filters HPF-R1 and HPF-C1 retain only very small details in the frame (of 5 1 – 4 pixels in size) while filter HPF-R3 retains much larger details (of up to 11 pixels).

In the following, the filtered frames will be labeled by the type of filter (HPF-X) used to create them.

Returning to Fig. 3, analyzer 50 also generates difference frames 10 between the current frame and another, earlier frame. The previous frame is typically at most 15 frames earlier. A "group" of pictures or frames (GOP) is a series of frames for which difference frames are generated.

Parameter estimator 52 takes the current frame and the filtered and difference frames and generates a set of parameters that describe the 15 information content of the current frame. The parameters are determined on a pixel-by-pixel basis or on a per frame basis, as relevant. It is noted that the parameters do not have to be calculated to great accuracy as they are used in combination to determine a per pixel, visual perception threshold THD_i.

At least some of the following parameters are determined:

20 Signal to noise ratio (SNR): this parameter can be determined by generating a difference frame between the current frame and the frame before it, high pass filtering of the difference frame, summing the intensities of the pixels in the filtered frame, normalized by both the number of pixels N in a frame and the maximum intensity I_{MAX} possible for the pixel. If the frame is a 25 television frame, the maximum intensity is 255 quanta (8 bits). The high

frequency filter selects only those intensities lower than 3σ , where σ indicates a level less than which the human eye cannot perceive noise. For example, σ can be 46dB, equivalent to a reduction in signal strength of a factor of 200.

Normalized $N\Delta_i$: this measures the change Δ_i , per pixel i , from the current frame to its previous frame. This value is then normalized by the maximum intensity I_{MAX} possible for the pixel.

Normalized volume of Intraframe change Nl_{xy} : this measures the volume of change in a frame l_{xy} (or how much detail there is in a frame), normalized by the maximum possible amount of information MAX_{INFO} within a frame (i.e. 8 bits per pixel x N pixels per frame). Since the highest frequency range indicates the amount of change in a frame, the volume of change l_{xy} is a sum of the intensities in the filtered frame having the highest frequency range, such as filtered frame HPF-R1.

Normalized volume of Interframe changes Nl_F : this measures the volume of changes l_F between the current frame and its previous frame, normalized by the maximum possible amount of information MAX_{INFO} within a frame. The volume of interframe changes l_F is the sum of the intensities in the difference frame.

Normalized volume of change within a group of frames Nl_{GOP} : this measures the volume of changes l_{GOP} over a group of frames, where the group is from 2 to 15 frames, as selected by the user. It is normalized by the maximum possible amount of information MAX_{INFO} within a frame and by the number of frames in the group.

Normalized luminance level NY_i : Y_i is the luminance level of a pixel in the current frame. It is normalized by the maximum intensity I_{MAX} possible for the pixel.

Color saturation p_i : this is the color saturation level of the i th pixel and it is determined by: $\left[0.78 \left(\frac{C_{r,i} - 128}{160} \right)^2 + 0.24 \left(\frac{C_{b,i} - 128}{126} \right)^2 \right]^{1/2}$ where $C_{r,i}$ and $C_{b,i}$ are the chrominance levels of the i th pixel.

Hue h_i : this is the general hue of the i th pixel and is determined by: $\arctan \left(1.4 \frac{C_{r,i} - 128}{C_{b,i} - 128} \right)$. Alternatively, hue h_i can be determined by interpolating

Table 1, below.

Response to hue $R_i(h_i)$: this is the human vision response to a given hue and is given by Table 1, below. Interpolation is typically used to produce a specific value of the response $R(h)$ for a specific value of hue h .

Table 1					
Color	Y	C_r	C_b	h ($^{\circ}$)	$R(h)$
White	235	128	128	-	-
Yellow	210	16	146	575	0.92
Cyan	170	166	16	490	0.21
Green	145	54	34	510	0.59
Magenta	106	202	222	-	0.2
Red	81	90	240	630	0.3
Blue	41	240	110	475	0.11
Black	15	128	128	-	-

Visual perception threshold determiner 54 determines the visual perception threshold THD_i per pixel as follows:

$$THD_i = THD_{min} \left(1 + N\Delta_i + NI_{xy} + NI_F + NI_{GOF} + NY_i + p_i + (1 - R_i(h_i)) + \frac{200}{SNR} \right)$$

Subclass determiner 56 compares each pixel i of each high pass filtered frame HPF-X to its associated threshold THD_i to determine whether or not that pixel is

significantly present in each filtered frame, where "significantly present" is defined by the threshold level and by the "detail dimension" (i.e. the size of the object or detail in the image of which the pixel forms a part). Subclass determiner 56 then defines the subclass to which the pixel belongs.

5 For the example provided above, if the pixel is not present in any of the filtered frames, the pixel must belong to an object of large size or the detail is only detected but not distinguished. If the pixel is only found in the filtered frame of HPF-C2 or in both frames HPF-C1 and HPF-C2, it must be a horizontal edge (an edge in the Y direction of the frame). If it is found in filtered frames HPF-R3
10 and HPF-C2, it is a single small detail. If the pixel is found only in filtered frames HPF-R1, HPF-R2 and HPF-R3, it is a very small vertical edge. If, in addition, it is also found in filtered frame HPF-C2, then the pixel is a very small, single detail.

The above logic is summarized and expanded in Table 2.

15

Table 2

Subclass	High Pass Filters					Remarks
	R1	R2	R3	C1	C2	
1	0	0	0	0	0	Large detail or detected detail only
2	0	0	0	0	1	Horizontal edge
3	0	0	0	1	1	Horizontal edge
4	0	0	1	0	0	Vertical edge
5	0	0	1	0	1	Single small detail
6	0	0	1	1	1	Single small detail
7	0	1	1	0	0	Vertical edge
8	0	1	1	0	1	Single small detail
9	0	1	1	1	1	Single small detail
10	1	1	1	0	0	Very small vertical edge
11	1	1	1	0	1	Very small single detail
12	1	1	1	1	1	Very small single detail

The output of subclass determiner 56 is an indication of the subclass to which each pixel of the current frame belongs. Intra-frame processor 44 performs spatial filtering of the frame, where the type of filter utilized varies in accordance
20 with the subclass to which the pixel belongs.

In accordance with a preferred embodiment of the present invention, intra-frame processor 44 filters each subclass of the frame differently and according to the information content of the subclass. The filtering limits the bandwidth of each subclass which is equivalent to sampling the data at different frequencies. Subclasses with a lot of content are sampled at a high frequency while subclasses with little content, such as a plain background area, are sampled at a low frequency.

Another way to consider the operation of the filters is that they smooth the data of the subclass, removing "noisiness" in the picture that the human eye does not perceive. Thus, intra-frame processor 44 changes the intensity of the pixel by an amount less than the visual distinguishing threshold for that pixel. Pixels whose contrast is lower than the threshold (i.e. details which were detected only) are transformed with non-linear filters. If desired, the data size of the detected only pixels can be reduced from 8 bits to 1 or 2 bits, depending on the visual threshold level and the detail dimension for the pixel. For the other pixels (i.e. the distinguished ones), 3 or 4 bits is sufficient.

Intra-frame processor 44 comprises a controllable filter bank 60 and a filter selector 62. Controllable filter bank 60 comprises a set of low pass and non-linear filters, shown in Figs. 5A and 5B to which reference is now made, which filter selector 62 activates, based on the subclass to which the pixel belongs. Selector 62 can activate more than one filter, as necessary.

Figs. 5A and 5B are two, alternative embodiments of controllable filter bank 60. Both comprise two sections 64 and 66 which operate on columns (i.e. line to line) and on rows (i.e. within a line), respectively. In each section 64 and 66, there is a choice of filters, each controlled by an appropriate switch, labeled

SW-X, where X is one of C1, C2, R1, R2, R3 (selecting one of the low pass filters (LPF)), D-C, D-R (selecting to pass the relevant pixel directly). Filter selector 62 switches the relevant switch, thereby activating the relevant filter. It is noted that the non-linear filters NLF-R and NLF-C are activated by switches
 5 R3 and C2, respectively. Thus, the outputs of non-linear filters NLF-R and NLF-C are added to the outputs of low pass filters LPF-R3 and LPF-C2, respectively.

Controllable filter bank 60 also includes time aligners (TA) which add any necessary delays to ensure that the pixel currently being processed
 10 remains at its appropriate location within the frame.

The low pass filters (LPF) are associated with the high pass filters used in analyzer 50. Thus, the cutoff frequencies of the low pass filters are close to those of the high pass filters. The low pass filters thus pass that which their associated high pass filters ignore.

15 Fig. 6, to which reference is now briefly made, illustrates exemplary low pass filters for the example provided hereinabove. Low pass filter LPF-R3 has a cutoff frequency of 0.5MHz and thus, generally does not retain anything which its associated high pass filter HPF-R3 (with a cutoff frequency of 1MHz) retains. Filter LPF-R2 has a cutoff frequency of 1MHz, filter LPF-C2 has a cutoff
 20 frequency of 1.25MHz and filters LPF-C1 and LPF-R1 have a cutoff frequency of about 2MHz. As for the high frequency filters, filters LPF-Cx operate on the columns of the frame and filters LPF-Rx operate on the rows of the frame.

Fig. 7, to which reference is now briefly made, illustrates an exemplary transfer function for the non-linear filters (NLF) which models the response of
 25 the eye when detecting a detail. The transfer function defines an output value

Vout normalized by the threshold level THD_i as a function of an input value Vin also normalized by the threshold level THD_i. As can be seen in the figure, the input – output relationship is described by a polynomial of high order. A typical order might be six, though lower orders, of power two or three, are also feasible.

Table 3 lists the type of filters activated per subclass, where the header for the column indicates both the type of filter and the label of the switch SW-X of Figs. 5A and 5B.

Table 3

Subclass	Low Pass Filters					D-R	D-C
	R1	R2	R3	C1	C2		
1	0	0	1	0	1	0	0
2	0	0	1	1	0	0	0
3	0	0	1	0	0	0	1
4	0	1	0	0	1	0	0
5	0	1	0	1	0	0	0
6	0	1	0	0	0	0	1
7	1	0	0	0	1	0	0
8	1	0	0	1	0	0	0
9	1	0	0	0	0	0	1
10	0	0	0	0	1	1	0
11	0	0	0	1	0	1	0
12	0	0	0	0	0	1	1

Fig. 5B includes rounding elements RND which reduce the number of bits of a pixel from eight to three or four bits, depending on the subclass to which the pixel belongs. Table 4 illustrates the logic for the example presented hereinabove, where the items which are not active for the subclass are indicated by "N/A".

Table 4

subclass	RND-R0 (Z1)	RND-R1 (Z2)	RND-R2 (Z3)	RND-C0 (Z4)	RND-C1 (Z5)
1	N/A	N/A	N/A	N/A	N/A
2	N/A	N/A	N/A	N/A	4bit
3	N/A	N/A	N/A	4bit	N/A
4	N/A	N/A	4bit	N/A	N/A
5	N/A	N/A	4bit	N/A	4bit
6	N/A	N/A	4bit	4bit	N/A
7	N/A	4bit	N/A	N/A	N/A
8	N/A	3bit	N/A	N/A	3bit
9	N/A	3bit	N/A	3bit	N/A
10	4bit	N/A	N/A	N/A	N/A

11	3bit	N/A	N/A	N/A	3bit
12	3bit	N/A	N/A	3bit	N/A

The output of intra-frame processor 44 is a processed version of the current frame which uses fewer bits to describe the frame than the original version.

5 ~~Reference is now made to Figs. 8A, 8B and 8C, which illustrate three alternative embodiments for inter-frame processor 48 which provides temporal filtering (i.e. inter-frame filtering) to further process the current frame. Since the present invention provides a full frame as output, inter-frame processor 48 determines which pixels have changed significantly from the previous frame and amends those only, storing the new version in the appropriate location in~~
~~frame memory 46.~~

The embodiments of Figs. 8A and 8B are open loop versions (i.e. the previous frame is the frame previously input into inter-frame processor 48) while the embodiment of Fig. 8C is a closed loop version (i.e. the previous frame is
15 the frame previously produced by inter-frame processor 48). All of the embodiments comprise a summer 68, a low pass filter (LPF) 70, a high pass filter (HPF) 72, two comparators 74 and 76, two switches 78 and 80, controlled by the results of comparators 74 and 76, respectively, and a summer 82. Figs. 8A and 8B additionally include an intermediate memory 84 for storing the
20 output of intra-frame processor 44.

Summer 68 takes the difference of the processed current frame, produced by processor 44, and the previous frame, stored in either in intermediate memory 84 (Figs. 8A and 8B) or in frame memory 46 (Fig. 8C). The difference frame is then processed in two parallel tracks.

parameter estimator 52, threshold determiner 54 and subclass determiner 56. As these figures are deemed to be self-explanatory, no further explanation will be included here.

It is noted that the present invention can be implemented with a field
5 programmable gate array (FPGA) and the frame memory can be implemented with SRAM or SDRAM.

The methods and apparatus disclosed herein have been described without reference to specific hardware or software. Rather, the methods and apparatus have been described in a manner sufficient to enable persons of
10 ordinary skill in the art to readily adapt commercially available hardware and software as may be needed to reduce any of the embodiments of the present invention to practice without undue experimentation and using conventional techniques.

It will be appreciated by persons skilled in the art that the present
15 invention is not limited by what has been particularly shown and described herein above. Rather the scope of the invention is defined by the claims that follow:

CLAIMS

1. A visual lossless encoder for processing a video frame prior to compression by a video encoder, the encoder comprising:
- threshold means for identifying a plurality of visual perception
5 threshold levels to be associated with the pixels of said video frame, wherein said threshold levels define contrast levels above which a human eye can distinguish a pixel from among its neighboring pixels of said frame;
- filter means for dividing said video frame into portions having
10 different detail dimensions;
- means, utilizing said threshold levels and said detail dimensions, for associating said pixels of said video frame into subclasses, wherein each subclass includes pixels related to the same detail and which generally cannot be distinguished from each other;
15 and
- means for altering the intensity of each pixel of said video frame according to its subclass.
2. An encoder according to claim 1 wherein said means for altering comprises an inter-frame processor and an intra-frame processor.
- 20 3. An encoder according to claim 2 and wherein said intra-frame processor comprises a controllable filter bank having a plurality of different filters and a filter selector which selects one of said filters for each pixel according to its subclass.
4. An encoder according to claim 2 and wherein said inter-frame processor
25 comprises a low pass filter and a high pass filter operative on a difference

frame between a current frame and a previous frame, large and small detail
threshold elements for thresholding the filtered difference frame with a large
detail threshold level and a small detail threshold level, respectively, and a
summer which sums the outputs of said two filters as amended by said
5 threshold elements.

5. An encoder according to claim 1 and wherein said threshold means
comprises means for generating a plurality of picture parameters describing at
least one of the following parameters: the volume of information in said frame,
the per pixel color and the cross-frame change of intensity and means for
10 generating said visual perception threshold from said at least one parameter.

6. A method of visual lossless encoding of frames of a video signal, the
method comprising steps of:

spatially and temporally separating and analyzing details of
said frames;

15 estimating parameters of said details;

defining a visual perception threshold for each of said
details in accordance with said estimated detail parameters;

classifying said frame picture details into subclasses in
accordance with said visual perception thresholds and said detail
parameters; and
20

transforming each said frame detail in accordance with its
associated subclass.

7. The method according to claim 6 and wherein said step of separating and
analyzing also includes the step of spatial high pass filtering of small
25 dimension details and temporal filtering for detail motion analysis.

8. The method according to claim 6 and wherein the step of estimating comprises at least one of the following steps:

determining NA_i , a per-pixel signal intensity change between a current frame and a previous frame, normalized by a maximum intensity;

determining a NI_{xy} , a normalized volume of intraframe change by high frequency filtering of said frame, summing the intensities of said filtered frame and normalizing said sum by the maximum possible amount of information within a frame;

generating NI_F , a volume of inter-frame changes between a current frame and its previous frame normalized by a maximum possible amount of information volume within a frame;

generating NI_{GOP} , a normalized volume of inter-frame changes for a group of pictures from the output of said previous step of generating;

evaluating a signal-to-noise ratio SNR by high pass filtering a difference frame between said current frame and said previous frame by selecting those intensities of said difference frame lower than a threshold defined as three times a noise level under which noise intensities are not perceptible to the human eye, summing the intensities of the pixels in the filtered frame and normalizing said sum by said maximum intensity and by the total number of pixels in a frame;

generating NY_i , a normalized intensity value per-pixel;

generating a per-pixel color saturation level p_i ;

generating a per-pixel hue value h_i ; and

determining a per-pixel response $R_i(h_i)$ to said hue value.

9. The method according to claim 8, wherein said step of defining includes the step of producing said visual perception thresholds, per-pixel, from a minimum threshold value and at least one of said parameters.

10. The method according to claim 8, wherein said step of defining includes the step of producing said visual perception thresholds, per-pixel, according to the following equation,

$$THD_i = THD_{min} \left(1 + NA_i + NI_{xy} + NI_p + NI_{GDP} + NY_i + P_i + (1 - R_i(h_i)) + \frac{200}{SNR} \right)$$

10 wherein THD_{min} is a minimum threshold level.

11. The method according to claim 6, wherein said step of classifying includes the steps of comparing multiple spatial high frequency levels of a pixel against its associated visual perception threshold and processing said comparison results to associate said pixel with one of said subclasses.

12. The method according to claim 6, wherein said step of transforming includes the step of filtering each subclass with an associated two dimensional low pass filter.

13. The method according to claim 12, wherein said step of transforming includes the steps of generating a difference frame between said current frame and a previous transformed frame, low and high pass filtering of said difference frame, comparing said filtered frames with a large detail threshold and a small detail threshold and summing those portions of said filtered frames which are greater than said thresholds.

14. The method according to claim 13 and wherein said large detail threshold is 2 to 5 percent.

15. The method according to claim 6 and also including the step of rounding the output of said step of transforming.
16. The method according to claim 6, and wherein said intensity is a luminance value.
- 5 17. The method according to claim 6, and wherein said intensity is a chrominance value.

09554618.031400

DECLARATION AND POWER OF ATTORNEY FOR PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below under my name.

I believe that I am the original, first and sole inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled

**METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE SYNTACTIC
ENCODING**

the Specification of which

☒ is attached hereto
☐ was filed on
as Application Serial No. _____
and was amended on _____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified Specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, 1.56(a).

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any provisional application filed in the United States in accordance with 35 U.S.C. §1.119(e), or any application for patent that has been converted to a Provisional Application within one (1) year of its filing date, or any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

PRIOR FILED APPLICATION(S)

<u>APPLICATION NUMBER</u>	<u>COUNTRY</u>	<u>(DAY/MONTH/YEAR FILED)</u>	<u>PRIORITY CLAIMED</u>
134182	Israel	23-Jan-00	YES

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application listed below, and, insofar as the subject matter of each of the claims of this application is not disclosed in any prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a), which occurred

Att. by Docket No.: P-2883-US

between the filing date of the prior application and the national or PCT international filing date of this application:

APPLICATION NO.	FILING DATE (DAY/MONTH/YEAR)	STATUS - PATENTED, PENDING, ABANDONED
-----------------	------------------------------	---------------------------------------

I hereby appoint as my attorney(s) and agent(s) Heidi M. Brun (Agent, Registration No. 35,104), or Jerome R. Smith, Jr. (Attorney, Registration No. 35,684), or Daniel J. Swirsky (Agent, Registration No. 45, 148) or Mark S. Cohen (Attorney, Registration No. 42, 425) said attorney(s) and agent(s) with full power of substitution and revocation to prosecute this application and transact all business in the Patent and Trademark Office connected therewith.

Please address all correspondence regarding this application to:

Heidi M. Brun
EITAN, PEARL, LATZER, & COHEN-ZEDEK
ONE CRYSTAL PARK, SUITE 210
2011 CRYSTAL DRIVE
ARLINGTON, VA 22202-3709

Direct all telephone calls to (703) 486-0800 and all facsimiles at (703) 486-0600.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further, that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

FULL NAME OF INVENTOR: SHERAIZIN, Semion →
FULL RESIDENCE ADDRESS: 28A Hamagen Street, Mazkeret Batya 76804, Israel
COUNTRY OF CITIZENSHIP: Israel
FULL POST OFFICE ADDRESS: same

SIGNATURE OF INVENTOR 

DATE 13/03/2000

Attorney Docket No.: P-2883-US

FULL NAME OF INVENTOR: SHERAIZIN, Vitaly

FULL RESIDENCE ADDRESS: 28A Hamagen Street, Mazkeret Batya 76804, Israel

COUNTRY OF CITIZENSHIP: Israel

FULL POST OFFICE ADDRESS: same

SIGNATURE OF INVENTOR  

DATE 13/03/2000

09524618 031400

PRINT OF DRAWINGS
AS ORIGINALLY FILED

004700-01542560

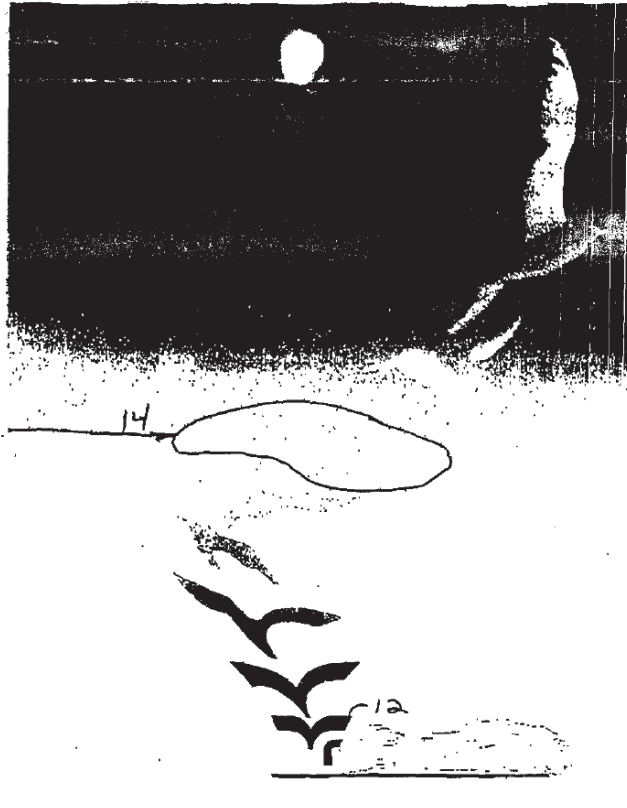


Fig 1

004750-87942560

PRINT OF DRAWINGS
AS ORIGINALLY FILED

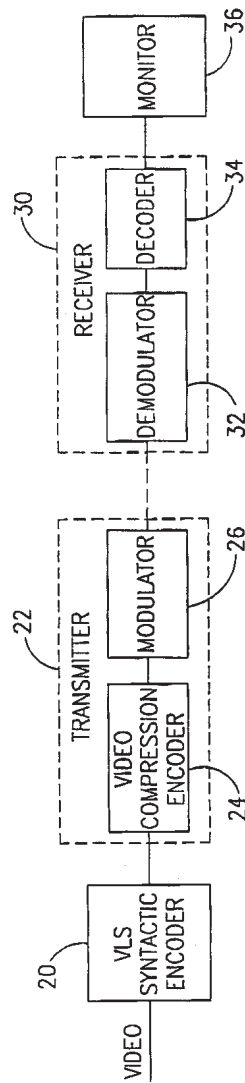


FIG.2

004120 81342560

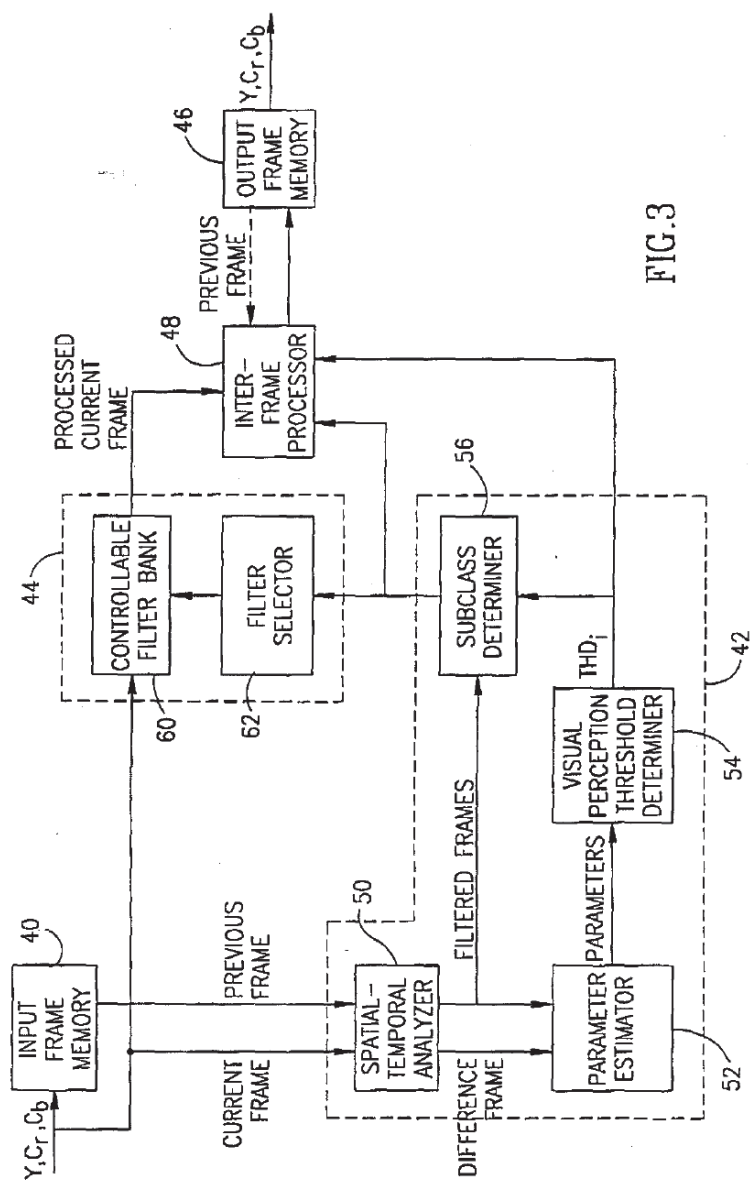


FIG. 3

0047ED-81942560

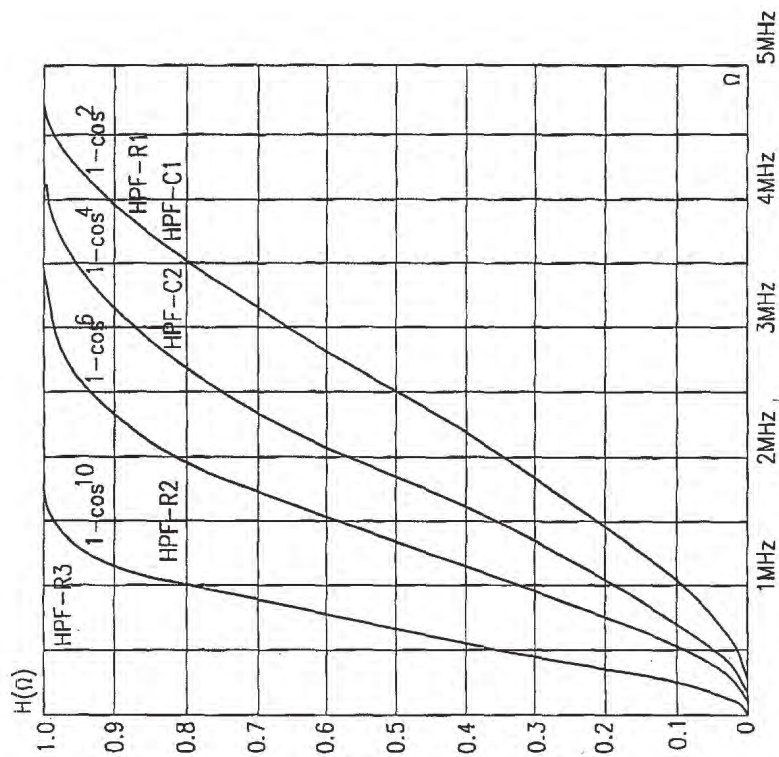


FIG. 4



0011078792500

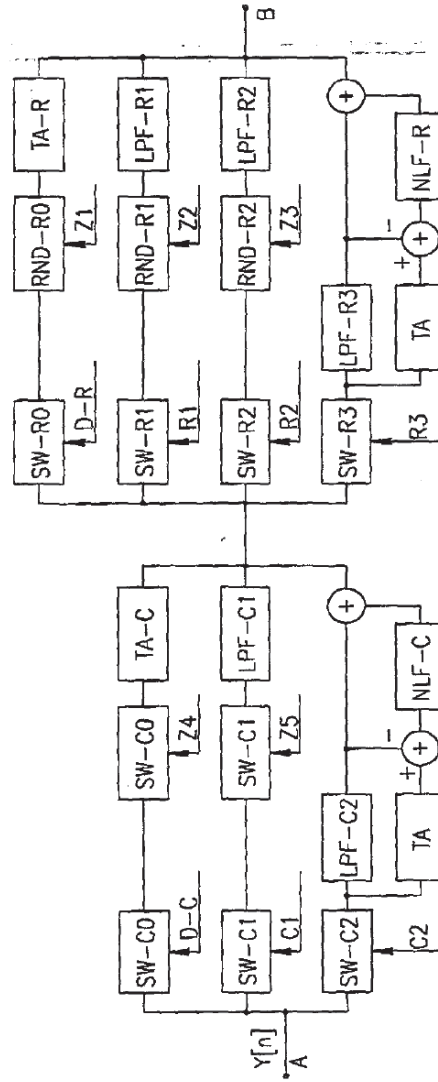


FIG. 5B

004750-87942560

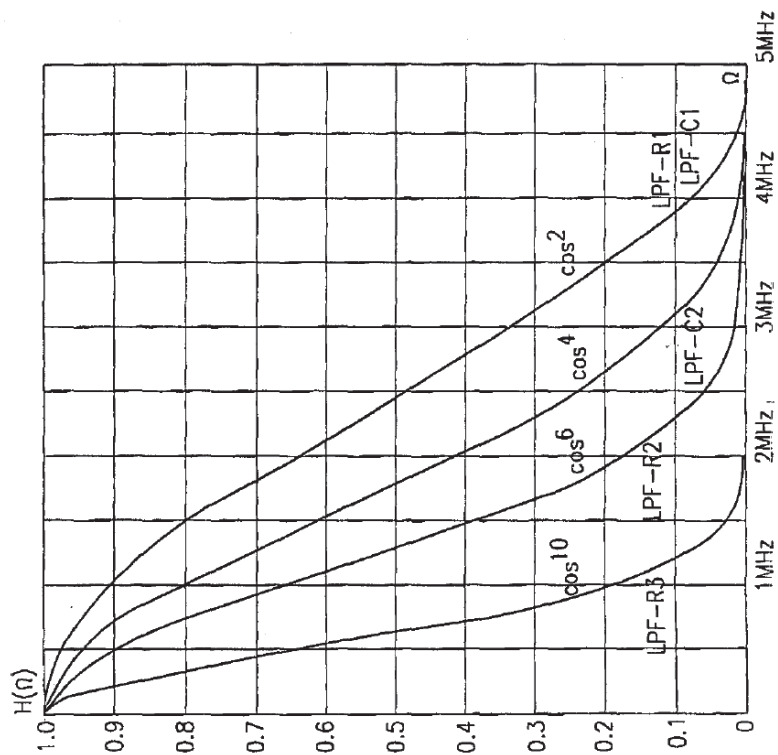


FIG. 6

00110-81942560

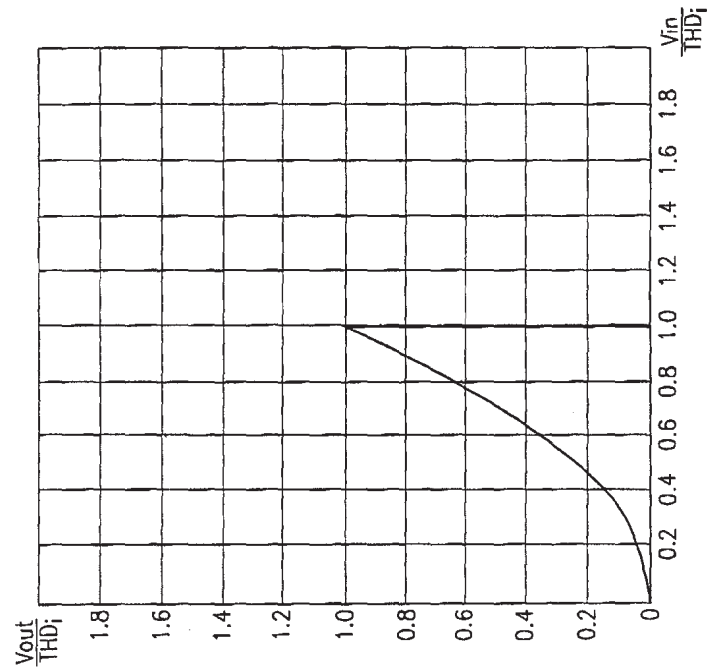


FIG. 7

NOTED 81942560

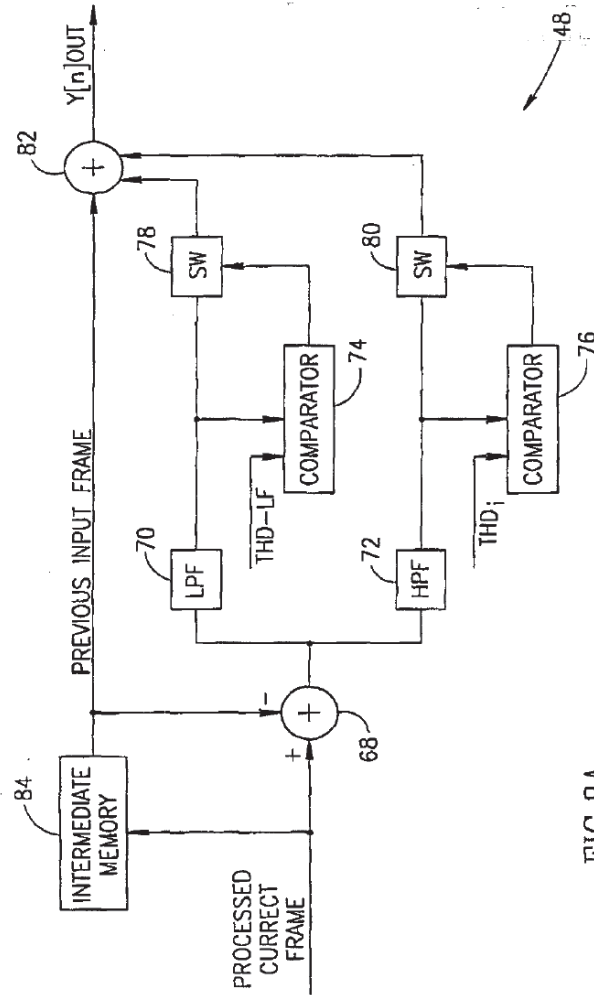


FIG. 8A

004100 81942560

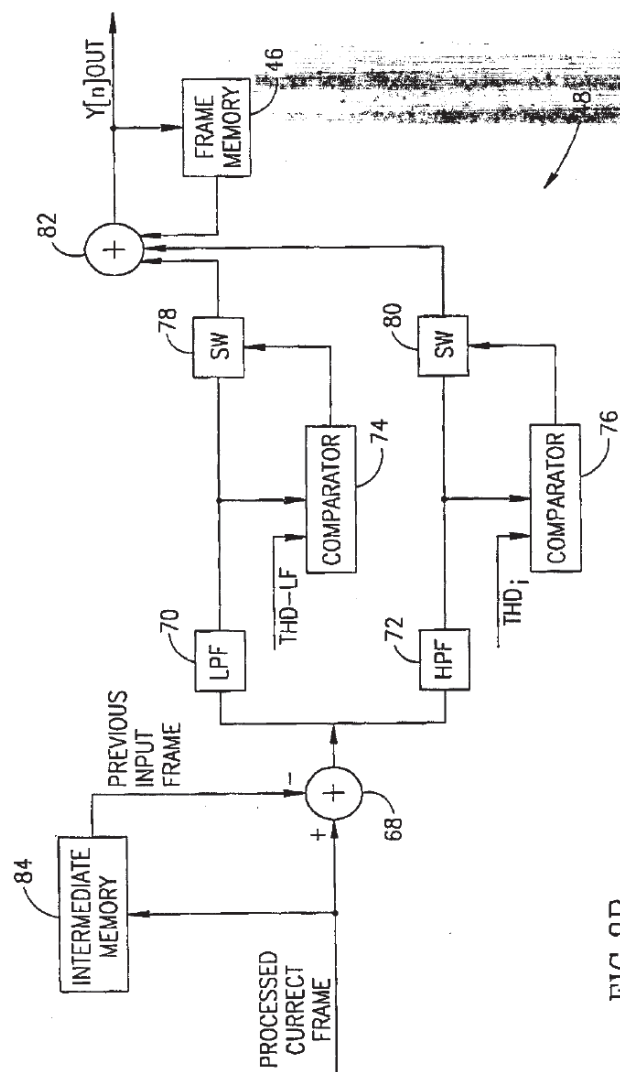


FIG. 8B

DOCKET # 81942560

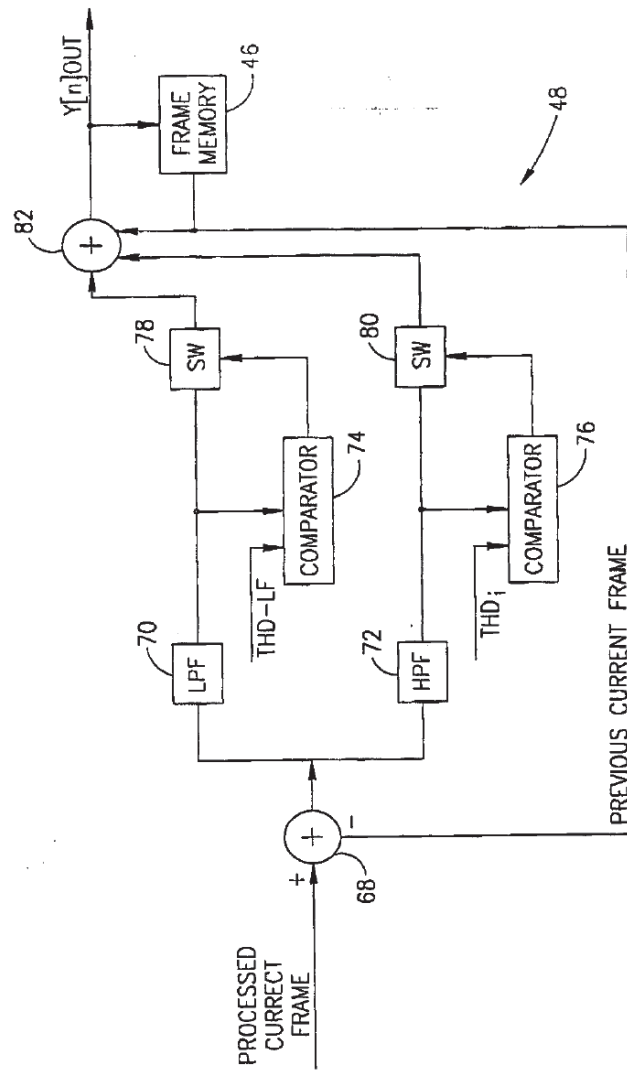


FIG.8C

004450-87942560

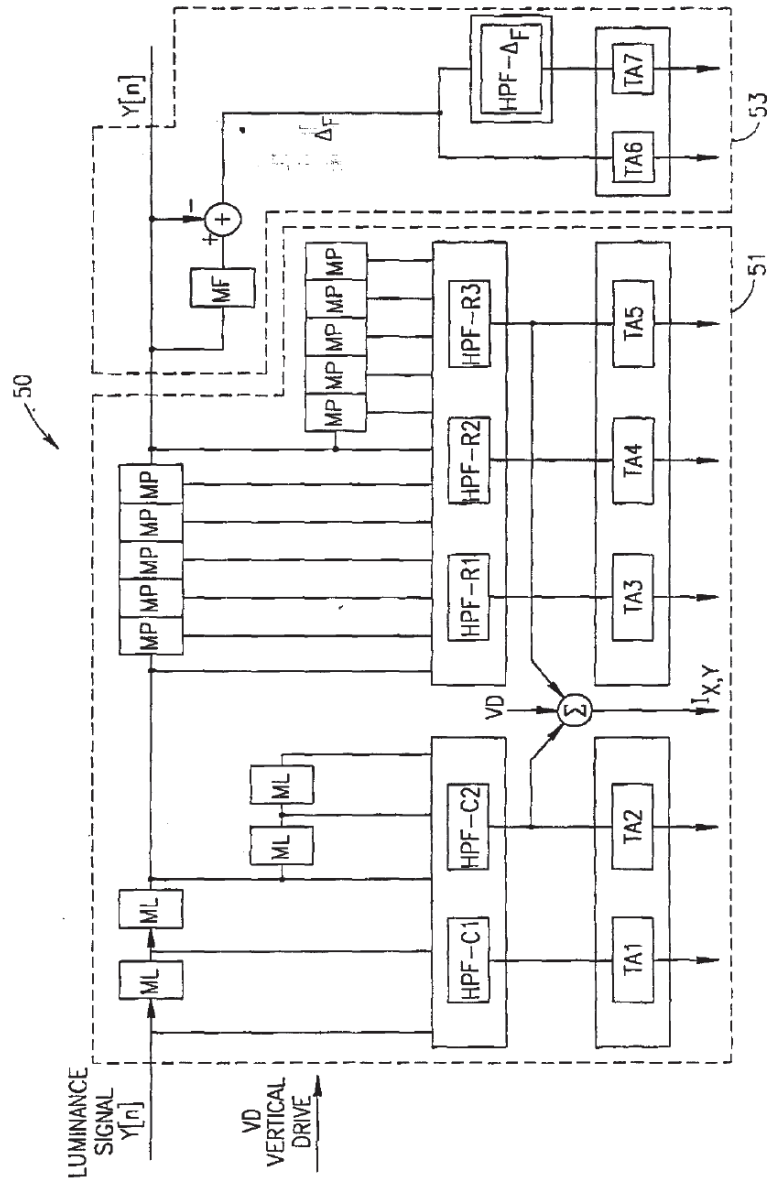


FIG. 9

NOTED 8792550

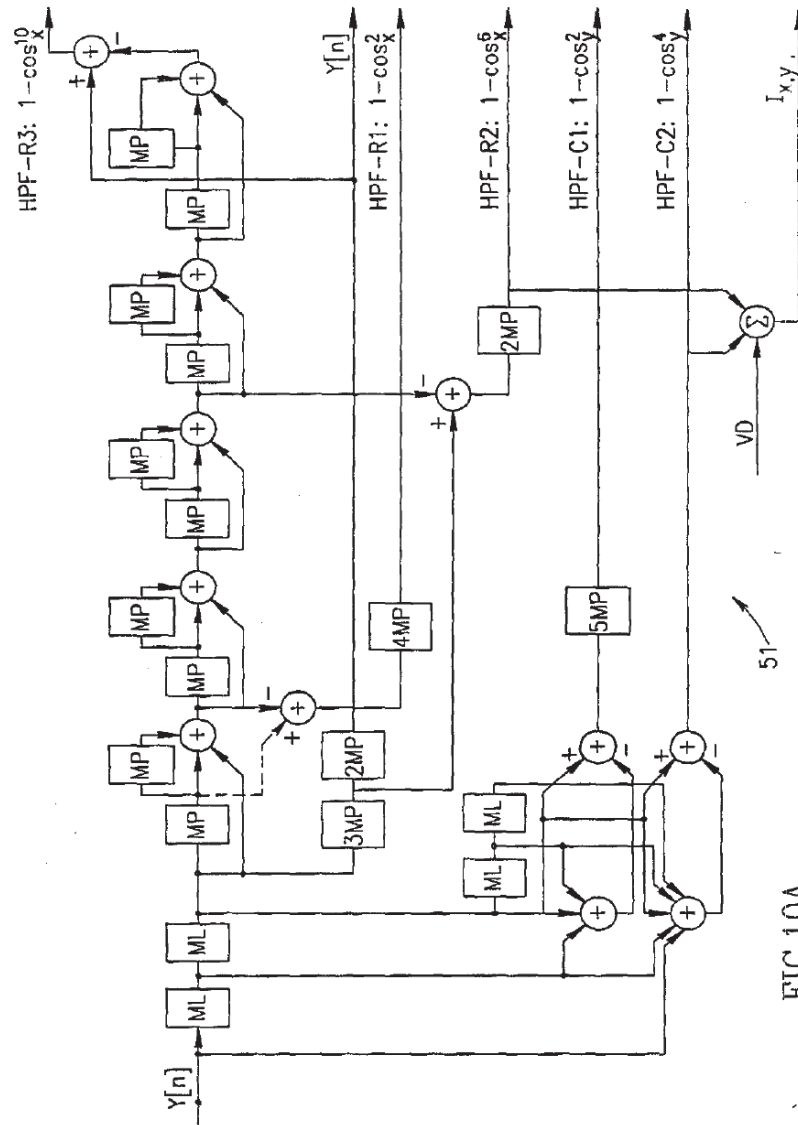
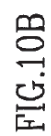
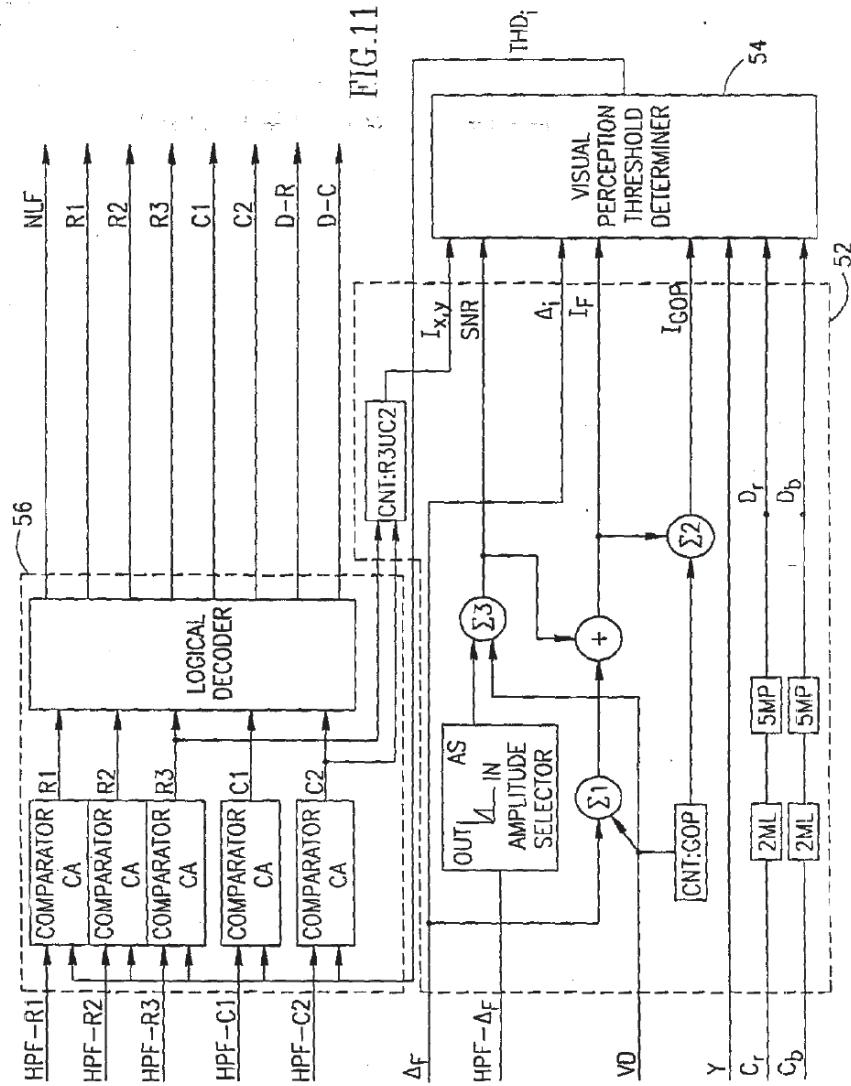


FIG. 10A



UNITED STATES



PTO/SB/17 (12-98)
 Approved for use through 09/30/2000. OMB 0831-0032
 Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

FEE TRANSMITTAL for FY 2000		<i>Complete if Known</i>	
<small>Patent fees are subject to annual revision. Small Entity payments must be supported by a small entity statement, otherwise large entity fees must be paid. See Forms PTO/SB/09-12.</small>			
TOTAL AMOUNT OF PAYMENT (\$)		345.00	
		Application Number	
		Filing Date	
		First Named Inventor	SHERAZIN, Semion
		Examiner Name	
		Group / Art Unit	
		Attorney Docket No.	P-2883-US

<p>METHOD OF PAYMENT (check one)</p> <p>1. <input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge indicated fees and credit any over payments to:</p> <p>Deposit Account Number: 05-0649</p> <p>Deposit Account Name: Eitan, Pearl, Latzer & Cohen-Zedek</p> <p><input checked="" type="checkbox"/> Charge Any Additional Fee Required Under 37 CFR 1.16 and 1.17</p> <p>2. Payment Enclosed:</p> <p><input type="checkbox"/> Check <input type="checkbox"/> Money Order <input type="checkbox"/> Other</p> <p style="text-align: center;">FEE CALCULATION</p> <p>1. BASIC FILING FEE</p> <table border="1" style="width: 100%;"> <thead> <tr> <th>Large Entity Fee Code</th> <th>Small Entity Fee Code</th> <th>Fee Description</th> <th>Fee Paid</th> </tr> </thead> <tbody> <tr> <td>101 890 201</td> <td>345</td> <td>Utility filing fee</td> <td>345.00</td> </tr> <tr> <td>108 310 208</td> <td>155</td> <td>Design filing fee</td> <td></td> </tr> <tr> <td>107 480 207</td> <td>240</td> <td>Plant filing fee</td> <td></td> </tr> <tr> <td>108 690 208</td> <td>345</td> <td>Reissue filing fee</td> <td></td> </tr> <tr> <td>114 150 214</td> <td>75</td> <td>Provisional filing fee</td> <td></td> </tr> <tr> <td colspan="3">SUBTOTAL (1)</td> <td>(\$345.00)</td> </tr> </tbody> </table> <p>2. EXTRA CLAIM FEES</p> <table border="1" style="width: 100%;"> <thead> <tr> <th>Total Claims</th> <th>Extra Claims</th> <th>Fee from Below</th> <th>Fee Paid</th> </tr> </thead> <tbody> <tr> <td>17</td> <td>20**</td> <td>X</td> <td>0</td> </tr> <tr> <td>Independent Claims</td> <td>2</td> <td>3**</td> <td>X</td> <td>0</td> </tr> <tr> <td>Multiple Dependent</td> <td></td> <td>X</td> <td>0</td> </tr> </tbody> </table> <p><small>**or number previously paid, if greater; For Reissues, see below</small></p> <table border="1" style="width: 100%;"> <thead> <tr> <th>Large Entity Fee Code</th> <th>Small Entity Fee Code</th> <th>Fee Description</th> <th>Fee Paid</th> </tr> </thead> <tbody> <tr> <td>103 18 203</td> <td>9</td> <td>Claims in excess of 20</td> <td></td> </tr> <tr> <td>102 78 202</td> <td>39</td> <td>Independent claims in excess of 3</td> <td></td> </tr> <tr> <td>104 260 204</td> <td>130</td> <td>Multiple dependent claim, if not paid</td> <td></td> </tr> <tr> <td>109 78 209</td> <td>39</td> <td>** Reissue independent claims over original patent</td> <td></td> </tr> <tr> <td>110 18 210</td> <td>9</td> <td>** Reissue claims in excess of 20 and over original patent</td> <td></td> </tr> <tr> <td colspan="3">SUBTOTAL (2)</td> <td>(\$0)</td> </tr> </tbody> </table>	Large Entity Fee Code	Small Entity Fee Code	Fee Description	Fee Paid	101 890 201	345	Utility filing fee	345.00	108 310 208	155	Design filing fee		107 480 207	240	Plant filing fee		108 690 208	345	Reissue filing fee		114 150 214	75	Provisional filing fee		SUBTOTAL (1)			(\$345.00)	Total Claims	Extra Claims	Fee from Below	Fee Paid	17	20**	X	0	Independent Claims	2	3**	X	0	Multiple Dependent		X	0	Large Entity Fee Code	Small Entity Fee Code	Fee Description	Fee Paid	103 18 203	9	Claims in excess of 20		102 78 202	39	Independent claims in excess of 3		104 260 204	130	Multiple dependent claim, if not paid		109 78 209	39	** Reissue independent claims over original patent		110 18 210	9	** Reissue claims in excess of 20 and over original patent		SUBTOTAL (2)			(\$0)	<p>3. ADDITIONAL FEES</p> <table border="1" style="width: 100%;"> <thead> <tr> <th>Large Entity Fee Code</th> <th>Small Entity Fee Code</th> <th>Fee Description</th> <th>Fee Paid</th> </tr> </thead> <tbody> <tr> <td>105 130 205</td> <td>65</td> <td>Surcharge - late filing fee or oath</td> <td></td> </tr> <tr> <td>127 50 227</td> <td>25</td> <td>Surcharge - late provisional filing fee or cover sheet</td> <td></td> </tr> <tr> <td>139 130 139</td> <td>130</td> <td>Non-English specification</td> <td></td> </tr> <tr> <td>147 2,520 147</td> <td>2,520</td> <td>For filing a request for reexamination</td> <td></td> </tr> <tr> <td>112 620* 112</td> <td>620*</td> <td>Requesting publication of SIR prior to Examiner action</td> <td></td> </tr> <tr> <td>113 1,840* 113</td> <td>1,840*</td> <td>Requesting publication of SIR after Examiner action</td> <td></td> </tr> <tr> <td>115 110 215</td> <td>55</td> <td>Extension for reply within first month</td> <td></td> </tr> <tr> <td>116 380 216</td> <td>190</td> <td>Extension for reply within second month</td> <td></td> </tr> <tr> <td>117 670 217</td> <td>435</td> <td>Extension for reply within third month</td> <td></td> </tr> <tr> <td>118 1,380 218</td> <td>890</td> <td>Extension for reply within fourth month</td> <td></td> </tr> <tr> <td>128 1,850 228</td> <td>925</td> <td>Extension for reply within fifth month</td> <td></td> </tr> <tr> <td>119 300 219</td> <td>150</td> <td>Notice of Appeal</td> <td></td> </tr> <tr> <td>120 300 220</td> <td>150</td> <td>Filing a brief in support of an appeal</td> <td></td> </tr> <tr> <td>121 280 221</td> <td>130</td> <td>Request for oral hearing</td> <td></td> </tr> <tr> <td>138 1,510 138</td> <td>1,510</td> <td>Petition to institute a public use proceeding</td> <td></td> </tr> <tr> <td>140 110 240</td> <td>55</td> <td>Petition to revive - unavoidable</td> <td></td> </tr> <tr> <td>141 1,210 241</td> <td>605</td> <td>Petition to revive - unintentional</td> <td></td> </tr> <tr> <td>142 1,210 242</td> <td>605</td> <td>Utility issue fee (or reissue)</td> <td></td> </tr> <tr> <td>143 430 243</td> <td>215</td> <td>Design issue fee</td> <td></td> </tr> <tr> <td>144 590 244</td> <td>290</td> <td>Plant issue fee</td> <td></td> </tr> <tr> <td>122 130 122</td> <td>130</td> <td>Petitions to the Commissioner</td> <td></td> </tr> <tr> <td>123 50 123</td> <td>50</td> <td>Petitions related to provisional applications</td> <td></td> </tr> <tr> <td>126 240 126</td> <td>240</td> <td>Submission of Information Disclosure Stmt</td> <td></td> </tr> <tr> <td>581 40 581</td> <td>40</td> <td>Recording each patent assignment per property (times number of properties)</td> <td></td> </tr> <tr> <td>146 760 246</td> <td>380</td> <td>Filing a submission after final rejection (37 CFR 1.129(a))</td> <td></td> </tr> <tr> <td>149 760 249</td> <td>380</td> <td>For each additional invention to be examined (37 CFR 1.129(b))</td> <td></td> </tr> <tr> <td colspan="3">Other fee (specify):</td> <td></td> </tr> <tr> <td colspan="3">Other fee (specify):</td> <td></td> </tr> <tr> <td colspan="3">SUBTOTAL (3)</td> <td>(\$)</td> </tr> </tbody> </table> <p>- Reduced by Basic Filing Fee Paid</p>	Large Entity Fee Code	Small Entity Fee Code	Fee Description	Fee Paid	105 130 205	65	Surcharge - late filing fee or oath		127 50 227	25	Surcharge - late provisional filing fee or cover sheet		139 130 139	130	Non-English specification		147 2,520 147	2,520	For filing a request for reexamination		112 620* 112	620*	Requesting publication of SIR prior to Examiner action		113 1,840* 113	1,840*	Requesting publication of SIR after Examiner action		115 110 215	55	Extension for reply within first month		116 380 216	190	Extension for reply within second month		117 670 217	435	Extension for reply within third month		118 1,380 218	890	Extension for reply within fourth month		128 1,850 228	925	Extension for reply within fifth month		119 300 219	150	Notice of Appeal		120 300 220	150	Filing a brief in support of an appeal		121 280 221	130	Request for oral hearing		138 1,510 138	1,510	Petition to institute a public use proceeding		140 110 240	55	Petition to revive - unavoidable		141 1,210 241	605	Petition to revive - unintentional		142 1,210 242	605	Utility issue fee (or reissue)		143 430 243	215	Design issue fee		144 590 244	290	Plant issue fee		122 130 122	130	Petitions to the Commissioner		123 50 123	50	Petitions related to provisional applications		126 240 126	240	Submission of Information Disclosure Stmt		581 40 581	40	Recording each patent assignment per property (times number of properties)		146 760 246	380	Filing a submission after final rejection (37 CFR 1.129(a))		149 760 249	380	For each additional invention to be examined (37 CFR 1.129(b))		Other fee (specify):				Other fee (specify):				SUBTOTAL (3)			(\$)
Large Entity Fee Code	Small Entity Fee Code	Fee Description	Fee Paid																																																																																																																																																																																															
101 890 201	345	Utility filing fee	345.00																																																																																																																																																																																															
108 310 208	155	Design filing fee																																																																																																																																																																																																
107 480 207	240	Plant filing fee																																																																																																																																																																																																
108 690 208	345	Reissue filing fee																																																																																																																																																																																																
114 150 214	75	Provisional filing fee																																																																																																																																																																																																
SUBTOTAL (1)			(\$345.00)																																																																																																																																																																																															
Total Claims	Extra Claims	Fee from Below	Fee Paid																																																																																																																																																																																															
17	20**	X	0																																																																																																																																																																																															
Independent Claims	2	3**	X	0																																																																																																																																																																																														
Multiple Dependent		X	0																																																																																																																																																																																															
Large Entity Fee Code	Small Entity Fee Code	Fee Description	Fee Paid																																																																																																																																																																																															
103 18 203	9	Claims in excess of 20																																																																																																																																																																																																
102 78 202	39	Independent claims in excess of 3																																																																																																																																																																																																
104 260 204	130	Multiple dependent claim, if not paid																																																																																																																																																																																																
109 78 209	39	** Reissue independent claims over original patent																																																																																																																																																																																																
110 18 210	9	** Reissue claims in excess of 20 and over original patent																																																																																																																																																																																																
SUBTOTAL (2)			(\$0)																																																																																																																																																																																															
Large Entity Fee Code	Small Entity Fee Code	Fee Description	Fee Paid																																																																																																																																																																																															
105 130 205	65	Surcharge - late filing fee or oath																																																																																																																																																																																																
127 50 227	25	Surcharge - late provisional filing fee or cover sheet																																																																																																																																																																																																
139 130 139	130	Non-English specification																																																																																																																																																																																																
147 2,520 147	2,520	For filing a request for reexamination																																																																																																																																																																																																
112 620* 112	620*	Requesting publication of SIR prior to Examiner action																																																																																																																																																																																																
113 1,840* 113	1,840*	Requesting publication of SIR after Examiner action																																																																																																																																																																																																
115 110 215	55	Extension for reply within first month																																																																																																																																																																																																
116 380 216	190	Extension for reply within second month																																																																																																																																																																																																
117 670 217	435	Extension for reply within third month																																																																																																																																																																																																
118 1,380 218	890	Extension for reply within fourth month																																																																																																																																																																																																
128 1,850 228	925	Extension for reply within fifth month																																																																																																																																																																																																
119 300 219	150	Notice of Appeal																																																																																																																																																																																																
120 300 220	150	Filing a brief in support of an appeal																																																																																																																																																																																																
121 280 221	130	Request for oral hearing																																																																																																																																																																																																
138 1,510 138	1,510	Petition to institute a public use proceeding																																																																																																																																																																																																
140 110 240	55	Petition to revive - unavoidable																																																																																																																																																																																																
141 1,210 241	605	Petition to revive - unintentional																																																																																																																																																																																																
142 1,210 242	605	Utility issue fee (or reissue)																																																																																																																																																																																																
143 430 243	215	Design issue fee																																																																																																																																																																																																
144 590 244	290	Plant issue fee																																																																																																																																																																																																
122 130 122	130	Petitions to the Commissioner																																																																																																																																																																																																
123 50 123	50	Petitions related to provisional applications																																																																																																																																																																																																
126 240 126	240	Submission of Information Disclosure Stmt																																																																																																																																																																																																
581 40 581	40	Recording each patent assignment per property (times number of properties)																																																																																																																																																																																																
146 760 246	380	Filing a submission after final rejection (37 CFR 1.129(a))																																																																																																																																																																																																
149 760 249	380	For each additional invention to be examined (37 CFR 1.129(b))																																																																																																																																																																																																
Other fee (specify):																																																																																																																																																																																																		
Other fee (specify):																																																																																																																																																																																																		
SUBTOTAL (3)			(\$)																																																																																																																																																																																															

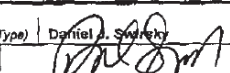
SUBMITTED BY		Complete (if applicable)	
Typed or Printed Name	Daniel J. Swireky	Reg. Number	45,148
Signature	<i>[Signature]</i>	Deposit Account	
Date	March 14, 2000	User ID	05-0649

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

Please type a plus sign (+) inside this box → ☐

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Appr. or use through 09/30/2000. OMB 0651-0002
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

UTILITY PATENT APPLICATION TRANSMITTAL		Attorney Docket No.	P-2883-US
First Inventor or Application Identifier		SHERAZIN, Semion	
Title		METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE SYNTACTIC ENCODING	
Express Mail Label No.			
Only for new nonprovisional applications under 37 C.F.R. § 1.53(b)			
APPLICATION ELEMENTS See MPEP chapter 600 concerning patent application contents		ADDRESS TO: Assistant Commissioner for Patents Box Patent Application Washington, DC 20231	
<p>1. <input checked="" type="checkbox"/> * Fee Transmittal Form (e.g., PTO/SB/17) (Submit an original and a duplicate for fee processing)</p> <p>2. <input checked="" type="checkbox"/> Specification [Total Pages 30] (preferred arrangement set forth below)</p> <ul style="list-style-type: none">- Descriptive title of the invention- Cross References to Related Applications- Reference to Microfiche Appendix- Background of the invention- Brief Summary of the invention- Brief Description of the Drawings (if filed)- Detailed Description- Claim(s)- Abstract of the Disclosure <p>3. <input checked="" type="checkbox"/> Drawing(s) [35 U.S.C. 113] [Total Sheets 15]</p> <p>4. Oath or Declaration [Total Pages 3]</p> <p>a. <input checked="" type="checkbox"/> Newly executed (original or copy)</p> <p>b. <input type="checkbox"/> Copy from a prior application (37 C.F.R. § 1.63(d)) (for continuation/divisional with Box 16 completed)</p> <p>c. <input type="checkbox"/> DELETION OF INVENTOR(S) Signed statement attached deleting inventor(s) named in the prior application, see 37 C.F.R. §§ 1.63(d)(2) and 1.33(b).</p> <p>NOTE FOR ITEMS 1 & 13: IN ORDER TO BE ENTITLED TO PAY SMALL ENTITY FEES, A SMALL ENTITY STATEMENT IS REQUIRED (37 C.F.R. § 1.27) EXCEPT IF ONE FILED IN A PRIOR APPLICATION IS RELIED UPON (37 C.F.R. § 1.28)</p>		<p>5. <input type="checkbox"/> Microfiche Computer Program (Appendix)</p> <p>6. Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary)</p> <p>a. <input type="checkbox"/> Computer Readable Copy</p> <p>b. <input type="checkbox"/> Paper Copy (identical to computer copy)</p> <p>c. <input type="checkbox"/> Statement verifying identity of above copies</p> <p>ACCOMPANYING APPLICATION PARTS</p> <p>7. <input type="checkbox"/> Assignment Papers (cover sheet & document(s))</p> <p>8. <input type="checkbox"/> 37 C.F.R. § 3.73(b) Statement (when there is an assignee) <input type="checkbox"/> Power of Attorney</p> <p>9. <input type="checkbox"/> English Translation Document (if applicable)</p> <p>10. <input type="checkbox"/> Information Disclosure Statement (IDS)/PTO-1449 <input type="checkbox"/> Copies of IDS Citations</p> <p>11. <input type="checkbox"/> Preliminary Amendment</p> <p>12. <input type="checkbox"/> Return Receipt Postcard (MPEP 5303) (Should be specifically itemized)</p> <p>13. <input type="checkbox"/> * Small Entity Statement(s) <input type="checkbox"/> Statement filed in prior application, (PTO/SB/09-12) Status still proper and desired</p> <p>14. <input type="checkbox"/> Certified Copy of Priority Document(s) (if foreign priority is claimed)</p> <p>15. <input checked="" type="checkbox"/> Other: postcard</p>	
16. If a CONTINUING APPLICATION, check appropriate box, and supply the requisite information below and in a preliminary amendment:			
<input type="checkbox"/> Continuation <input type="checkbox"/> Divisional <input type="checkbox"/> Continuation-in-part (CIP) of prior application No.: _____			
Prior application information: Examiner Group/Art Unit:			
For CONTINUATION or DIVISIONAL APPS only: The entire disclosure of the prior application, from which an oath or declaration is supplied under Box 4b, is considered a part of the disclosure of the accompanying continuation or divisional application and is hereby incorporated by reference. The incorporation can only be relied upon when a portion has been inadvertently omitted from the submitted application parts.			
17. CORRESPONDENCE ADDRESS			
<input type="checkbox"/> Customer Number or Bar Code (Insert Customer No. or Attach bar code label here) or <input checked="" type="checkbox"/> Correspondence address below			
Name Eitan, Pearl, Latzer & Cohen-Zedek			
Address One Crystal Park, Suite 210, 2011 Crystal Drive			
City	Arlington	State	VA
Country	USA	Telephone	(703) 486-0800
Zip Code	22202-3709	Fax	(703) 486-0800
Name (Print/Type)	Daniel A. Shtarky	Registration No. (Attorney/Agent)	45,148
Signature		Date	14 March 2000

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any
comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office,
Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents,
Box Patent Application, Washington, DC 20231.



#2

PATENT
P-2883-USIN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT(S): SHERAIZIN, Semion; and SHERAIZIN, Vitaly
SERIAL NO.: 09/524,618 EXAMINER: Not known
FILED: March 14, 2000 GROUP ART UNIT: Not known
FOR: METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE
SYNTACTIC ENCODING

RECEIVED
MAY 23 2000
TC 2700 MAIL ROOMASSISTANT COMMISSIONER FOR PATENTS
WASHINGTON, D.C. 20231


COMMUNICATION FOR FORWARDING
STATEMENT CLAIMING SMALL ENTITY STATUS -
SMALL BUSINESS CONCERN

Sir:

Applicant(s) attach hereby a Statement Claiming Small Entity Status For a Business Concern for the above identified patent Application.

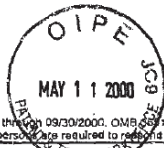
If any additional fee is required, the undersigned attorney hereby authorizes the Patent Office to charge such additional fee to Deposit Account 05-0649.

Respectfully submitted,


Heidi M. Brun
Attorney for Applicant(s)
Registration No. 34,504

Dated: 11 May 2000

Eitan, Pearl, Latzer & Cohen-Zedek
One Crystal Park, Suite 210
2011 Crystal Drive
Arlington, VA 22202-3709
Tel: 703.486.0600
Fax: 703.486.0800



Approved for use through 09/30/2000. OMB 0591-0031 Patent and Trademark Office, U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no person is required to respond to a collection of information unless it displays a valid OMB control number.

PTO/SF/10 (1-95)

STATEMENT CLAIMING SMALL ENTITY STATUS (37 CFR 1.9(f) & 1.27(c))--SMALL BUSINESS CONCERN	Docket Number (Optional) P-2883-US
---	---------------------------------------

Applicant, Patentee, or Identifier: SHERAIZIN, Semion et al.
Application or Patent No.: March 14, 2000
Filed or Issued: 09/524,618
Title: METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE SYNTACTIC ENCODING

I hereby state that I am
☐ the owner of the small business concern identified below:
☒ an official of the small business concern empowered to act on behalf of the concern identified below:

NAME OF SMALL BUSINESS CONCERN VLS COM LTD.

ADDRESS OF SMALL BUSINESS CONCERN 4 Pikeris Street, (Lev Hatamar Building), Rehovot 76545, Israel

I hereby state that the above identified small business concern qualifies as a small business concern as defined in 13 CFR Part 121 for purposes of paying reduced fees to the United States Patent and Trademark Office. Questions related to size standards for a small business concern may be directed to: Small Business Administration, Size Standards Staff, 409 Third Street, SW, Washington, DC 20416.

I hereby state that rights under contract or law have been conveyed to and remain with the small business concern identified above with regard to the invention described in:

☐ the specification filed herewith with title as listed above.
☒ the application identified above.
☐ the patent identified above.

If the rights held by the above identified small business concern are not exclusive, each individual, concern, or organization having rights in the invention must file separate statements as to their status as small entities, and no rights to the invention are held by any person, other than the inventor, who would not qualify as an independent inventor under 37 CFR 1.9(c) if that person made the invention, or by any concern which would not qualify as a small business concern under 37 CFR 1.9(d), or a nonprofit organization under 37 CFR 1.9(e).

Each person, concern, or organization having any rights in the invention is listed below:
☒ no such person, concern, or organization exists.
☐ each such person, concern, or organization is listed below.

Separate statements are required from each named person, concern, or organization having rights to the invention stating their status as small entities. (37 CFR 1.27)

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 CFR 1.28(b))

NAME OF PERSON SIGNING ZAMIR ABERMAN
TITLE OF PERSON IF OTHER THAN OWNER CHAIRMAN & CEO
ADDRESS OF PERSON SIGNING GRANITZKY ST RABINOVITZ ISRAEL
SIGNATURE [Signature] DATE 23/4/2000

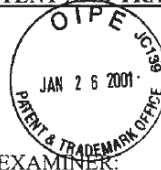
RECEIVED
MAY 23 2000
TO 100 MAIL ROOM

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

#3

PATENT
P-2883-US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE



APPLICANT(S): SHERAIZIN, et al.

SERIAL NO.: 09/524,618

EXAMINER:

unknown

FILED: March 14, 2000

GROUP ART UNIT: unknown

FOR.: METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE
SYNTACTIC ENCODING

ASSISTANT COMMISSIONER FOR PATENTS
WASHINGTON, D.C. 20231

RECEIVED

FEB 01 2001

Technology Center 2600

Sir:

INFORMATION DISCLOSURE STATEMENT

Pursuant to 37 C.F.R. §§1.56, 1.97 and 1.98, this Information Disclosure Statement includes:

1. ☒ Documents including patents, publications, and other information listed on the attached Form PTO-1449 for consideration by the Examiner;
2. ☐ Form PTO-1449 which lists documents including patents, publications and other information for consideration by the Examiner but in accordance with 37 C.F.R. 1.98(d) does not include those documents which have been previously cited or submitted to the Patent Office in the following prior application U.S. Serial No. _____, filed _____ which is properly identified and relied on.
3. ☐ Other information for the Examiner's consideration which was cited in a communication from a foreign patent office in a counterpart foreign application.

The information herein cited is only in fulfillment of Applicant(s) duty of candor in disclosing all information brought to Applicant(s) attention. This submission does not represent that a search has been made or that no better art exists and does not constitute an admission that

APPLICANTS:
SERIAL NO.: 09/524,618
FILED: March 14, 2000
Page 2

each or all of the listed documents are material or constitute "prior art". If it should be determined that any of the listed documents do not constitute "prior art" under United States law, Applicant(s) reserve the right to present to the office the relevant facts and law regarding the appropriate status of such documents.

Applicant(s) further reserve(s) the right to take appropriate action to establish the patentability of the disclosed invention over the listed documents, should one or more of the documents be applied against the claims of the present application.

In accordance with MPEP Sections 609 and 707.05(b), it is requested that each and every document cited (including any cited in applicant's specification which is not repeated on the attached Form PTO-1449) be given thorough consideration and that it be cited of record in the prosecution history of the present application by initialing on Form PTO-1449. Such initialing is requested even if the Examiner does not consider it to be prior art for any reason, or even if the Examiner does not believe that the guidelines for citation have been fully complied with. This is requested so that each document becomes listed on the face of the patent issuing on the present application and is evidence that the Examiner has considered the document.

This Information Disclosure Statement is being filed:

I) ☒ Within three (3) months of filing the subject Application or entry of the subject Application into the national stage or before mailing of the first Office Action on the merits whichever event occurs last pursuant to of 37 C.F.R §1.97 (b); or

II) ☐ After the period specified in (I) but before the mailing date of either a final Official Action under 37 C.F.R §1.113 or a notice of allowance under 37 C.F.R §1.311 whichever occurs first;

1. ☐ The undersigned hereby states that each item of information listed on the Form PTO-1449 was cited in a communication from a foreign Patent Office in a counterpart foreign application not more than three (3) months prior to the filing of this Information Disclosure Statement; or

2. ☐ the undersigned hereby authorizes the Patent Office to charge the fee in the amount of \$240.00 under 37 C.F.R §1.17 (p) to Deposit Account 05-0649.

III) ☐ After the period in (I) and (II) but before the payment of the issue fee,

I, The Undersigned hereby states:

APPLICANTS:

SERIAL NO.:

09/524,618

FILED:

March 14, 2000

Page 3

a) ☐ that each item of information cited on the form PTO-1449 was cited in a communication from a foreign Patent Office in a counterpart foreign application not more than three (3) months prior to the filing of this Information Disclosure Statement; or

b) ☐ that no items of information contained in Form PTO-1449 was cited in a communication from a foreign patent office in a counterpart foreign application, and to the knowledge of the undersigned after making reasonable inquiry, no item of information contained in this Information Disclosure Statement was known to any individual designated in 37 C.F.R. § 1.56(c) more than three months prior to the filing of this Information Disclosure Statement; and

2. The Undersigned submits the following Petition by the Applicant under 37 CFR §197(d) 2:

PETITION

Applicant(s) hereby petition(s) the Patent Office to consider the attached Information Disclosure Statement

3. The Undersigned hereby authorizes the Patent Office to charge the Petition fee in the Amount of \$130.00 under 37 C.F.R. §1.17 (i) to Deposit Account 05-0649.

Except for issue fees payable under 37 C.F.R. §1.18, the Commissioner is hereby authorized by this paper to charge any additional fees during the entire pendency of this application including fees due under 37 C.F.R. §§1.16 and 1.17 which may be required, including any required extension of time fees, or credit any overpayment to Deposit Account No. 05-0649.

Applicants further reserve the right to take appropriate action to establish the patentability of the disclosed invention over the listed documents, should one or more of the documents be applied against the claims of the present application.

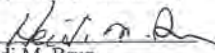
In accordance with MPEP Sections 609 and 707.05(b), it is requested that each document cited (including any cited in applicant's specification which is not repeated on the attached Form PTO-1449) be given thorough consideration and that it be cited of record in the prosecution history of the present application by initialing on Form PTO-1449. Such initialing is requested even if the Examiner does not consider it to be prior art for any reason, or even if the Examiner does not

APPLICANTS:
SERIAL NO.: 09/524,618
FILED: March 14, 2000
Page 4

believe that the guidelines for citation have been fully complied with. This is requested so that each document becomes listed on the face of the patent issuing on the present application.

Except for issue fees payable under 37 C.F.R. §1.18, the Commissioner is hereby authorized by this paper to charge any additional fees during the entire pendency of this application including fees due under 37 C.F.R. §§1.16 and 1.17 which may be required, including any required extension of time fees, or credit any overpayment to Deposit Account No. 05-0649.

Respectfully submitted,


Heidi M. Brun
Attorney for Applicant(s)
Registration No. 34,504

Dated: January 11, 2001

Eitan, Pearl, Latzer & Cohen-Zedek
One Crystal Park, Suite 210
2011 Crystal Drive
Arlington, VA 22202-3709
Tel: 703.486.0600
Fax: 703.486.0800

262

PTO/SB/03A (10-96)

Use through 10/31/99. OMB 0651-0031
U.S. DEPARTMENT OF COMMERCE

Patent and Trademark U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Substitute for form 1449A/PTO		Complete if Known	
INFORMATION DISCLOSURE STATEMENT BY APPLICANT (use as many sheets as necessary)		Application Number	09/524,618
		Filing Date	March 14, 2000
		First Named Inventor	SHERAZIN, et al
		Group Art Unit	unknown 2621
		Examiner Name	unknown PHUOC TRAN
Attorney Docket Number	P-2883-US		
Sheet	1	of	2

[illegible][illegible]

Examiner Signature	<i>Phyllis Tran</i>	Date Considered	12/18/01
-----------------------	---------------------	--------------------	----------

***EXAMINER:** Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

considered. Include copy of this form with next communication to applicant.

¹ Unique citation designation number. ² See attached kinds of U.S. Patent Documents. ³ Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). ⁴ For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. ⁵ Kind of document by the appropriate symbol(s) as indicated on the document under WIPO Standard ST. 16 if possible. ⁶ Applicant is to place a check mark here if English language translation is attached.

Burden Hour Statement: This form is estimated to take 2.0 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. **DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.**

Please type a plus sign (+) inside this box →

Under the Paperwork Reduction Act of 1995, no person is required to respond to a collection of information unless it displays a valid OMB control number. PTO/SB/08A (10-96)
 Patent and Trademark Office U.S. DEPARTMENT OF COMMERCE
 Use through 10/31/99, OMB 0651-0031

Substitute for form 1449B/PTO JAN 26 2001 INFORMATION DISCLOSURE STATEMENT BY APPLICANT (use as many sheets as necessary)		Complete If Known	
Application Number		09/524,618	
Filing Date		March 14, 2000	
First Named Inventor		SHERAZIN, et al	
Group Art Unit		unknown 2621	
Examiner Name		unknown PHUOTRAN	
Attorney Docket Number		P-2883-US	
Sheet	2	of	2

OTHER PRIOR ART - NON PATENT LITERATURE DOCUMENTS			
Examiner Initials*	Cite No. ¹	Include name of the author (in CAPITAL LETTERS), title of the article (where appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T ²
PH	AI	Raj Malluri, et al., "A Robust, Scalable, Object-Based Video Compression Technique for Very Low Bit-Rate Coding", IEEE Transaction of Circuit and Systems for Video Technology, Vol. 7, No.1, Feb.1997	<input type="checkbox"/>
PH	AJ	AwadKh. Al-Asmari, "An Adaptive Hybrid Coding Scheme for HDTV and Digital Sequences," IEEE Transactions on Consumer Electronics, Vol. 42, No. 3, pp. 926-936, Aug. 1995	<input type="checkbox"/>
PH	AK	Kwok-tung Lo & Jian Feng, "Predictive Mean Search Algorithms for Fast VQ Encoding of Images," IEEE Transactions On Consumer Electronics, Vol. 41, No. 2, pp. 327-331, May 1995	<input type="checkbox"/>
PH	AL	James Goei et al., "Pre-processing for MPEG Compression Using Adaptive Spatial Filtering", IEEE Transactions On Consumer Electronics, Vol. 41, No. 3, pp. 687-698, Aug. 1995	<input type="checkbox"/>
PH	AM	Jian Feng, et al., "Motion Adaptive Classified Vector Quantization for ATM Video Coding", IEEE Transactions on Consumer Electronics, Vol. 41, No. 2, p. 322-326, May 1995	<input type="checkbox"/>
PH	AN	Austin Y. Lian, et al., "Scene-Context Dependent Reference - Frame Placement for MPEG Video Coding," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9, No. 3, pp. 478-489, Apr. 1999	<input type="checkbox"/>
PH	AO	Kuo-Chin Fan, Kou-Sou Kan, "An Active Scene Analysis-Based approach for Pseudoconstant Bit-Rate Video Coding", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 8, No. 2, pp. 159-170, April 1998	<input type="checkbox"/>
PH	AP	Takashi Ida and Yoko Sambansugi, "Image Segmentation and Contour Detection Using Fractal Coding", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 8, No. 8, pp. 968-975, December 1998	<input type="checkbox"/>
PH	AQ	Liang Shen & RangarajM. Rangayyan, "A Segmentation-Based Lossless Image Coding Method for High-Resolution Medical Image Compression", IEEE Transactions on Medical Imaging, Vol. 16, No. 3, pp. 301-316, June 1997	<input type="checkbox"/>
PH	AR	Adrian Munteanu et al., "Wavelet-Based Lossless Compression of Coronary Angiographic Images", IEEE Transactions on Medical Imaging, Vol. 18, No. 3, pp. 272-281, March 1999	<input type="checkbox"/>
PH	AS	Akira Okumura, et al., "Signal Analysis and Compression performance Evaluation of Pathological Microscopic Images," IEEE Transactions on Medical Imaging, Vol. 16, No. 6, pp. 701-710, December 1997	<input type="checkbox"/>

Examiner Signature	PHUOTRAN	Date Considered	12/18/01
--------------------	----------	-----------------	----------

* EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

¹ Unique citation designation number. ² Applicant is to place a check mark here if English language Translation is attached.


Burden Hour Statement: This form is estimated to take 2.0 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

RECEIVED

FEB 01 2001

Technology Center 2600

Notice of Allowability	Application No. 09/524,618	Applicant(s) Sheralzin et al.
	Examiner Phuoc Tran	Art Unit 2621



--The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance and Issue Fee Due or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1 ☒ This communication is responsive to paper filed 1/26/01

2 ☒ The allowed claim(s) is/are 1-17

3 The drawings filed on _____ are acceptable as formal drawings.

4 ☒ Acknowledgement is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d).

a) ☒ All b) ☐ Some c) ☐ None of the:

1 ☒ Certified copies of the priority documents have been received.

2 ☐ Certified copies of the priority documents have been received in Application No. _____

3 ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a))

*Certified copies not received: _____

5 ☐ Acknowledgement is made of a claim for domestic priority under 35 U.S.C. § 119(e).

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application. **THIS THREE-MONTH PERIOD IS NOT EXTENDABLE FOR SUBMITTING NEW FORMAL DRAWINGS, OR A SUBSTITUTE OATH OR DECLARATION. This three-month period for complying with the REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL is extendable under 37 CFR 1.136(a).**

6 ☐ Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL APPLICATION (PTO-152) which gives reason(s) why the oath or declaration is deficient. A SUBSTITUTE OATH OR DECLARATION IS REQUIRED.

7 ☐ Applicant MUST submit NEW FORMAL DRAWINGS

(a) ☐ including changes required by the Notice of Draftsperson's Patent Drawing Review (PTO-948) attached

1) ☐ hereto or 2) to Paper No. _____

(b) ☐ including changes required by the proposed drawing correction filed _____, which has been approved by the examiner.

(c) ☐ including changes required by the attached Examiner's Amendment/Comment or in the Office action of Paper No. _____

Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings. The drawings should be filed as a separate paper with a transmittal letter addressed to the Official Draftsperson.

8 ☐ Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Any reply to this letter should include, in the upper right hand corner, the APPLICATION NUMBER (SERIES CODE/SERIAL NUMBER). If applicant has received a Notice of Allowance and Issue Fee Due, the ISSUE BATCH NUMBER and DATE of the NOTICE OF ALLOWANCE should also be included.

Attachment(s)

1 <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) 3 <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) 5 <input checked="" type="checkbox"/> Information Disclosure Statement(s) (PTO-1449), Paper No(s) <u>3</u> 7 <input type="checkbox"/> Examiner's Comment Regarding Requirement for Deposit of Biological Material 9 <input type="checkbox"/> Other	2 <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) 4 <input type="checkbox"/> Interview Summary (PTO-413), Paper No. _____ 6 <input type="checkbox"/> Examiner's Amendment/Comment 8 <input checked="" type="checkbox"/> Examiner's Statement of Reasons for Allowance
---	---

Application/Control Number: 09/524,618

Page 2

Art Unit: 2621

1. The following is an examiner's statement of reasons for allowance: None of the prior art whether taken singly or in combination teaches a visual lossless encoder for processing a video frame according to the claimed invention. The encoder comprises threshold means for identifying a plurality of visual perception threshold levels to be associated with the pixels of the video frame, wherein the threshold levels above which a human eye can distinguish a pixel from among its neighboring pixels of the frame; filter means for dividing the video frame into portions having details dimensions; means, utilizing the threshold levels and detail dimensions, for associating the pixels of the video frame into subclasses, wherein each subclass includes pixels related to the same detail and which generally cannot be distinguished from each other; and means for altering the intensity of each pixel of the video frame according to its subclass. The prior art also fails to teach the corresponding visual lossless encoding method according to the claimed invention.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

2. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Phuoc Tran whose telephone number is (703) 305-4861. The examiner can normally be reached on 9:30 AM-6:00 PM from Monday to Friday.

Application/Control Number: 09/524,618

Page 3

Art Unit: 2621

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Leo H. Boudreau, can be reached on (703) 305-4706.

Any response to this action should be mailed to:

Commissioner of Patents and Trademarks

Washington, D.C. 20231

or faxed to:

(703) 872-9314 (for Technology Center 2600 only)

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is (703) 306-0377.

phuc Tran
[Signature]

Notice of References Cited	Applicant/Patent Sheraizin et al	Application/Control No. 09/524,618	
	Examiner Phuoc Tran	Art Unit 2621	Page 1 of 1

U.S. PATENT DOCUMENTS				
	Document Number <small>Country Code-Number-Kind Code</small>	Date <small>MM-YYYY¹</small>	Name	Classification ²
A	5,901,178	5/1999	Lee et al	375 240
B	5,787,203	7/1998	Lee et al	382 232
C				
D				
E				
F				
G				
H				
I				
J				
K				
L				
M				

FOREIGN PATENT DOCUMENTS					
	Document Number <small>Country Code-Number-Kind Code</small>	Date <small>MM-YYYY¹</small>	Country	Name	Classification ²
N					
O					
P					
Q					
R					
S					
T					

NON-PATENT DOCUMENTS				
Include, as applicable: Author, Title, Date, Publisher, Edition or Volume, Pertinent Pages				
U				
V				
W				
X				

¹ A copy of this reference is not being furnished with this Office action. See MPEP § 707.05(a).

² Dates in MM-YYYY format are publication dates.

³ Classifications may be U.S. or foreign.



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231
www.uspto.gov

NOTICE OF ALLOWANCE AND FEE(S) DUE

7590 12/18/2001

Heidi M Brun
Eitan Pearl Latzer & Cohen-Zedek
One Crystal Park Suite 210
2011 Crystal Drive
Arlington, VA 22202-3709

EXAMINER	
TRAN, PHUOC	
ART UNIT	CLASS-SUBCLASS
2621	382-244000

DATE MAILED: 12/18/2001

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/524,618	03/14/2000	Semion Sheraizin	P-2883-US	8670

TITLE OF INVENTION: METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE SYNTACTIC ENCODING

TOTAL CLAIMS	APPLN. TYPE	SMALL ENTITY	ISSUE FEE	PUBLICATION FEE	TOTAL FEE(S) DUE	DATE DUE
14	nonprovisional	YES	\$640	\$0	\$640	03/18/2002

THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT.

PROSECUTION ON THE MERITS IS CLOSED. THIS NOTICE OF ALLOWANCE IS NOT A GRANT OF PATENT RIGHTS. THIS APPLICATION IS SUBJECT TO WITHDRAWAL FROM ISSUE AT THE INITIATIVE OF THE OFFICE OR UPON PETITION BY THE APPLICANT. SEE 37 CFR 1.313 AND MPEP 1308.

THE ISSUE FEE AND PUBLICATION FEE (IF REQUIRED) MUST BE PAID WITHIN **THREE MONTHS** FROM THE MAILING DATE OF THIS NOTICE OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. **THIS STATUTORY PERIOD CANNOT BE EXTENDED.** SEE 35 U.S.C. 151.

HOW TO REPLY TO THIS NOTICE:

I. Review the SMALL ENTITY status shown above. If the SMALL ENTITY is shown as YES, verify your current SMALL ENTITY status:

- A. If the status is changed, pay the PUBLICATION FEE (if required) and twice the amount of the ISSUE FEE shown above and notify the United States Patent and Trademark Office of the change in status, or
- B. If the status is the same, pay the TOTAL FEE(S) DUE shown above.

If the SMALL ENTITY is shown as NO:

- A. Pay TOTAL FEE(S) DUE shown above, or
- B. If applicant claimed SMALL ENTITY status before, or is now claiming SMALL ENTITY status, check the box below and enclose the PUBLICATION FEE and 1/2 the ISSUE FEE shown above.
- ☐ Applicant claims SMALL ENTITY status.
See 37 CFR 1.27.

II. PART B - FEE(S) TRANSMITTAL should be completed and returned to the United States Patent and Trademark Office (USPTO) with your ISSUE FEE and PUBLICATION FEE (if required). Even if the fee(s) have already been paid, Part B - Fee(s) Transmittal should be completed and returned. If you are charging the fee(s) to your deposit account, section "4b" of Part B - Fee(s) Transmittal should be completed and an extra copy of the form should be submitted.

III. All communications regarding this application must give the application number. Please direct all communications prior to issuance to Box ISSUE FEE unless advised to the contrary.

IMPORTANT REMINDER: Utility patents issuing on applications filed on or after Dec. 12, 1980 may require payment of maintenance fees. It is patentee's responsibility to ensure timely payment of maintenance fees when due.

Page 1 of 3

PTOL-85 (REV. 07-01) Approved for use through 01/31/2004.

PART B - FEE(S) TRANSMITTAL

Complete and mail this form, together with applicable fee(s), to:

Box ISSUE FEE
Assistant Commissioner for Patents
Washington, D.C. 20231

Bc8

MAILING INSTRUCTIONS: This form should be used for transmitting the **ISSUE FEE** and **PUBLICATION FEE** (if required). Blocks 1 through 4 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Legibly mark-up with any corrections or use Block 1)

7590 12/18/2001

Heidi M Brun
 Eitan Pearl Latzer & Cohen-Zedek
 One Crystal Park Suite 210
 2011 Crystal Drive
 Arlington, VA 22202-3709

Note: The certificate of mailing below can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing.

Certificate of Mailing

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Box Issue Fee address above on the date indicated below.

(Depositor's name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/524,618	03/14/2000	Samson Shenzin	P-2893-US	8670

TITLE OF INVENTION: METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE SYNTACTIC ENCODING

TOTAL CLAIMS	APPLN. TYPE	SMALL ENTITY	ISSUE FEE	PUBLICATION FEE	TOTAL FEE(S) DUE	DATE DUE
14	nonprovisional	YES	\$640	\$0	\$640	03/18/2002

EXAMINER	ART UNIT	CLASS-SUBCLASS
TRAN, PHUOC	2621	382-244000

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363). Use of PTO form(s) and Customer Number are recommended, but not required.

- ☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB-122) attached.
☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB-47) attached.

2. For printing on the patent front page, list (1) the names of up to 3 registered patent attorneys or agents OR, alternatively, (2) the name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

- 1 _____
 2 _____
 3 _____

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. Inclusion of assignee data is only appropriate when an assignment has been previously submitted to the USPTO or is being submitted under separate cover. Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE

(B) RESIDENCE: (CITY and STATE OR COUNTRY)

Please check the appropriate assignee category or categories (will not be printed on the patent): ☐ individual ☐ corporation or other private group entity ☐ government

4a. The following fee(s) are enclosed:

- ☐ Issue Fee
☐ Publication Fee
☐ Advance Order - # of Copies _____

4b. Payment of Fee(s):

- ☐ A check in the amount of the fee(s) is enclosed.
☐ Payment by credit card. Form PTO-2038 is attached.
☐ The Commissioner is hereby authorized by charge the required fee(s), or credit any overpayment, to Deposit Account Number _____ (enclose an extra copy of this form).

The COMMISSIONER OF PATENTS AND TRADEMARKS is requested to apply the Issue Fee and Publication Fee (if any) to the application identified above.

(Authorized Signature)

(Date)

NOTE: The Issue Fee and Publication Fee (if required) will not be accepted from anyone other than the applicant; a registered attorney or agent; or the assignee or other party in interest as shown by the records of the United States Patent and Trademark Office.

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending on the needs of the individual case. Any comments on the amount of time required to complete this form should be sent to the Chief Information Officer, United States Patent and Trademark Office, Washington, D.C. 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND FEES AND THIS FORM TO: Box Issue Fee, Assistant Commissioner for Patents, Washington, D.C. 20231

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMIT THIS FORM WITH FEE(S)

Page 2 of 3

PTOL-85 (REV. 07-01). Approved for use through 01/31/2004. OMB 0651-0033

U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/524,618	03/14/2000	Semion Sherniav	P-3883-US	8670
7500 12/18/2001				
Heidi M Brun Eitan Pearl Latzer & Cohen-Zedek One Crystal Park Suite 210 2011 Crystal Drive Arlington, VA 22202-3709 UNITED STATES			EXAMINER TRAN, PHUOC	
			ART UNIT	PAPER NUMBER
			1621	
			DATE MAILED: 12/18/2001	

Determination of Patent Term Extension under 35 U.S.C. 154 (b)
(application filed after June 7, 1995 but prior to May 29, 2000)

The patent term extension is 0 days. Any patent to issue from the above identified application will include an indication of the 0 day extension on the front page.

If a continued prosecution application (CPA) was filed in the above-identified application, the filing date that determines patent term extension is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Application Information Retrieval (PAIR) system. (<http://pair.uspto.gov>)

PART B - FEE(S) TRANSMITTAL

Complete and mail this form, together with applicable fee(s), to:

Box ISSUE FEE
Assistant Commissioner for Patents
Washington, D.C. 20231

MAILING INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 4 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Legibly make-up with any corrections or use Block 1)

7590 12/18/2001

Heidi M Brun
Eitan Pearl Latzer & Cohen-Zedek
One Crystal Park Suite 210
2011 Crystal Drive
Arlington, VA 22202-3709



Note: The certificate of mailing below can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing.

Certificate of Mailing

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Box Issue Fee address above on the date indicated below.

(Depositor's name)

(Signature)

(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/524,618	03/14/2000	Samson Sheralzin	P-2883-US	8670

TITLE OF INVENTION: METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE SYNTACTIC ENCODING

TOTAL CLAIMS	APPLN. TYPE	SMALL ENTITY	ISSUE FEE	PUBLICATION FEE	TOTAL FEE(S) DUE	DATE DUE
14	nonprovisional	YES	\$640	\$0	\$640	03/18/2002

EXAMINER	ART UNIT	CLASS-SUBCLASS
TRAN, PHUOC	2621	382-244000

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.303). Use of PTO form(s) and Customer Number are recommended, but not required.

☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.

☐ "Fee Address" indication (or "Fee Address" indication form PTO/SB/47) attached.

2. For printing on the patent front page, list (1) the names of up to 3 registered patent attorneys or agents OR, alternatively, (2) the name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

Eitan, Pearl, Latzer
& Cohen-Zedek
2
3

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. Inclusion of assignee data is only appropriate when an assignment has been previously submitted to the USPTO or is being submitted under separate cover. Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE

VLS COM LTD.

(B) RESIDENCE: (CITY and STATE OR COUNTRY)

Rechovot, ISRAEL

Please check the appropriate assignee category or categories (will not be printed on the patent) ☐ individual ☒ corporation or other private group entity ☐ government

4a. The following fee(s) are enclosed:

☒ Issue Fee

☐ Publication Fee

☒ Advance Order - # of Copies 7

4b. Payment of Fee(s):

☐ A check in the amount of the fee(s) is enclosed.

☐ Payment by credit card. Form PTO-2038 is attached.

☒ The Commissioner is hereby authorized by charge the required fee(s), or credit any overpayment, to Deposit Account Number 05-0649 (enclose an extra copy of this form).

The COMMISSIONER OF PATENTS AND TRADEMARKS is requested to apply the Issue Fee and Publication Fee (if any) to the application identified above.

(Authorized Signature)

Caleb Pollack

(Date)

March 14, 2002

NOTE: The Issue Fee and Publication Fee (if required) will not be accepted from anyone other than the applicant; a registered attorney or agent, or the assignee or other party in interest as shown by the records of the United States Patent and Trademark Office.

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending on the needs of the individual case. Any comments on the amount of time required to complete this form should be sent to the Chief Information Officer, United States Patent and Trademark Office, Washington, D.C. 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND FEES AND THIS FORM TO: Box Issue Fee, Assistant Commissioner for Patents, Washington, D.C. 20231.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMIT THIS FORM WITH FEE(S)

Page 2 of 3

PTOL-85 (REV. 07-01) Approved for use through 01/31/2004. OMB 0651-0033

U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE



4
Attorney Docket No: P-2883-US

6P/2621
MAR 21 2002
5302
515A

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: SHERAIZIN et al. Examiner: TRAN, P.
Serial No.: 09/524,618 Group Art Unit: 2621
Filed: March 14, 2000
Title: METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE
SYNTACTIC ENCODING

RECEIVED

MAR 21 2002

AMENDMENT UNDER 37 CFR 1.312

Assistant Commissioner for Patents
Washington, DC 20231

Technology Center 2600

Match & Return

Sir:

In response to the Notice of Allowance mailed December 18, 2001, kindly amend the above identified application as follows:

In the Specifications:

✓
Please replace the paragraph beginning at page 12, line 4, with the following rewritten paragraph:

AI
-- Reference is now made to Fig. 2, which illustrates the present invention within an image transmission system. Thus, Fig. 2 shows a visual lossless (VLS) syntactic encoder 20 connected to a standard video transmitter 22 which includes a video compression encoder 24, such as a standard MPEG encoder, and a modulator 26. VLS syntactic encoder 20 transforms an incoming video signal such that video compression encoder 24 can compress the video signal two to five times more than video compression encoder 24 can do on its own, resulting in a significantly reduced volume bit stream to be transmitted.--

APPLICANT: SHERAIZIN, Semion et al.
SERIAL NO.: 09/524,618
FILED: March 14, 2000
Page 2



Please replace the paragraph beginning at page 22, line 5, with the following rewritten paragraph:

A2
--Reference is now made to Figs. 8A, 8B and 8C, which illustrate three alternative embodiments for inter-frame processor 48 which provides temporal filtering (i.e. inter-frame filtering) to further process the current frame. Since the present invention provides a full frame as output, inter-frame processor 48 determines which pixels have changed significantly from the previous frame and amends those only, storing the new version in the appropriate location in output frame memory 46.--

Attached hereto is a marked-up version of the changes made to the specification by the current amendment. The attached page is captioned "Version with markings to show changes made".

RECEIVED
MAR 21 2002
Technology Center 2600

REMARKS

The foregoing amendments and the following remarks are directed to formalities, in order to move the above-referenced patent application to issue. Entry of this paper is respectfully requested.

Amendments to Specification

The amendments to the Specification are either editorial in nature or are provided to match the text with the drawings. No new matter has been added.

Amendments to Drawings

Figs. 5B, 8B and 8C have been amended to match the drawings to the text.

Specifically, in Fig. 5A, frames 64 and 66 were added and in Figs. 8B and 8C, the title "FRAME MEMORY" of element 46 has been corrected to "OUTPUT FRAME MEMORY".


Approval is requested.

APPLICANT: SHERAIZIN, Semion et al.
SERIAL NO.: 09/524,618
FILED: March 14, 2000
Page 3



Should the Examiner have any question or comment as to the form, content, or entry of this Paper, the Examiner is requested to contact the undersigned at the telephone number below. Similarly, if there are any further issues to be resolved to advance the prosecution of this application to issue, the Examiner is requested to telephone the undersigned.

Respectfully submitted,


Caleb Pollack
Attorney for Applicant(s)
Registration No. 37,912

Dated: March 14, 2002

Eitani, Pearl, Latzer & Cohen-Zedek
One Crystal Park, Suite 210, 2011 Crystal Drive
Arlington, VA, USA 22202-3709
Telephone: (703) 486-0600
Fax: (703) 486-0800

RECEIVED
MAR 21 2002
Technology Center 2600

APPLICANT: SHERAIZIN, Semion et al.
SERIAL NO.: 09/524,618
FILED: March 14, 2000
Page 4



RECEIVED

Version with markings to show changes made MAR 21 2002

Technology Center 2600

In the Specification:

Paragraph beginning at page 12, line 4, has been amended as follows:

--Reference is now made to Fig. 2, which illustrates the present invention within an image transmission system. Thus, Fig. 2 shows a visual lossless (VLS) syntactic encoder 20 connected to a standard video transmitter 22 which includes a video compression encoder 24, such as a standard MPEG encoder, and a modulator 26. VLS syntactic encoder 20 transforms an incoming video signal such that video compression encoder 24 can compress the video signal two to five times more than video compression encoder 24 can do on its own, resulting in a significantly reduced volume bit stream to be transmitted.--

Paragraph beginning at page 22, line 5, has been amended as follows:

--Reference is now made to Figs. 8A, 8B and 8C, which illustrate three alternative embodiments for inter-frame processor 48 which provides temporal filtering (i.e. inter-frame filtering) to further process the current frame. Since the present invention provides a full frame as output, inter-frame processor 48 determines which pixels have changed significantly from the previous frame and amends those only, storing the new version in the appropriate location in output frame memory 46.--



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/524,618	03/14/2000	Serijon Steraizin	P-2883-US	8670

7590 05/03/2002

Heidi M Brun
Eitan Pearl Latzer & Cohen-Zedek
One Crystal Park Suite 210
2011 Crystal Drive
Arlington, VA 22202-3709

EXAMINER

TRAN, PHUOC

ART UNIT

PAPER NUMBER

2521

#6

DATE MAILED: 05/03/2002

Please find below and/or attached an Office communication concerning this application or proceeding.



UNITED STATES DEPARTMENT OF COMMERCE
Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

APPLICATION NUMBER	FILING DATE	FIRST NAMED APPLICANT	ATTORNEY DOCKET NO.
--------------------	-------------	-----------------------	---------------------

EXAMINER

ART UNIT	PAPER NUMBER
----------	--------------

#6

DATE MAILED:

Response to Rule 312 Communication

- ☐ The petition filed _____ under 37 CFR 1.312(b) is granted. The paper has been forwarded to the examiner for consideration on the merits.

Director,
Patent Examining Group _____

- ☒ The amendment filed 3/18/02 under 37 CFR 1.312 has been considered, and has been:

- ☒ entered.
☐ entered as directed to matters of form not affecting the scope of the invention (Order 3311).
☐ disapproved. See explanation below.
☐ entered in part. See explanation below.

Melissa Roberts
Publishing Division



OK to Enter

PATENT
P-2883-US

MH
53-00
#7

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT(S): SHERAIZIN, Semion et al.

SERIAL NO: 09/524,618

EXAMINER:

TRAN, Phuoc

FILED:

March 14, 2000

GROUP ART UNIT: 2621

FOR:

METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE
SYNTACTIC ENCODING

ASSISTANT COMMISSIONER OF PATENTS
AND TRADEMARKS
WASHINGTON, DC 20231

RECEIVED

MAR 19 2002

Technology Center 2600


CLAIM FOR PRIORITY

Sir:

Applicant(s) hereby claims priority under 35 U.S.C. Section 119 based on Israeli Application No. 134182 filed January 23, 2000.

A certified copy of the priority document is submitted herewith.

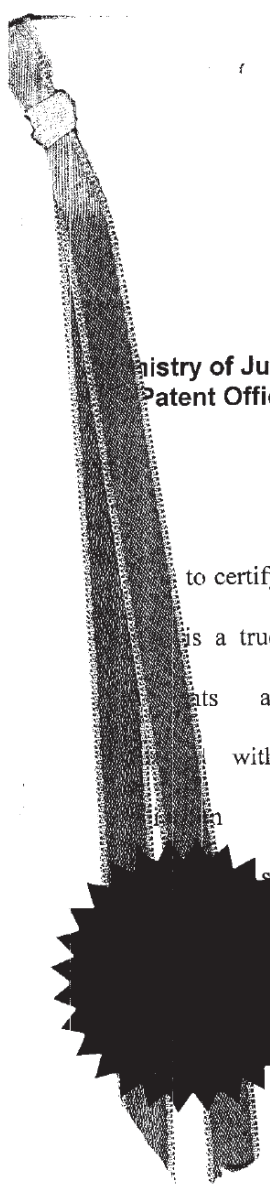
Respectfully submitted,


Caleb Pollack
Attorney for Applicants
Registration No. 37,912

Dated: March 12, 2002

Eitan, Pearl, Latzer & Cohen-Zedek
One Crystal Park, Suite 210, 2011 Crystal Drive
Arlington, VA 22202-3709
Tel: 703.486.0600
Fax: 703.486.0800

12.01



מדינת ישראל
STATE OF ISRAEL

Ministry of Justice
Patent Office

משרד המשפטים
לשכת הפטנטים

to certify that annexed
is a true copy of the
as originally
with the patent
of which
specified on the
nex.

זאת לתעודה כי רצופים
בוזז העתקים נכונים של
המסמכים שהופקדו
לכתחילה עם הבקשה
לפטנט לפי הפרטים
הרשומים בעמוד הראשון
של הנספח.

RECEIVED

MAR 19 2002

Technology Center 2600

This 31-01-2002 היום

רשם הפטנטים

Commissioner of Patents

נתאשר
Certified

לשימוש הלשכה
For Office Use

מספר: Number	134182
תאריך: Date	23-01-2000
הוקדם/נדחה: Ante/Post-dated	

VLS COM LTD.
4 Pikeris St. (Lev Hatamar Building)
Rehovot, 76545
Israel

Israeli Company

Inventor: Semion Sheraizin, Vitaly Sheraizin

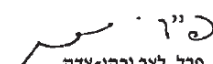
בעל אמצאה מכח המחאה זכויות ששמה הוא
Owner, by virtue of Assignment of an invention the title of which is

שיטה והתקן לדחיסה מקדימה ללא איבוד על ידי העין

(בעברית)
(Hebrew)

METHOD AND APPARATUS FOR VISUAL LOSSLESS PRE-PROCESSING

(באנגלית)
(English)

*בקשת חלוקה - Application of Division		*בקשת פטנט מוסף - Application for Patent Addition		דרישה דין קדימה Priority Claim	
*מבקשת פטנט from Application		*לבקשה/לפטנט to Patent/Appl.		מספר/סימן Number/Mark	תאריך Date
No. מס'		No. מס'			
dated מיום		dated מיום			
יפוי-כח: כללי / עוד יוגש POA: general/to be filed later					
הוגש בענין filed in case					
המען למסירת מסמכים בישראל Address for Service in Israel					
איתן, פרל, לצר וכהן-צדק עורכי דין, עורכי פטנטים ונוטריון רח' שנקר 7, הרצליה 46725					
חתימת המבקש Signature of the Applicant					
עבור המבקש, , For the Applicant,					
 איתן, פרל, לצר וכהן-צדק EITAN, PEARL, LATZER & COHEN-ZEDEK P-2883-IL					
				לשימוש הלשכה For Official Use	

נופס זה, כשהוא מוסבע בהותם לשכת הפטנטים ומושלם במספר ובתאריך ההגשה, הנו אישור להגשת הבקשה שפרטית רשומים לעיל.
This form, impressed with the Seal of the Patent Office and indicating the number and date of filing, certifies the filing of the application the particulars of which are set out above.

* מחק את המיותר Delete whatever is inapplicable

**METHOD AND APPARATUS FOR VISUAL LOSSLESS
PRE-PROCESSING**

שיטה והתקן לדחיסה מקדימה ללא איבוד על ידי העין

Eitan, Pearl, Latzer & Cohen-Zedek

P-2883-IL

FIELD OF THE INVENTION

The present invention relates generally to processing of video images and, in particular, to pre-processing of images for later compression by
5 standard compression techniques.

BACKGROUND OF THE INVENTION

There are many types of video signals, such as broadcast television (TV), digital, video conferencing, interactive TV, etc. All of these signals, in their digital form, are divided into frames, each of which consists of many pixels
10 (image elements), each of which requires 8 – 24 bytes to describe them. The result is megabytes of data per frame.

Before storing and / or transmitting these signals, they typically are compressed, using one of many standard video compression techniques, such as JPEG, MPEG, H-compression, etc. These compression standards use
15 video signal transforms and intra- and inter-frame coding that exploits spatial and temporal correlations among pixels of a frame and across frames.

However, these compression techniques create a number of well-known, undesirable and unacceptable artifacts, such as blockiness, low resolution and wiggles, among others. These are particularly problematic for
20 broadcast TV (satellite TV, cable TV, etc.) or for systems with very low bit rate (video conferencing, video phone).

Much research has been performed to try and improve the standard compression techniques. The following articles and patents discuss various prior art methods to do so:

1. Raj Malluri et al, "A Robust, Scalable, Object-Based Video
Compression Technique for Very Low Bit-Rate Coding," IEEE Transactions of
Circuit and Systems for Video Technology, vol. 7, No. 1, Feb. 1997;

AwadKh, Al-Asmari, "An Adaptive Hybrid Coding Scheme for HDTV
5 and Digital Sequences," IEEE Transactions on Consumer Electronics, vol. 42,
No. 3, pp. 926-936, Aug. 1995;

Kwok-tung Lo and Jian Feng, "Predictive Mean Search Algorithms for
Fast VQ Encoding of Images," IEEE Transactions On Consumer Electronics,
vol. 41, No. 2, pp. 327-331, May 1995;

10 James Goel et al. "Pre-processing for MPEG Compression Using
Adaptive Spatial Filtering", IEEE Transactions On Consumer Electronics, " vol.
41, No. 3, pp. 687-698, Aug. 1995;

Jian Feng et al. "Motion Adaptive Classified Vector Quantization for
ATM Video Coding", IEEE Transactions on Consumer Electronics, vol. 41, No.
15 2, p. 322-326, May 1995;

Austin Y. Lan et al., "Scene-Context Dependent Reference – Frame
Placement for MPEG Video Coding," IEEE Transactions On Circuits and
Systems for Video Technology, vol. 9, No. 3, pp. 478-489, Apr. 1999;

Kuo-Chin Fan, Kou-Sou Kan, "An Active Scene Analysis-Based
20 approach for Pseudoconstant Bit-Rate Video Coding", IEEE Transactions On
Circuits and Systems for Video Technology, vol. 8 No. 2, pp. 159-170, Apr.
1998;

Takashi Ida and Yoko Sambansugi, "Image Segmentation and Contour
Detection Using Fractal Coding", IEEE Transactions On Circuits and Systems
25 for Video Technology, vol. 8, No. 8, pp. 968-975, Dec. 1998;

· Liang Shen and Rangaraj M. Rangayyan, "A Segmentation-Based Lossless Image Coding Method for High-Resolution Medical Image Compression," IEEE Transactions On Medical Imaging, vol. 16, No. 3, pp. 301-316, June 1997;

5 Adrian Munteanu et al., "Wavelet-Based Lossless Compression of Coronary Angiographic Images", IEEE Transactions On Medical Imaging, vol. 18, No. 3, p. 272-281, March 1999; and

 Akira Okumura et al., "Signal Analysis and Compression Performance Evaluation of Pathological Microscopic Images," IEEE Transactions on Medical
10 Imaging, vol. 16, No. 6, pp. 701-710, Dec. 1997,

SUMMARY OF THE INVENTION

An object of the present invention is to provide a method and apparatus for vision compression which is generally lossless vis-a-vis what the human eye perceives.

5 There is therefore provided, in accordance with a preferred embodiment of the present invention, a visual lossless pre-processor for processing a video frame prior to compression by a video encoder. The pre-processor includes a unit which identifies a plurality of threshold levels, one per pixel of the video frame, above which a human eye can distinguish the pixel
10 from among the other pixels of the frame, a unit which associates the pixels of the video frame into subclasses and a unit which alters the intensity of each pixel of the video frame according to subclass and to whether or not the pixel is distinguished.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be understood and appreciated more fully from the following detailed description taken in conjunction with the appended drawings in which:

5 Fig. 1 is an example of a video frame;

Fig. 2 is a block diagram illustration of a video compression system having a visual lossless pre-processor, constructed and operative in accordance with a preferred embodiment of the present invention;

10 Fig. 3 is a block diagram illustration of the details of the visual lossless pre-processor of Fig. 2;

Fig. 4 is a graphical illustration of the transfer functions for a number of high pass filters useful in the pre-processor of Fig. 3;

Figs. 5A and 5B are block diagram illustrations of alternative embodiments of a filter bank forming part of the pre-processor of Fig. 3;

15 Fig. 6 is a graphical illustration of the transfer functions for a number of low pass filters useful in the filter bank of Figs. 5A and 5B;

Fig. 7 is a graphical illustration of the transfer function for a non-linear filter useful in the filter bank of Figs. 5A and 5B;

20 Figs. 8A, 8B and 8C are block diagram illustrations of alternative embodiments of an inter-frame processor forming part of the pre-processor of Fig. 3;

Fig. 9 is a block diagram illustration of a spatial-temporal analyzer forming part of the pre-processor of Fig. 3;

Figs. 10A and 10B are detail illustrations of the analyzer of Fig. 9; and

Fig. 11 is a detail illustration of a frame analyzer forming part of the pre-processor of Fig. 3.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

Applicants have realized that there are different levels of texture detail in a image and that the human eye perceives these details in different ways. In particular, Applicants have realized the following:

1. Texture details whose detection mainly depends on the level of noise in the image occupy approximately 50 – 80% of a image.
2. A visual perception threshold T for texture details does not depend on the shape of the object in the image.
3. Visual perception threshold T depends on a number of things, including the general brightness of the image. It does not depend on the noise spectrum.

Using the above facts, the present invention is a method for describing, and then encoding, images based on which objects in the image can be clearly perceived by the human eye and which ones can only be detected by it.

Reference is now made to Fig. 1, which is a grey-scale image of a plurality of shapes of a bird in flight, ranging from a photograph of one (labeled 10) to a very stylized version of one (labeled 12). The background of the image is a vignette which is very dark at the top of the image and very light at the bottom of the image.

The human eye can distinguish most of the birds of the image. However, there is at least one bird, labeled 14, which the eye can detect but cannot determine all of its details. Furthermore, there are large swaths of the image (in the background) which have no objects in them.

The present invention is a method and system for pre-processing images before they are sent to a standard video compression unit. The present

invention separates the details of an image into two different types, those that can only be detected (for which only one bit will suffice to describe each of their pixels) and those which can be distinguished (for which at least three bits are needed to describe each of their pixels).

5 Reference is now made to Fig. 2, which illustrates the present invention within an image transmission system. Thus, Fig. 2 shows a visual lossless (VLS) pre-processor 20 connected to a standard image transmitter 22 which includes a video compression encoder 24, such as a standard MPEG encoder, and a modulator 26. VLS pre-processor 20 preprocesses an incoming video
10 signal such that video compression encoder 24 can significantly compress the video signal, resulting in a reduced size signal to be transmitted. Typical compression ratios for the combined VLS pre-processor / video compression encoder are 60 – 100 which is generally twice that of the video compression encoder by itself.

15 Modulator 26 modulates the reduced size signal and transmits it to a receiver 30, which, as in the prior art, includes a demodulator 32 and a decoder 34. Demodulator 32 demodulates the transmitted signal and decoder 34 decodes and decompresses the demodulated signal. The result is provided to a monitor 36 for display.

20 It will be appreciated that, although the compression rates are high in the present invention, the resultant video displayed on monitor 36 is not significantly degraded. This is because pre-processor 20 attempts to quantify each frame of the video signal according to which sections of the frame are more or less distinguished by the human eye. For the less-distinguished
25 sections, pre-processor 20 provides pixels of a minimum bit size, thus reducing

the overall size (in bits) of the frame. Since the human eye does not distinguish these sections, the reproduced frame does not appear significantly different than the original frame, despite its smaller size (in bits).

Reference is now made to Fig. 3, which details the elements of VLS pre-processor 20. Pre-processor 20 comprises a memory 40, a frame analyzer 42, a processor 44, a frame memory 46 and a frame memory updater 48. Analyzer 42 analyzes each frame to separate it into subclasses, where subclasses define areas which are distinguished and those which are not as well as areas which describe objects of large size and which of small size. Processor 44 spatially filters each pixel of the frame according to its subclass and, optionally, also provides each pixel of the frame with the appropriate number of bits. Interframe updater 48 provides temporal filtering (i.e. inter-frame filtering) and updates frame memory 46 with the elements of the current frame which are different than those of the previous frame.

It is noted that frames are composed of pixels, each having three color components, each of which is typically defined by eight bits. VLS pre-processor 20 generally separately processes the three components. The following discusses the operation on the luminance signal Y , it being appreciated that the operation for the chrominance signals C_r and C_b is similar. However, the bandwidth of the chrominance signals is half as wide as that of the luminance signal. Thus, the filters (in the x direction of the frame) for chrominance have a narrower bandwidth. The following discussion shows the filters for the luminance signal.

Frame analyzer 42 comprises a spatial-temporal analyzer 50, a parameter estimator 52, a visual perception threshold determiner 54 and a

subclass determiner 56. Details of these elements are provided in Figs. 9 – 11, discussed hereinbelow.

As discussed hereinabove, areas which the human eye distinguishes are ones of high contrast or high information content. Areas of high contrast are areas with a lot of high frequency content. Thus, spatial-temporal analyzer 50 generates a plurality of filtered frames from the current frame, each filtered through a different high pass filter (HPF), where each high pass filter retains a different range of frequencies therein.

Fig. 4, to which reference is now briefly made, is an amplitude vs. frequency graph illustrating the transfer functions of an exemplary set of high pass filters. Four graphs are shown. It can be seen that the curve labeled HPF-R3 has a cutoff frequency of 1MHz and thus, retains portions of the frame with information above 1MHz. Similarly, curve HPF-R2 has a cutoff frequency of 2 MHz, HPF-C2 has a cutoff frequency of 3MHz and HPF-R1 and HPF-C1 have a cutoff frequency of 4MHz. As will be discussed hereinbelow, the terminology "Rx" refers to operations on a row of pixels while the terminology "Cx" refers to operations on a column of pixels.

In particular, the filters of Fig. 4 implement the following finite impulse response (FIR) filters on either a row of pixels (the x direction of the frame) or a column of pixels (the y direction of the frame), where the number of pixels used in the filter defines the power of the cosine. For example, a filter implementing $\cos^{10}x$ takes 10 pixels around the pixel of interest, five to one side and five to the other side of the pixel of interest.

$$\text{HPF-R3: } 1 - \cos^{10}x$$

$$\text{HPF-R2: } 1 - \cos^5x$$

HPF-R1: $1 - \cos^2 x$

HPF-C2: $1 - \cos^4 y$

HPF-C1: $1 - \cos^2 y$

The high pass filters can also be considered as digital equivalents of optical apertures. The higher the frequency range, the smaller the aperture. Thus, filters HPF-R1 and HPF-C1 retain only very small details in the frame (of 1 – 4 pixels in size) while filter HPF-R3 retains much larger details (of up to 10 pixels).

In the following, the filtered frames will be labeled by the type of filter (HPF-X) used to create them.

Returning to Fig. 3, analyzer 50 also generates difference frames between the current frame and another, earlier frame. The previous frame is typically at most 15 frames earlier. A "group" of frames is a series of frames for which difference frames are generated.

Parameter estimator 52 takes the current frame and the filtered and difference frames and generates a set of parameters that describe the information content of the current frame. The parameters are determined on a pixel-by-pixel basis or on a per frame basis, as relevant. It is noted that the parameters do not have to be calculated to great accuracy as they are used in combination to determine a per pixel, visual perception threshold THD_i .

At least some of the following parameters are determined:

Signal to noise ratio (SNR): this parameter can be determined by generating a difference frame between the current frame and the frame before it, filtering out the high frequency content of the difference frame and summing the intensities of the pixels in the filtered frame.

Normalized amount of change $N\Delta_i$: this measures the amount of change Δ_i per pixel i , from the current frame to its previous frame, normalized by the maximum intensity I_{MAX} possible for the pixel. If the frame is a television frame, the maximum intensity is 255.

5 Normalized volume of change NI_{XY} : this measures the volume of change in a frame I_{XY} , normalized by the maximum possible amount of information MAX_{INFO} within a frame (i.e. 8 bits per pixel x N pixels per frame). Since the highest frequency range indicates the amount of change in a frame, the volume of change I_{XY} is a sum of the intensities in the filtered frame having
10 the highest frequency range, such as filtered frame HPF-R1.

Normalized volume of interframe changes NI_F : this measures the volume of changes I_F between the current frame and its previous frame, normalized by the maximum possible amount of information MAX_{INFO} within a frame. The volume of interframe changes I_F is the sum of the intensities in the
15 difference frame.

Normalized volume of change within a group of frames NI_{GOP} : this measures the volume of changes I_{GOP} over a group of frames, where the group is from 2 to 15 frames, as selected by the user. It is normalized by the maximum possible amount of information MAX_{INFO} within a frame and by the
20 number of frames in the group.

Normalized luminance level NY_i : Y_i is the luminance level of a pixel in the current frame. It is normalized by the maximum intensity I_{MAX} possible for the pixel.

Color saturation p_i : this is the color saturation level of the i th pixel and it is determined by: $\left[0.78 \left(\frac{C_{r,i} - 128}{160} \right)^2 + 0.24 \left(\frac{C_{b,i} - 128}{126} \right)^2 \right]^{1/2}$ where $C_{r,i}$ is the amount of red in the i th pixel and $C_{b,i}$ is the amount of blue in the i th pixel.

Hue h_i : this is the general hue of the i th pixel and is determined by: $\arctan \left(1.4 \frac{C_{r,i} - 128}{C_{b,i} - 128} \right)$.

Visual perception threshold determiner 54 determines the visual perception threshold THD_i as follows:

$$THD_i = THD_{min} \left(1 + N\Delta_i + NI_{xy} + NI_F + NI_{GDP} + NY_i + p_i + h_i + \frac{200}{SNR} \right)$$

Subclass determiner 56 compares each pixel i of each high pass filtered frame HPF-X to its associated threshold THD_i to determine whether or not that pixel is significantly present in each filtered frame, where "significantly present" is defined by the threshold level and by the "detail dimension" (i.e. the size of the object or detail in the picture of which the pixel forms a part). Subclass determiner 56 then defines the subclass to which the pixel belongs.

For the example provided above, if the pixel is not present in any of the filtered frames, the pixel must belong to an object of large size or it. If the pixel is only found in the filtered frame of HPF-C2 or in both frames HPF-C1 and HPF-C2, it must be a horizontal edge (an edge in the Y direction of the frame). If it is found in filtered frames HPF-R3 and HPF-C2, it is a single small detail. If the pixel is found only in filtered frames HPF-R1, HPF-R2 and HPF-R3, it is a very small vertical edge. If, in addition, it is also found in filtered frame HPF-C2, then the pixel is a very small, single detail.

The above logic is summarized and expanded in Table 1.

Table 1

Subclass	High Pass Filters					Remarks
	R1	R2	R3	C1	C2	
1	0	0	0	0	0	Extended detail or detail detected only
2	0	0	0	0	1	Horizontal edge
3	0	0	0	1	1	Horizontal edge
4	0	0	1	0	0	Vertical edge
5	0	0	1	0	1	Single small detail
6	0	0	1	1	1	Single small detail
7	0	1	1	0	0	Vertical edge
8	0	1	1	0	1	Single small detail
9	0	1	1	1	1	Single small detail
10	1	1	1	0	0	Very small vertical edge
11	1	1	1	0	1	Very small single detail
12	1	1	1	1	1	Very small single detail

The output of subclass determiner 56 is an indication of the subclass to which each pixel of the current frame belongs. Processor 44 performs spatial filtering of the frame, where the type of filter utilized varies in accordance with the subclass to which the pixel belongs. Processor 44 changes the intensity of the pixel by an amount less than the visual distinguishing threshold for that pixel. Pixels whose contrast is lower than the threshold are transformed with the non-linear filters. If desired, the data size of the pixel can be reduced from 8 bits to 3 or 4 bits, depending on the visual threshold level and the detail dimension for the pixel.

Processor 44 comprises a filter bank 60 and a filter selector 62. Filter bank 60 comprises a set of low pass and non-linear filters, shown in Figs. 5A and 5B to which reference is now made, which filter selector 62 activates, based on the subclass to which the pixel belongs. Selector 62 can activate more than one filter, as necessary.

Figs. 5A and 5B are two, alternative embodiments of filter bank 60. Both comprise two sections 64 and 66 which operate on columns and rows, respectively. In each section 64 and 66, there is a choice of filters, each

controlled by an appropriate switch, labeled SW-X, where X is one of C1, C2, R1, R2, R3 (selecting one of the low pass filters (LPF)) or D-C, D-R (selecting to pass the relevant pixel directly). Filter selector 62 switches the relevant switch, thereby activating the relevant filter. It is noted that the non-linear filters (NLF) are activated together with low pass filter LPF-R3 and the outputs of the two filters are summed.

Filter bank 60 also includes time adjusters (TA) which add any necessary delays to ensure that the pixel currently being processed remains at its appropriate location within the frame.

10 The low pass filters (LPF) are associated with the high pass filters used in analyzer 50. Thus, the cutoff frequencies of the low pass filters are close to those of the high pass filters. The low pass filters thus pass that which their associated high pass filters ignore.

Fig 6, to which reference is now briefly made, illustrates exemplary low pass filters for the example provided hereinabove. Low pass filter LPF-R3 has a cutoff frequency of 0.5MHz and thus, generally does not retain anything which its associated high pass filter HPF-R3 (with a cutoff frequency of 1MHz) retains. Filter LPF-R2 has a cutoff frequency of 1MHz, filter LPF-C2 has a cutoff frequency of 1.25MHz and filters LPF-C1 and LPF-R1 have a cutoff frequency of about 2MHz. As for the high frequency filters, filters LPF-Cx operate on the columns of the frame and filters LPF-Rx operate on the rows of the frame.

Fig. 7, to which reference is now briefly made, illustrates an exemplary transfer function for the non-linear filters (NLF). The transfer function defines an output value V_{out} normalized by the threshold level THD_i as a function of an input value V_{in} also normalized by the threshold level THD_i . As can be seen in

the figure, the input – output relationship is described by a polynomial of high order. A typical order might be six, though lower orders, of power two or three, are also feasible.

Table 2 lists the type of filters activated per subclass, where the header for the column indicates both the type of filter and the label of the switch SW-X of Figs. 5A and 5B.

Table 2							
Subclass	Low Pass Filters						
	R1	R2	R3	C1	C2	D-R	D-C
1	0	0	1	0	1	0	0
2	0	0	1	1	0	0	0
3	0	0	1	0	0	0	1
4	0	1	0	0	1	0	0
5	0	1	0	1	0	0	0
6	0	1	0	0	0	0	1
7	1	0	0	0	1	0	0
8	1	0	0	1	0	0	0
9	1	0	0	0	0	0	1
10	0	0	0	0	1	1	0
11	0	0	0	1	0	1	0
12	0	0	0	0	0	1	1

Fig. 5B includes rounding elements RND which reduce the number of bits of a pixel from eight to three or four bits, depending on the subclass to which the pixel belongs. Table 3 illustrates the logic for the example presented hereinabove.

Table 3					
subclass	RND-P0 (Z1) "1" equal 8 bit "0" equal 3 bit	RND-P1 (Z2) "1" equal 8 bit "0" equal 4 bit	RND-P2 (Z3) "1" equal 8 bit "0" equal 4 bit	RND-L0 (Z4) "1" equal 8 bit "0" equal 3 bit	RND-L1 (Z5) "1" equal 8 bit "0" equal 4 bit
1	1	1	1	1	1
2	1	1	1	1	0
3	1	1	1	1	1
4	1	1	1	1	1
5	1	1	0	1	0
6	1	1	1	1	1
7	1	1	1	1	1
8	1	0	1	0	1
9	1	1	1	1	1
10	1	1	1	1	1
11	1	1	1	1	1
12	1	1	1	1	1

The output of processor 44 is a processed version of the current frame, typically having fewer bits than the original version.

Reference is now made to Figs. 8A, 8B and 8C, which illustrate three alternative embodiments for interframe updater 48 which provides temporal filtering (i.e. inter-frame filtering) to further process the current frame. Since the present invention provides a full frame as output and since frame memory 46 stores the previous frame, interframe updater 48 determines which pixels have changed significantly from the previous frame and amends those only, storing the new version in the appropriate location in frame memory 46.

10 The embodiments of Figs. 8A and 8B are open loop versions while the embodiment of Fig. 8C is a closed loop version. All of the embodiments comprise a summer 68, a low pass filter (LPF) 70, a high pass filter (HPF) 72, two comparators 74 and 76, two switches 78 and 80, controlled by the results of comparators 74 and 76, respectively, and a summer 82.

15 Summer 68 takes the difference of the processed current frame, produced by processor 44, and the previous frame, stored in frame memory 46. The difference frame is then processed in two parallel tracks.

In the first track, it is low pass filtered with either a $\cos^6 x$ filter (for broadcast frames) or a $\cos^{10} x$ filter (for video conferencing). Each pixel of the filtered frame is compared to a general threshold THD-LF which is typically set to 5% of the maximum expected intensity for the frame. Thus, the pixels which are kept are only those which changed by more than 5%.

In the second track, the difference frame is high pass filtered with either a $1-\cos^6 x$ filter (for broadcast frames) or a $1-\cos^{10} x$ filter (for video conferencing). Since high pass filtering retains the small details, each pixel of

the high pass filtered frame is compared to the particular threshold THD_i for that pixel, as produced by threshold determiner 54. If the difference pixel has an intensity above the threshold THD_i (i.e. the change in the pixel is significant), it is allowed through (i.e. switch 80 is set to pass the pixel).

5 Summer 82 adds the filtered difference pixels passed by switches 78 and / or 80 with the pixel of the previous frame to produce the new pixel. If switches 78 and 80 did not pass anything, the new pixel is the same as the previous pixel. Otherwise, the new pixel is the sum of the previous pixel and one or more difference pixels.

10 Reference is now briefly made to Figs. 9, 10A, 10B and 11 which detail elements of frame analyzer 42. In these figures, the term "ML" indicates a line of the current frame, "MP" indicates a pixel of the current frame, "MF" indicates a frame, "VD" indicates the vertical drive, "TA" indicates a time adjustment, e.g. a delay, and CNT indicates a counter.

15 Fig. 9 generally illustrates the operation of spatial-temporal analyzer 50 and Figs. 10A and 10B provide one detailed embodiment for the spatial analysis and temporal analysis portions 51 and 53, respectively. Fig. 11 details parameter estimator 52, threshold determiner 54 and subclass determiner 56. As these figures are deemed to be self-explanatory, no further explanation will
20 be included here.

The methods and apparatus disclosed herein have been described without reference to specific hardware or software. Rather, the methods and apparatus have been described in a manner sufficient to enable persons of ordinary skill in the art to readily adapt commercially available hardware and
25 software as may be needed to reduce any of the embodiments of the present

invention to practice without undue experimentation and using conventional techniques.

It will be appreciated by persons skilled in the art that the present invention is not limited by what has been particularly shown and described
5 herein above. Rather the scope of the invention is defined by the claims that follow:

CLAIMS

1. A visual lossless pre-processor for processing a video frame prior to compression by a video encoder, the pre-processor comprising:
 - 5 Means for identifying a plurality of threshold levels, one per pixel of said video frame, above which a human eye can distinguish said pixel from among the other pixels of said frame;
 - Means for associated said pixels of said video frame into subclasses; and
 - 10 Means for altering the intensity of each pixel of said video frame according to subclass and to whether or not said pixel is distinguished.
2. A pre-processor according to claim 1 substantially as described hereinabove.
- 15 3. A pre-processor according to claim 1 substantially as illustrated in any of the drawings.

For the Applicant,

20 
Eitan, Pearl, Latzer & Cohen-Zedek

Lawyers, Patent Attorneys & Notaries

P-2883-IL



14



12

robomatix

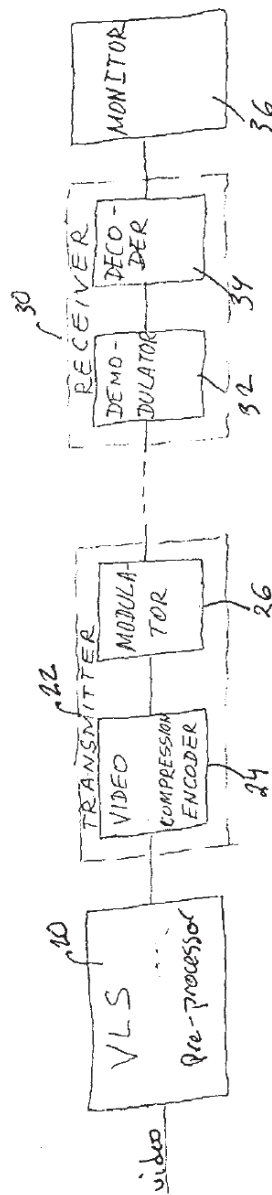
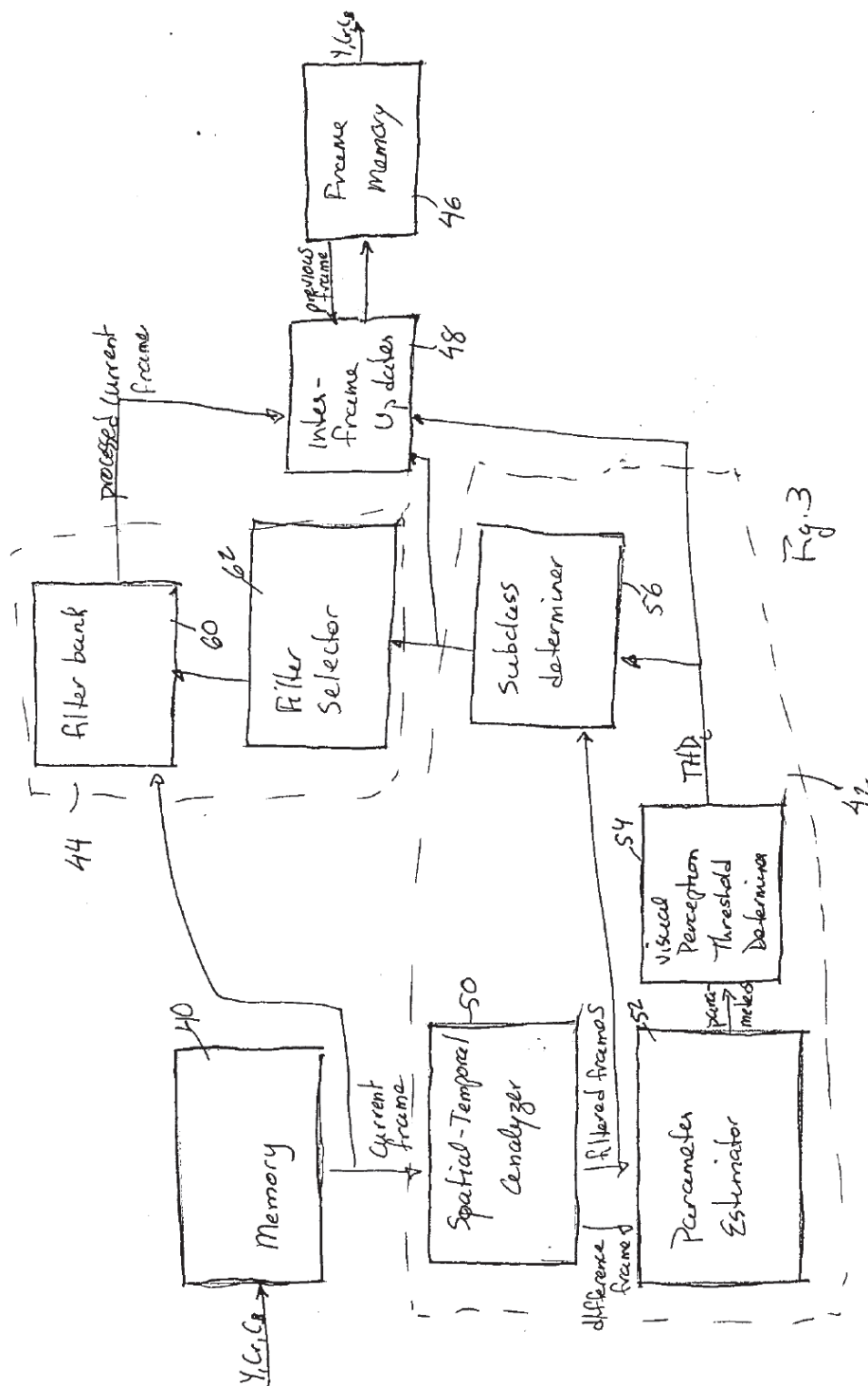
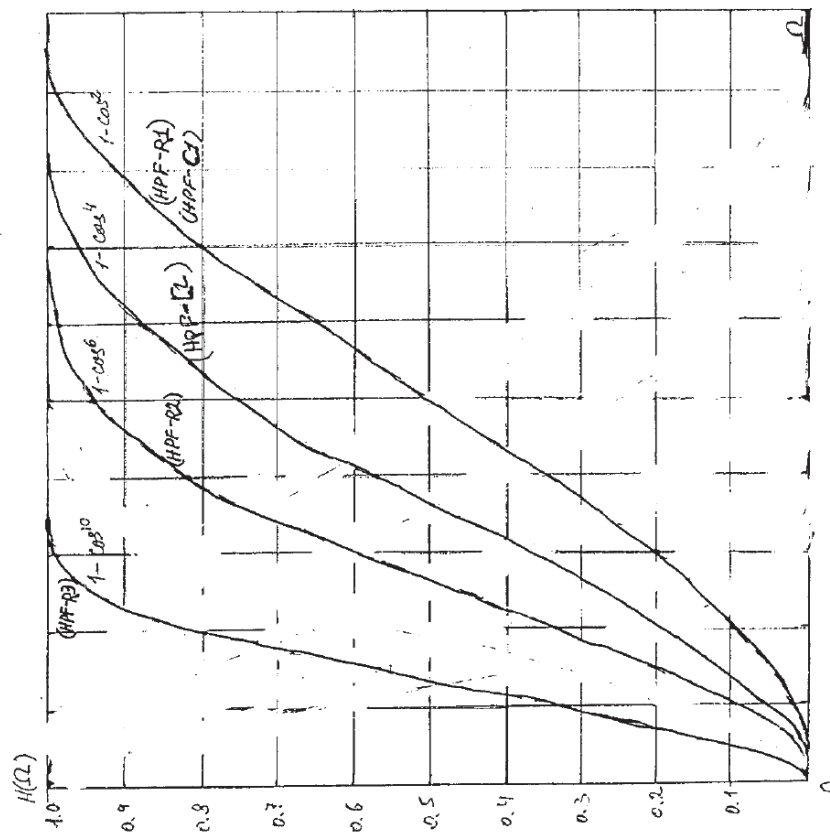


FIG. 2





5 MHz

1 MHz

5 MHz

1 MHz

Fig. 4

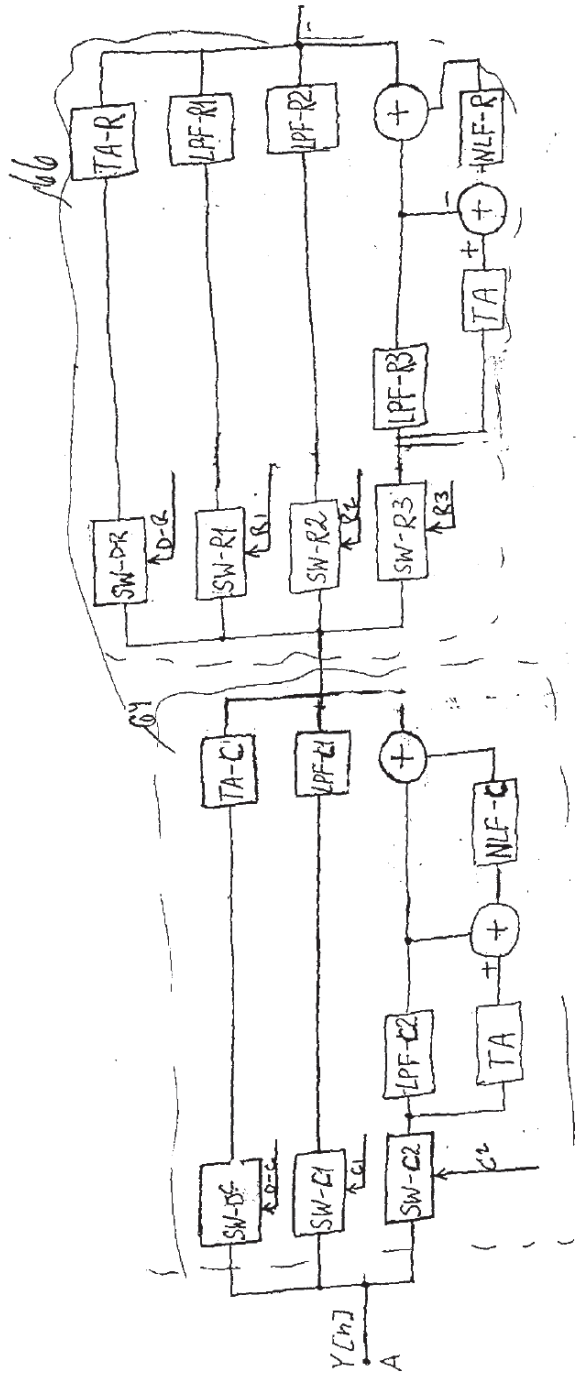


Fig. 5A

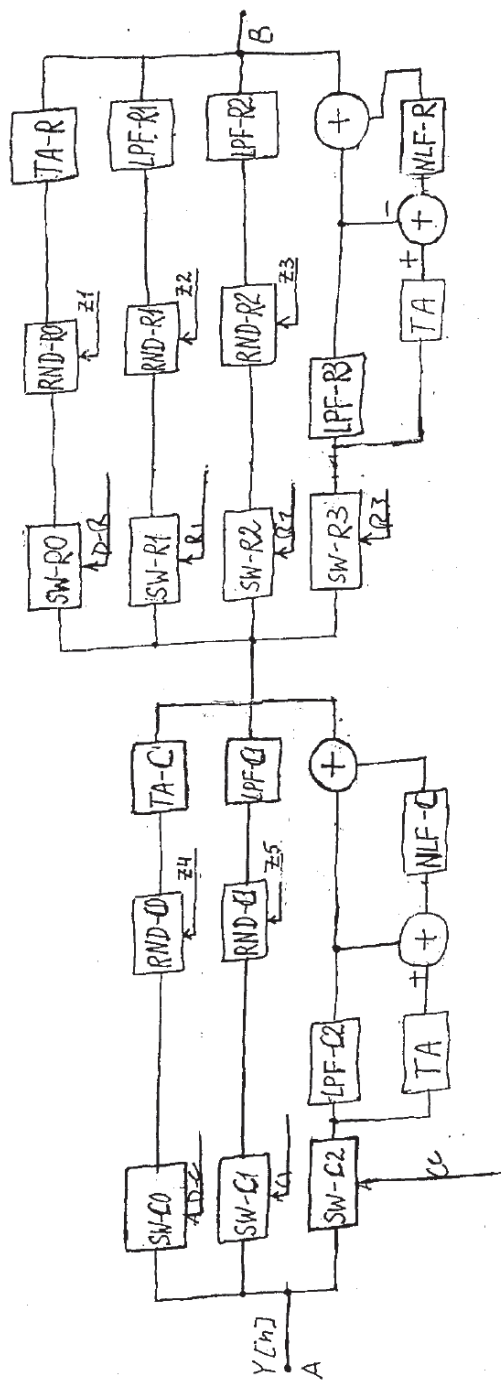


Fig. 5B

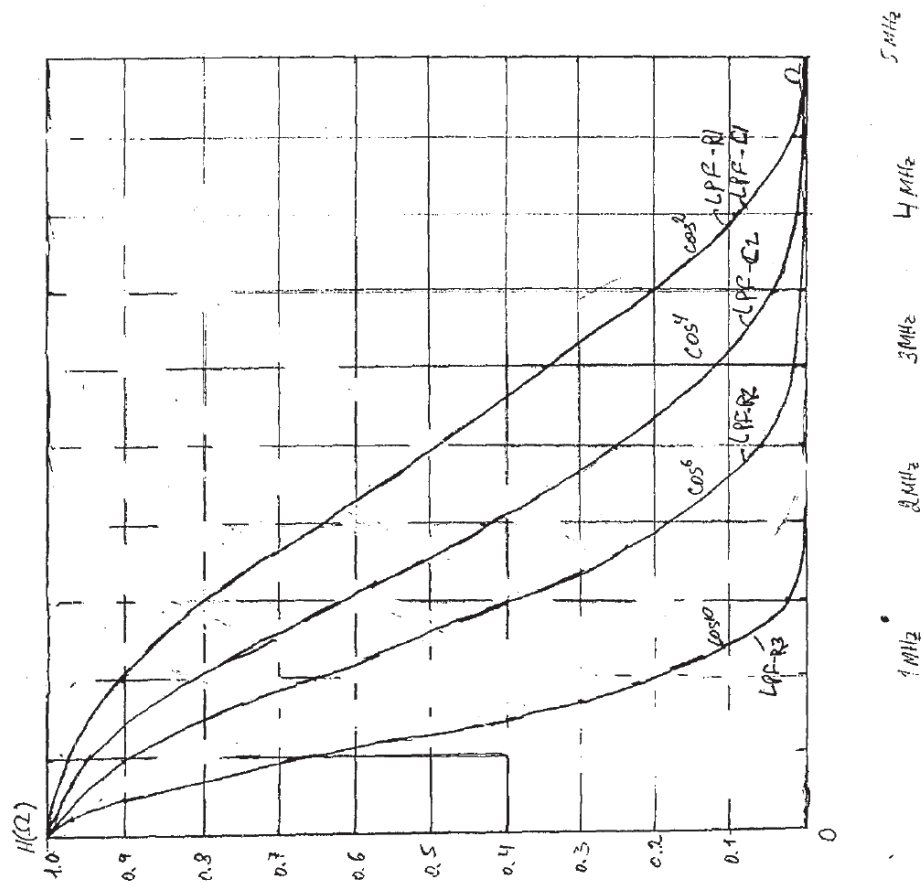


Fig. 6

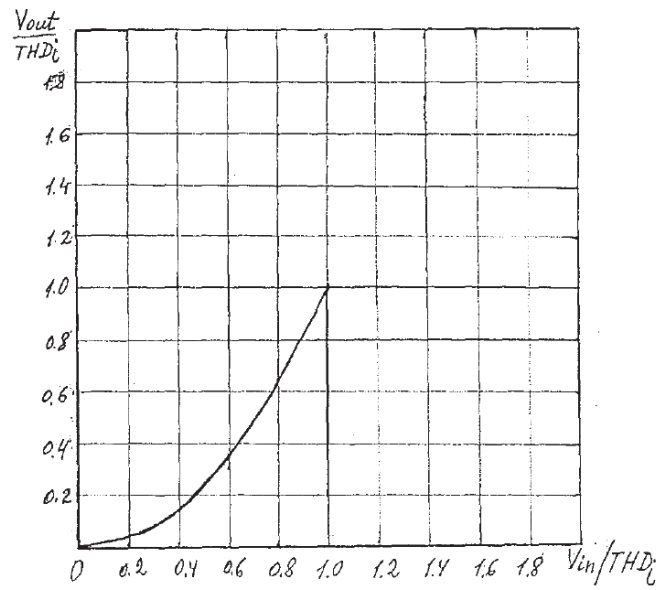
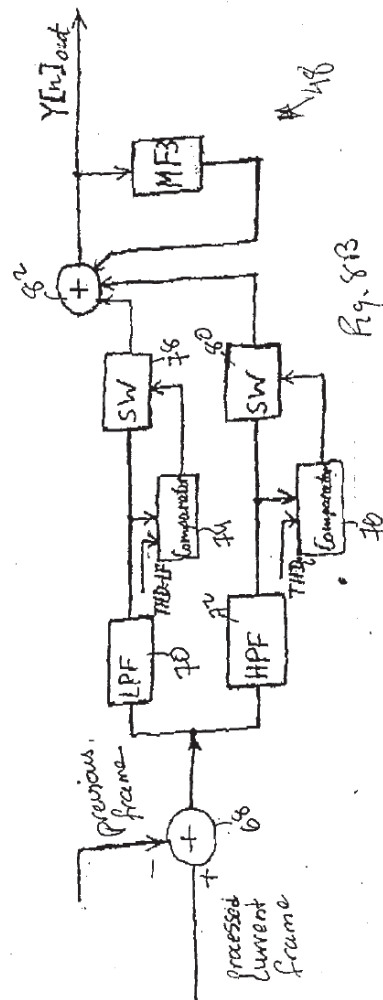
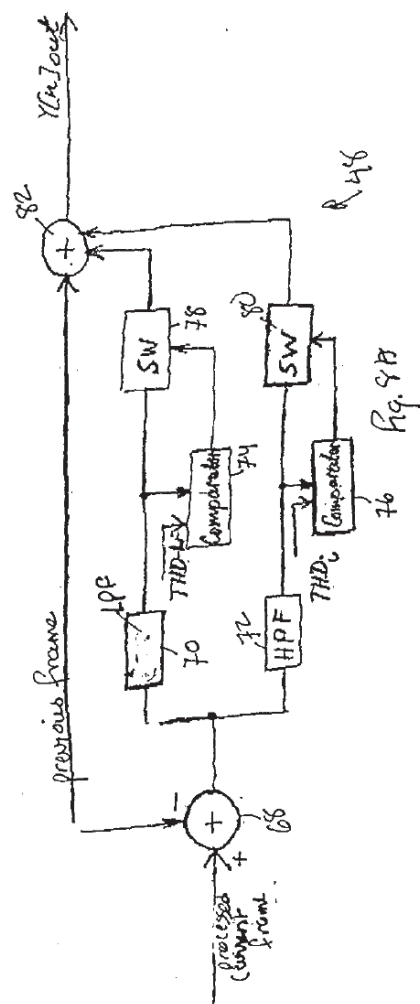


Fig. 7



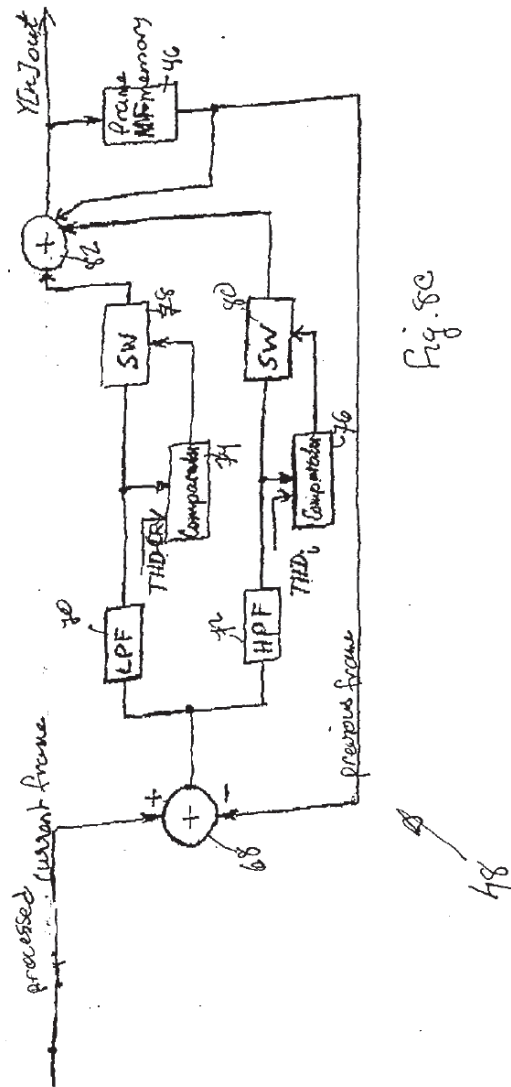


Fig. 8C

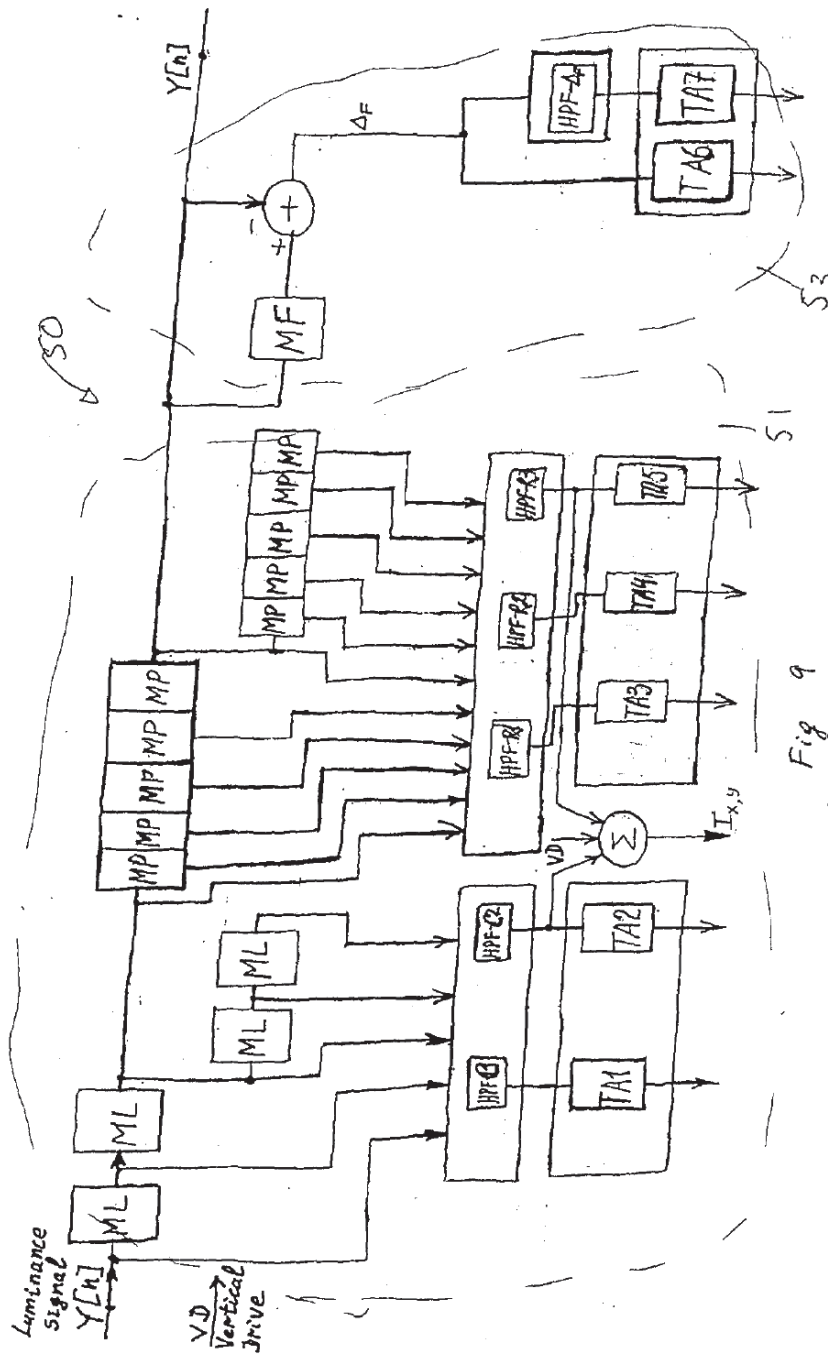


Fig. 9

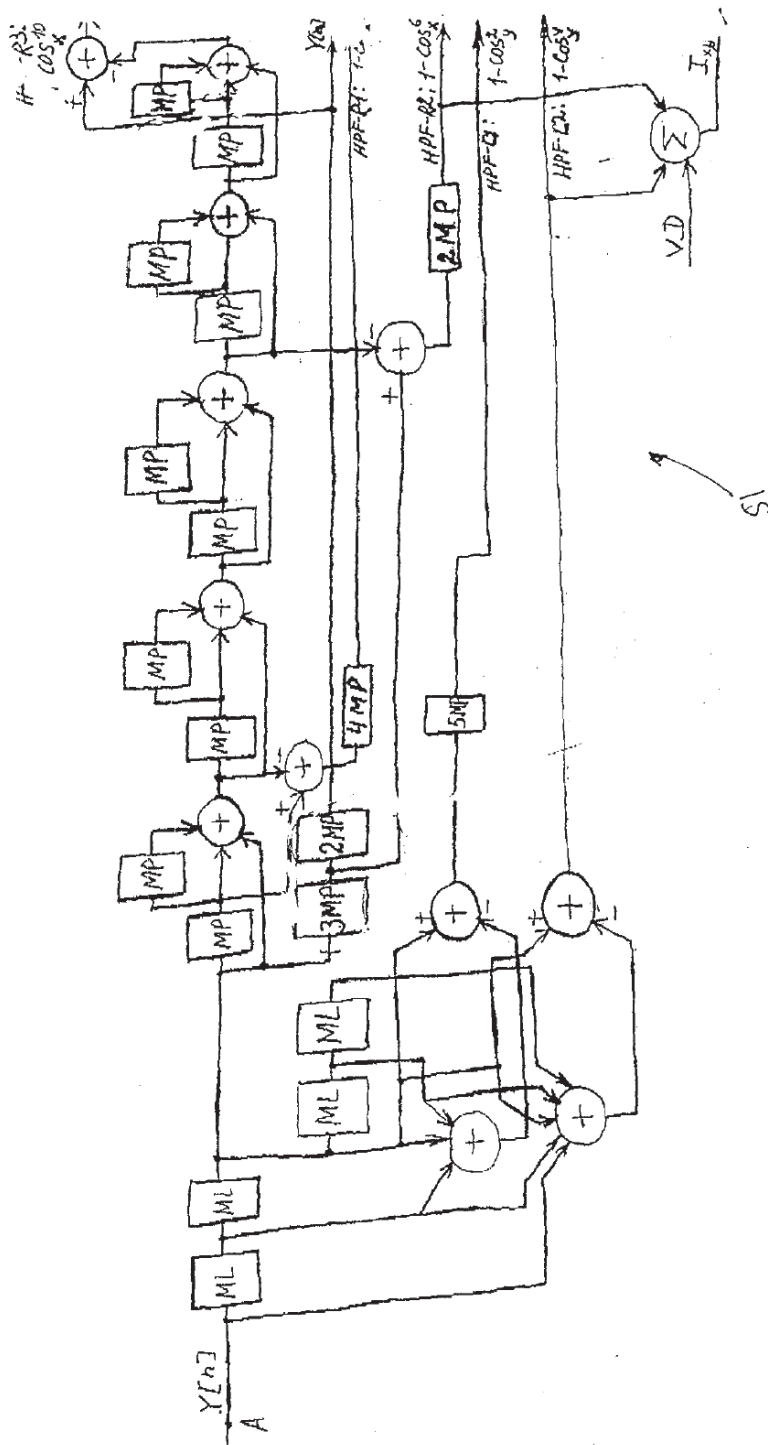


Fig. 10A

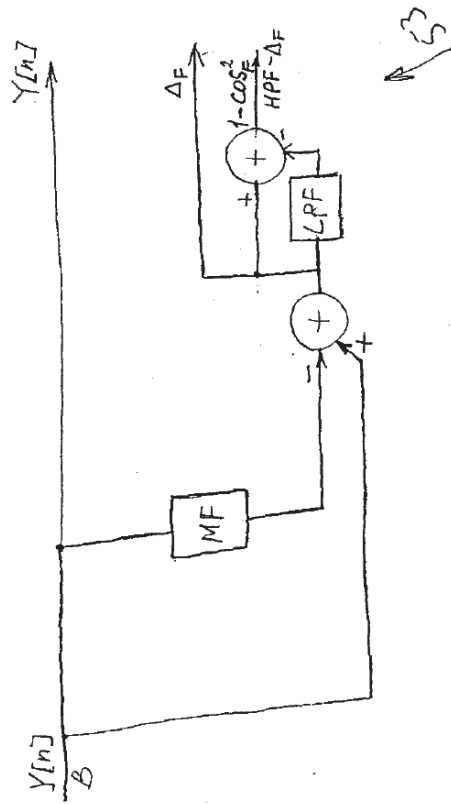


Fig. 10B

56

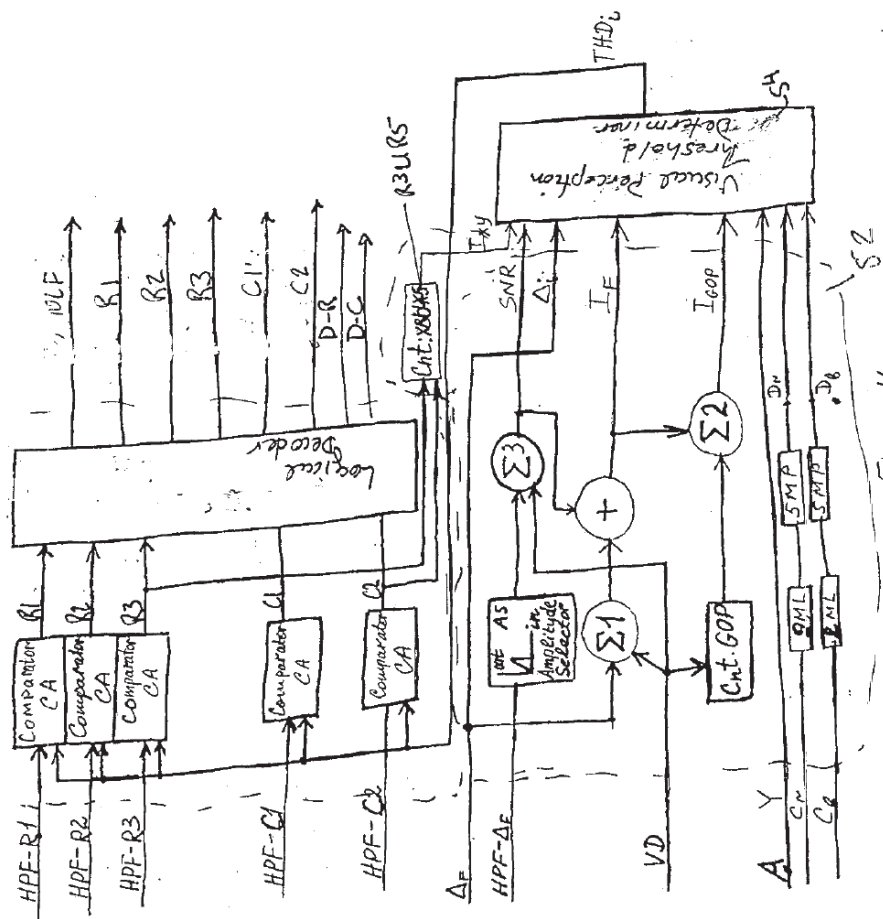


Fig. 11



MATCH & RETURN

Attorney Docket No: P-2883-US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: SHERAIZIN et al. Examiner: TRAN, P.
Serial No.: 09/524,618 Group Art Unit: 2621
Filed: March 14, 2000
Title: METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE
SYNTACTIC ENCODING

*approved
PT*

*#8
Diana
0-13-0*

ATTENTION: Official Draftsman
Commissioner of Patents and Trademarks
Washington, D.C. 20231

RECEIVED

MAR 21 2002

CORRECTED DRAWINGS

Technology Center 2600

Sir:

Enclosed please find red-lined corrected drawings.


In Fig. 5A, frames 64 and 66 were added.

In Fig. 8B and 8C, the title "FRAME MEMORY" of element 46 has been corrected to
"OUTPUT FRAME MEMORY".

Approval is requested.

If there are any questions regarding these drawings, please call the undersigned.

Respectfully submitted,


Caleb Pollack
Attorney for Applicant(s)
Registration No. 37,912

Dated: March 14, 2002

Eitan, Pearl, Latzer & Cohen-Zedek
One Crystal Park, Suite 210, 2011 Crystal Drive
Arlington, VA, USA 22202-3709
Telephone: (703) 486-0600
Fax: (703) 486-0800

12-01

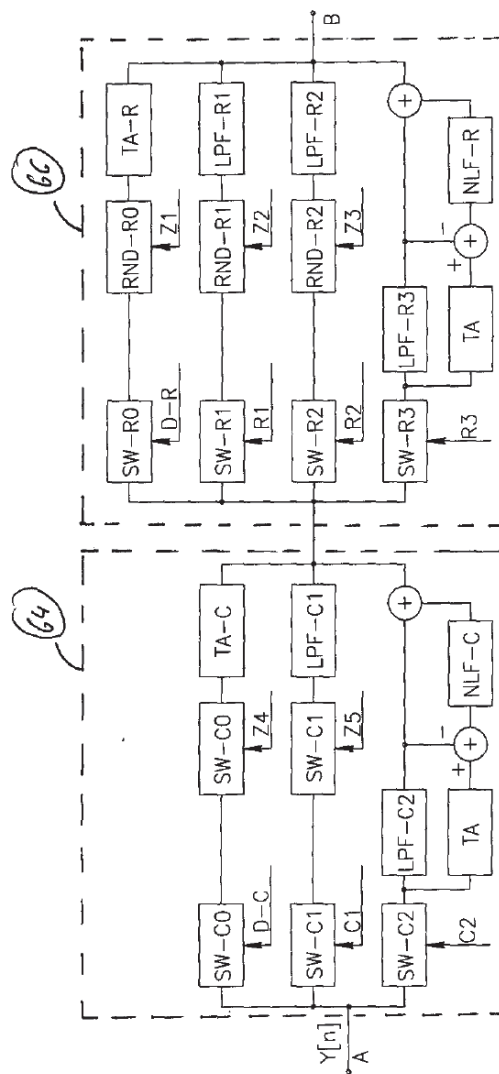


FIG. 5B

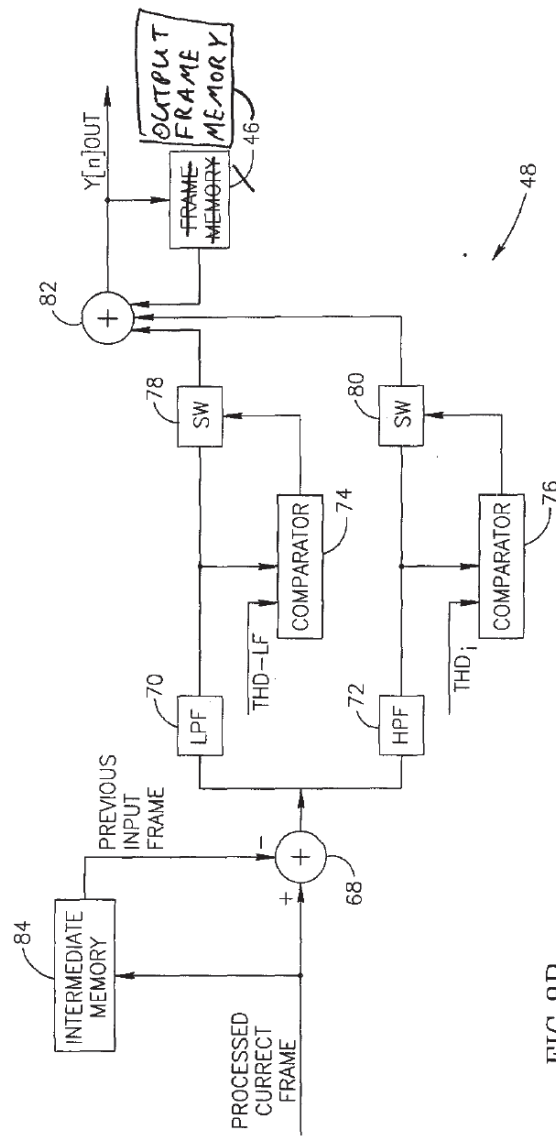


FIG.8B

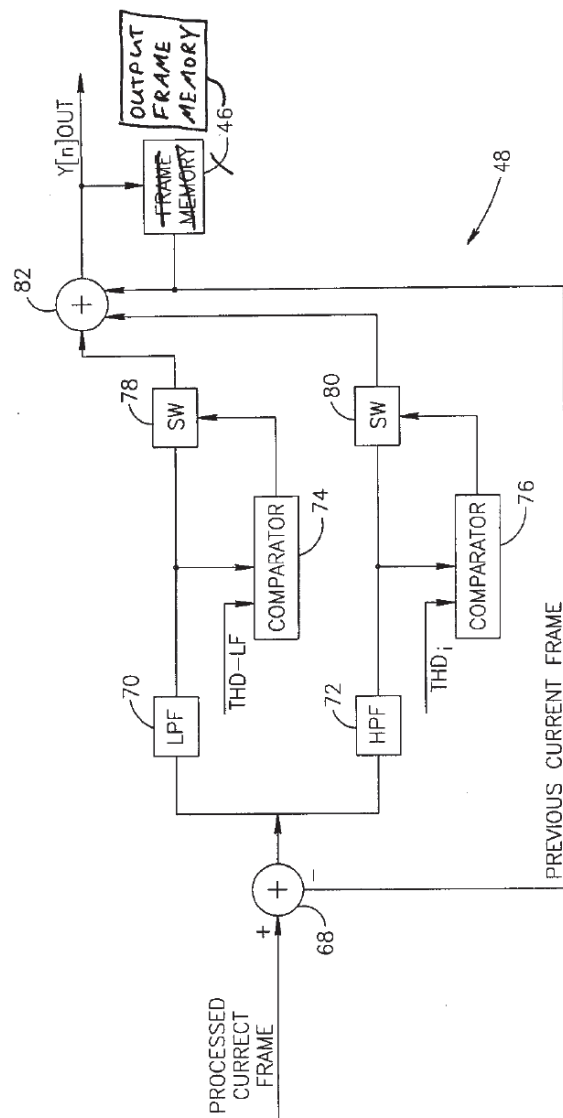


FIG. 8C



B

PATENT
P-2883-US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT(S): SHERAIZIN, Semion M. et al.

SERIAL NO: 09/524,618

Examiner:

TRAN, Phuoc

#9
DRIVER
0-13-00

FILED: March 14, 2000

GROUP ART UNIT: 2621

FOR: METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE
SYNTACTIC ENCODING

1022

ASSISTANT COMMISSIONER FOR PATENTS
WASHINGTON, DC 20231

BOX ISSUE FEE

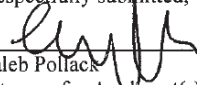
Attn: Official Draftsperson

SUBMISSION OF FORMAL DRAWINGS

Sir:

Applicant(s) submit(s) two sets of formal drawings, each set consisting of fifteen (15) pages in the above-identified Application. If there are any questions regarding these drawings, please call the undersigned attorney of record at (703) 486-0600.

Respectfully submitted,


Caleb Pollack
Attorney for Applicant(s)
Registration No. 37,912

Dated: March 14, 2002

Eitan, Pearl, Latzer & Cohen-Zedek
One Crystal Park, Suite 210
2011 Crystal Drive
Arlington, VA 22202-3709
U.S.A.

Tel: (703) 486-0600

Fax: (703) 486-0800

12/10/

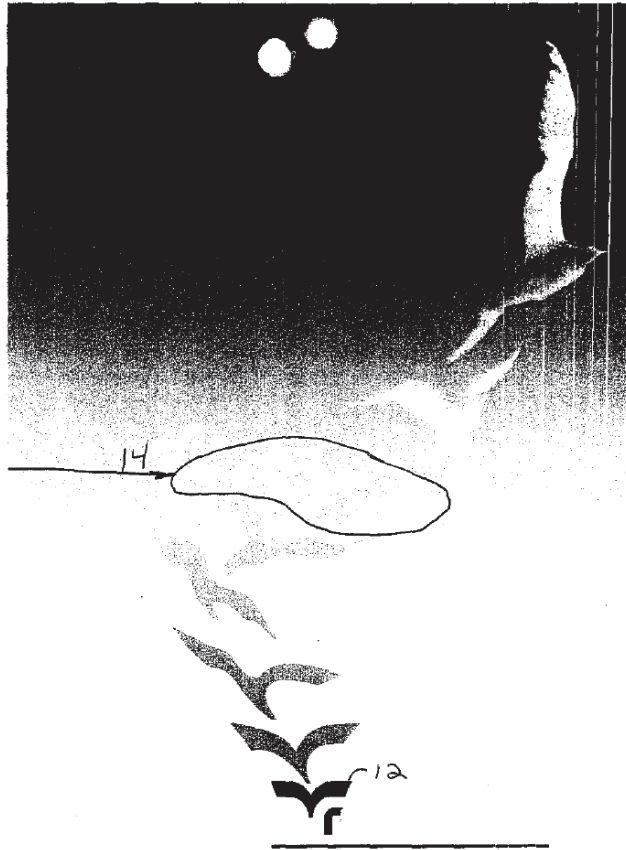


Fig 1

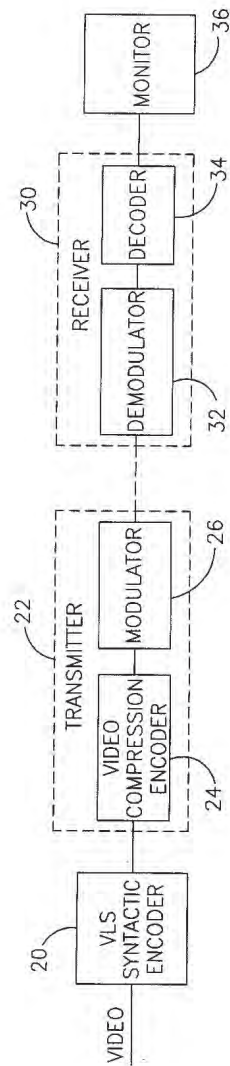


FIG. 2

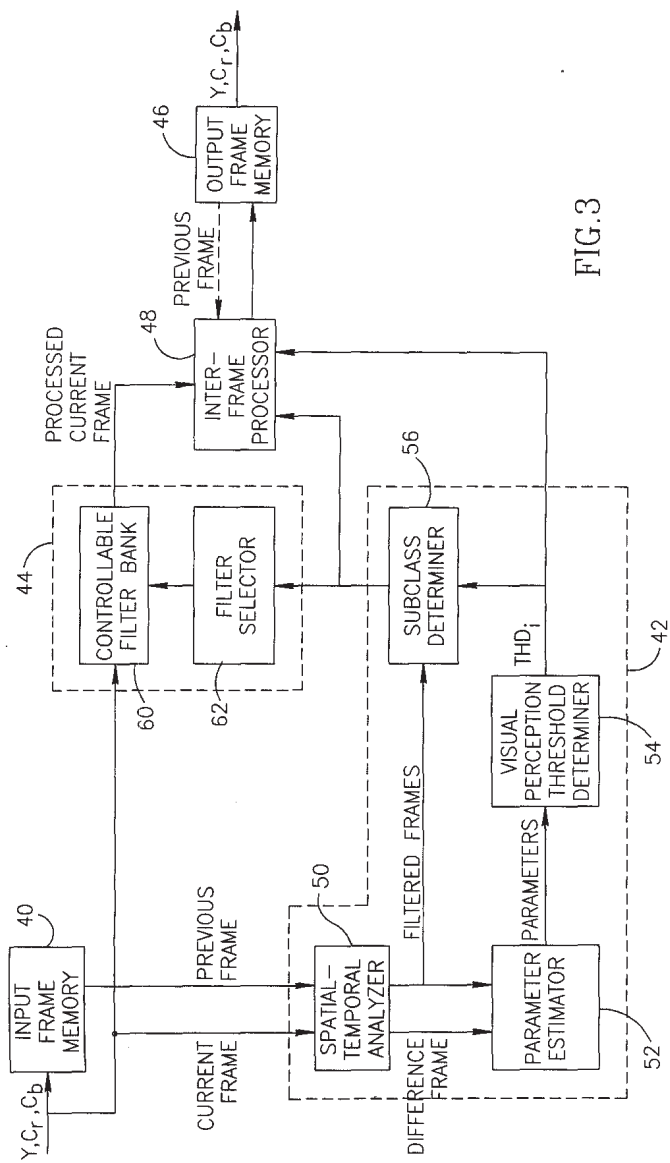


FIG. 3

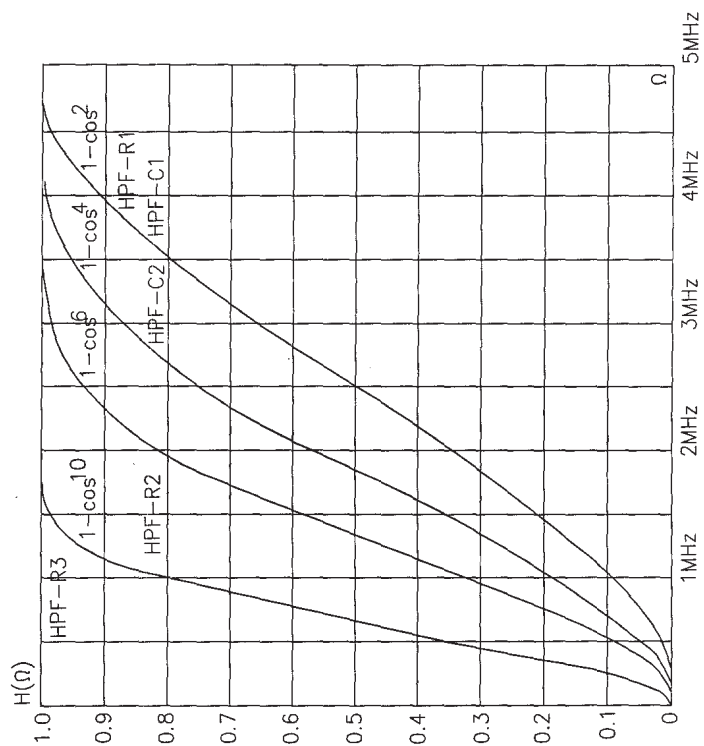


FIG. 4

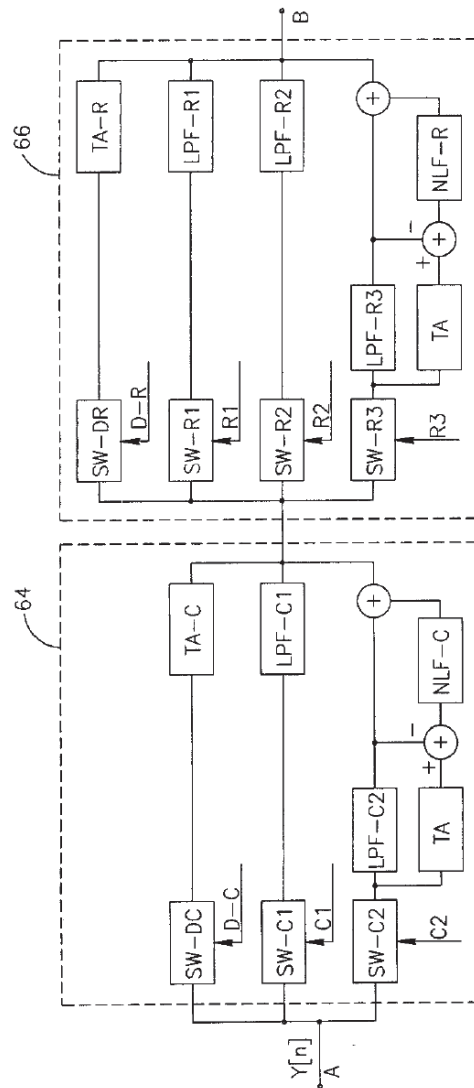


FIG. 5A

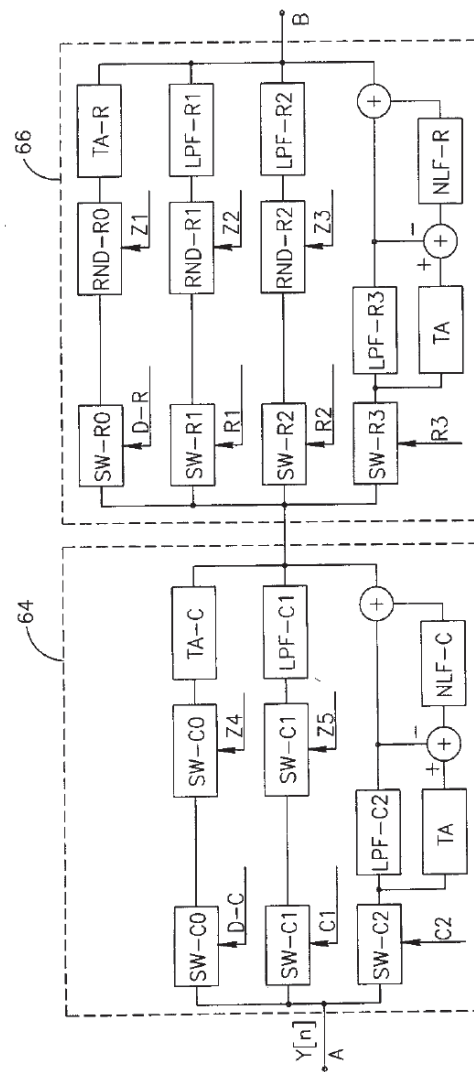


FIG. 5B

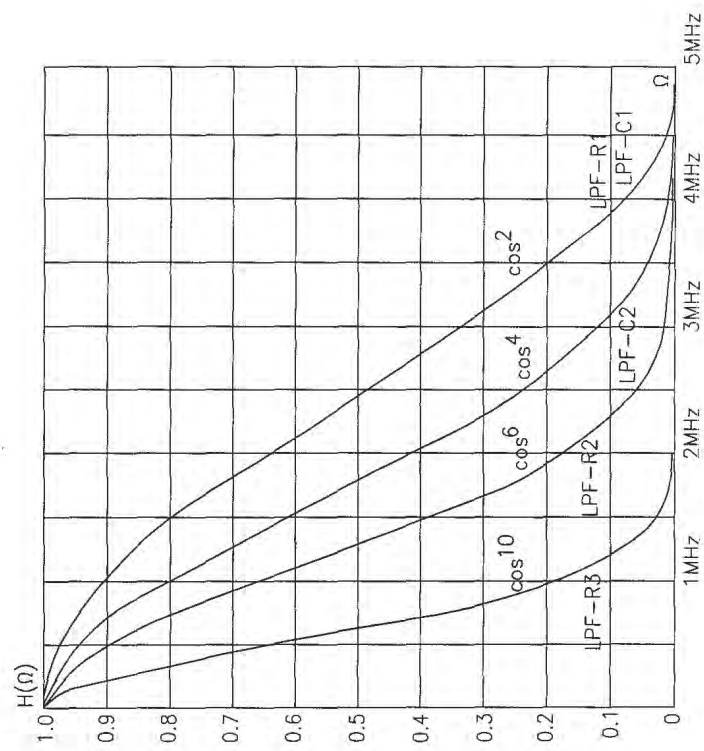


FIG. 6

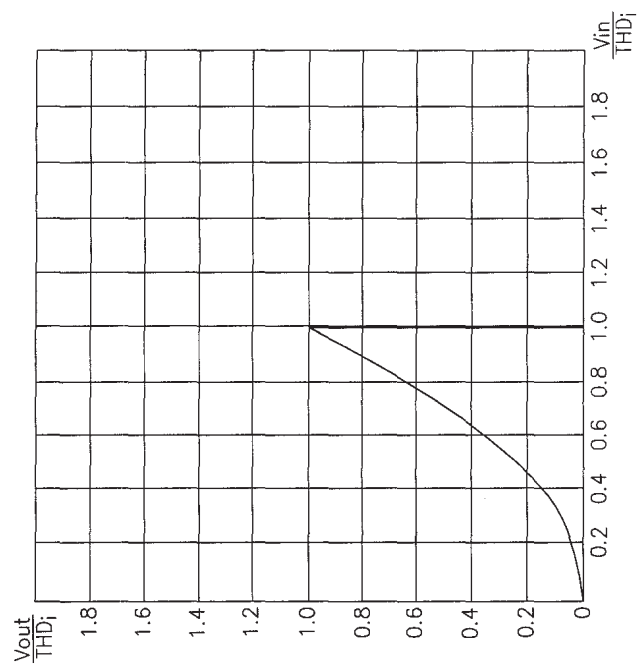
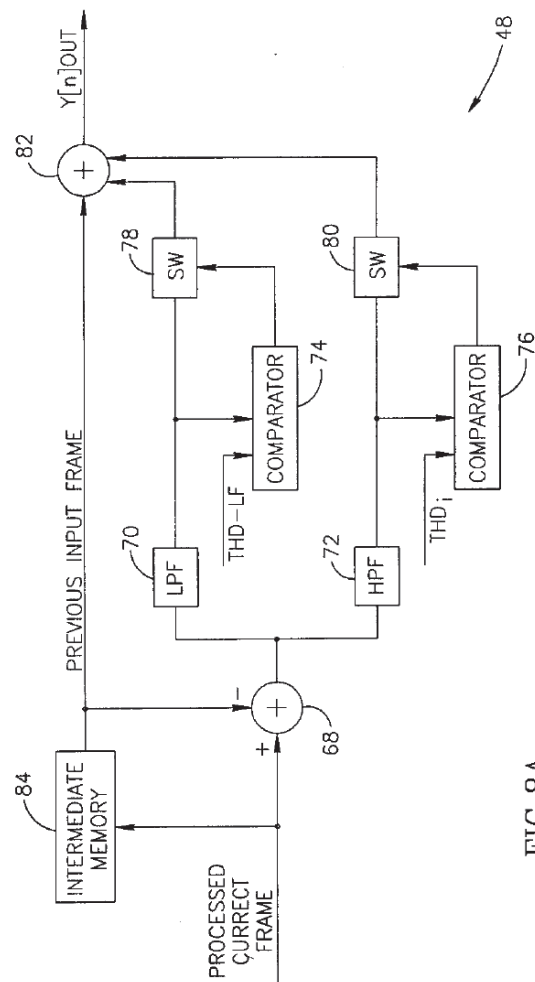


FIG. 7



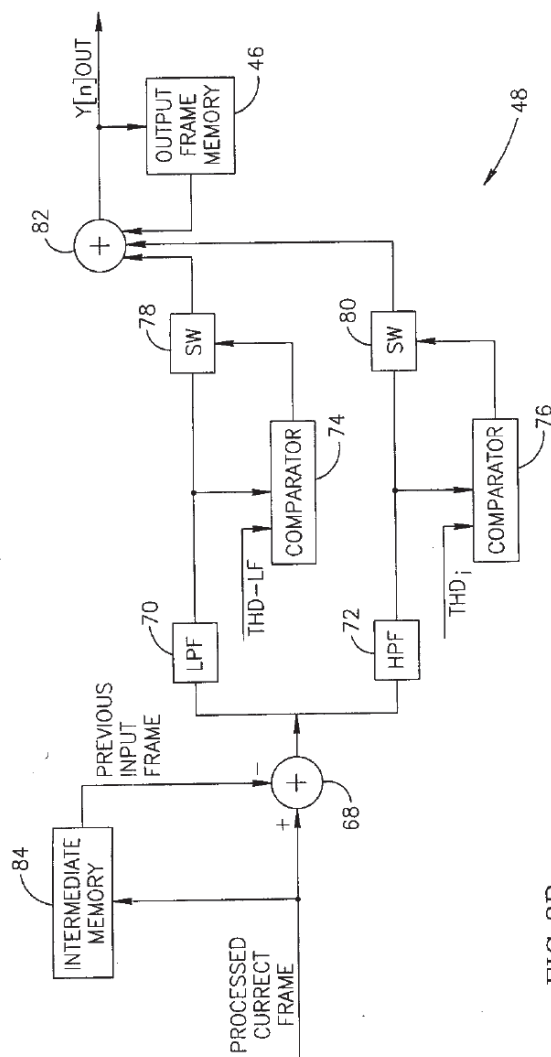


FIG. 8B

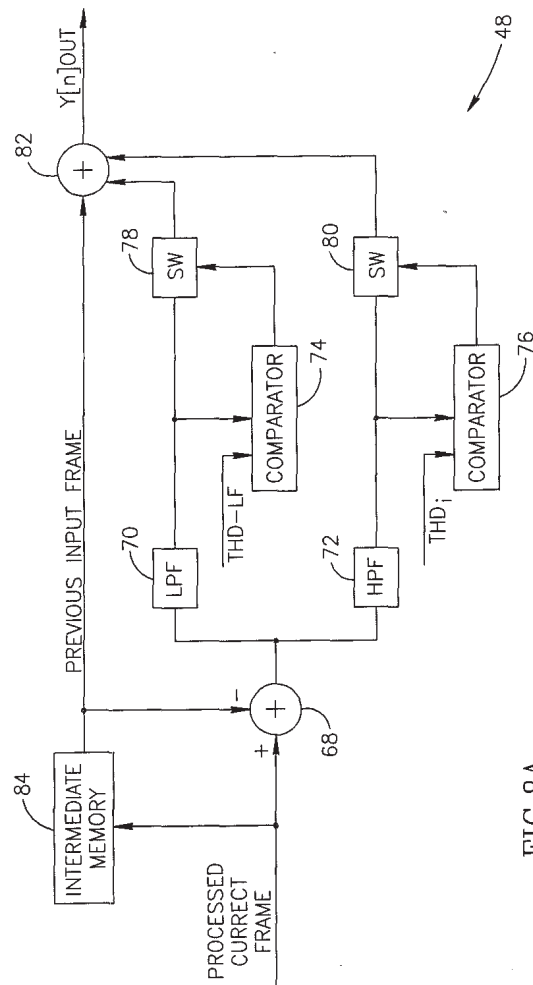


FIG.8A

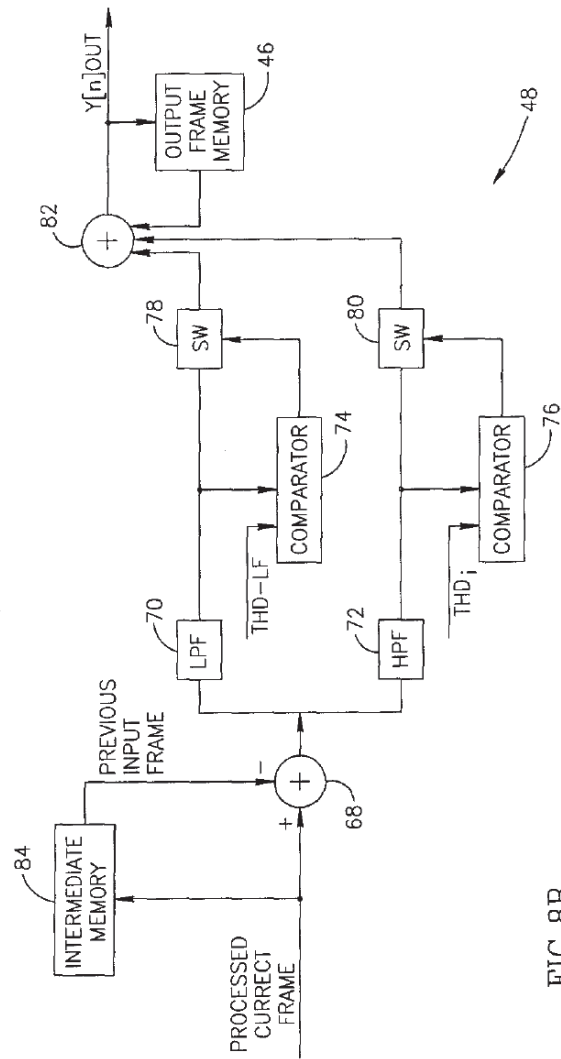


FIG. 8B

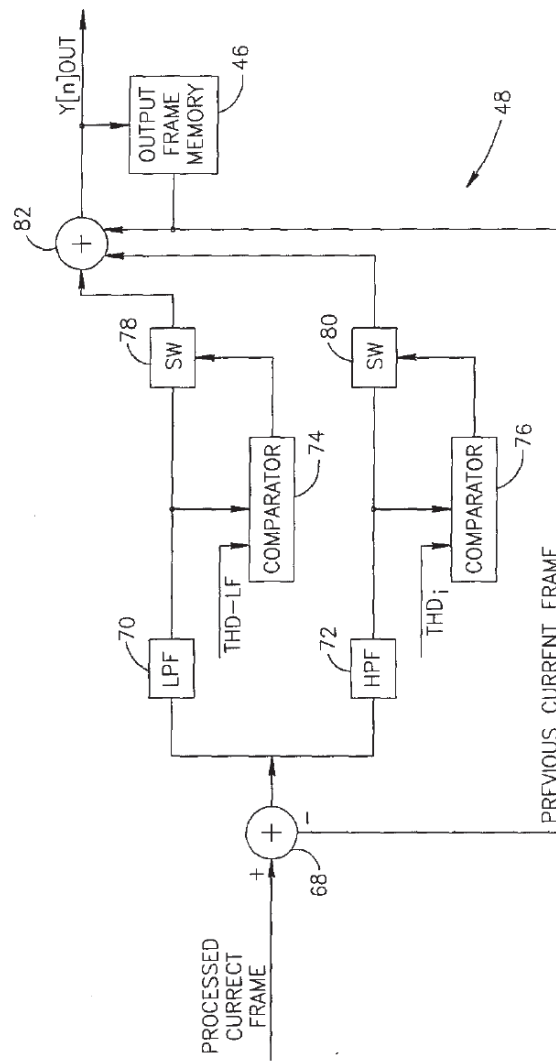


FIG. 8C

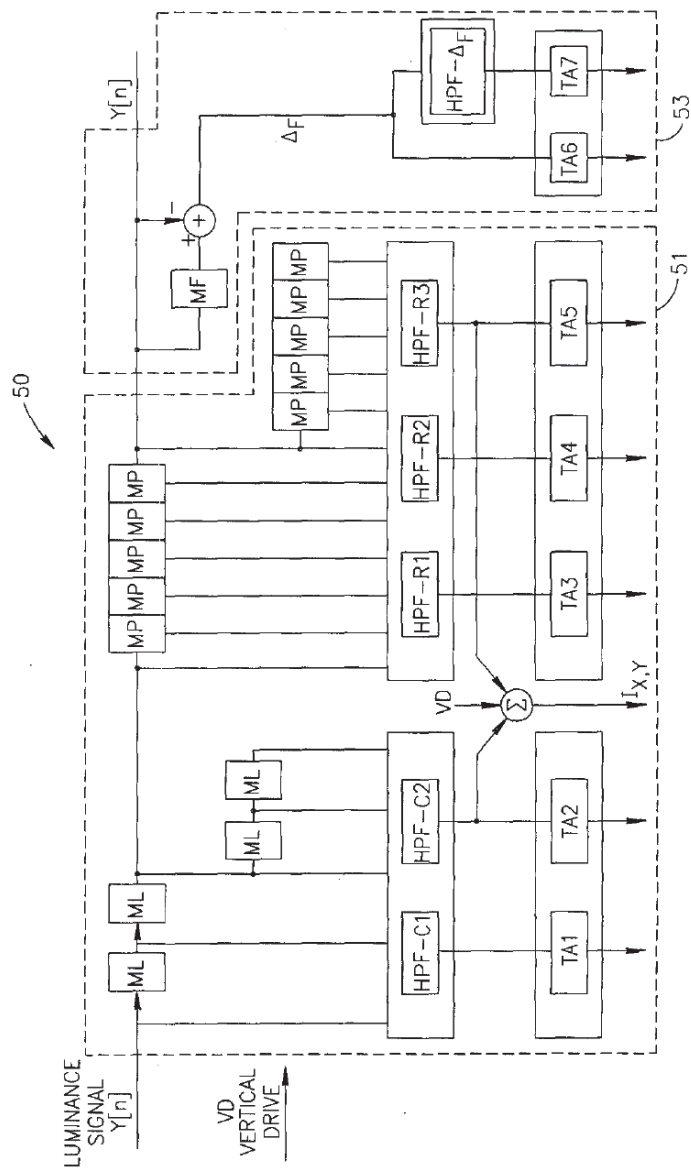


FIG. 9

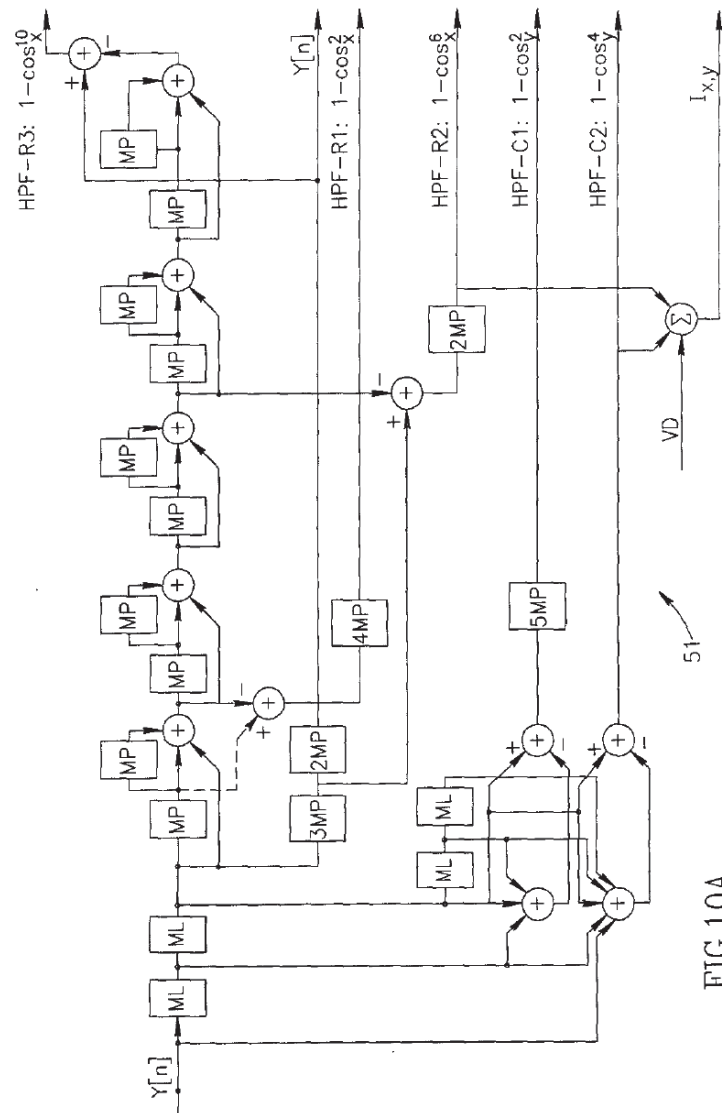


FIG.10A

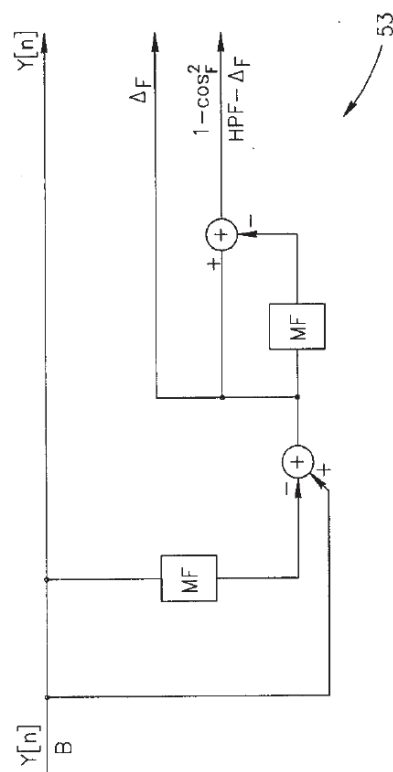
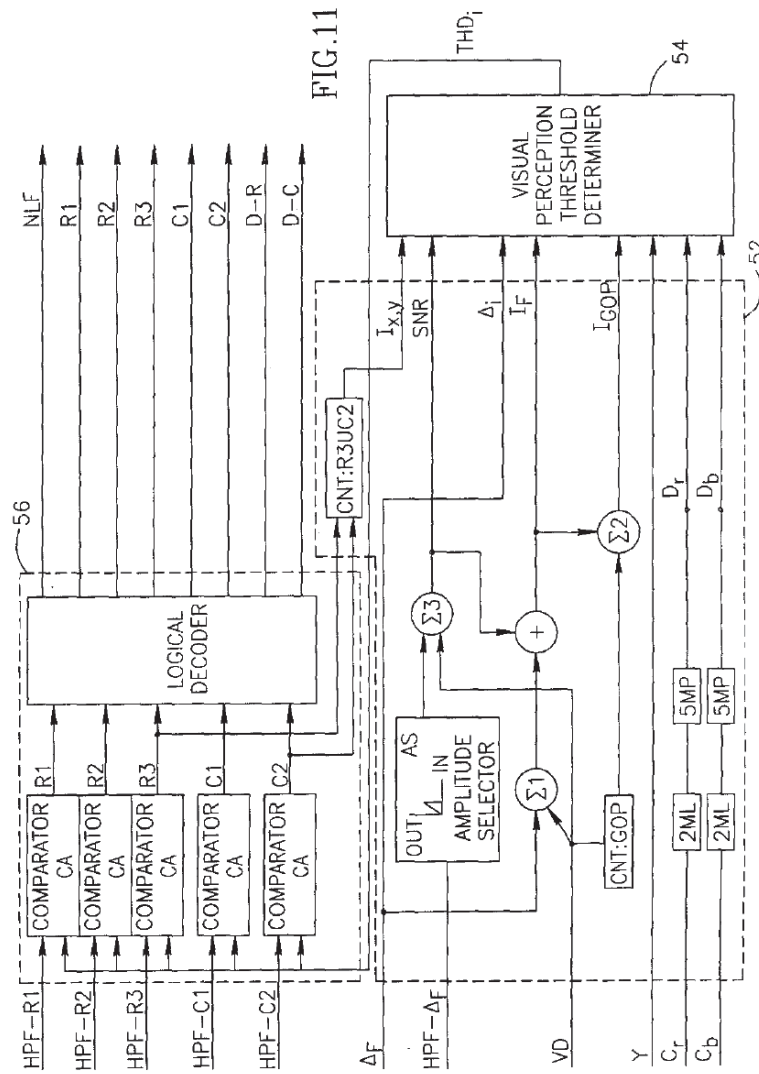


FIG.10B





UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

APPLICATION NUMBER	FILING/RECEIPT DATE	FIRST NAMED APPLICANT	ATTORNEY DOCKET NUMBER
09/524618	03/14/2000	Semion Sheraizim	P-2883-US

Heidi M Brun
Eitan Pearl Latzer & Cohen-Zedek
One Crystal Park Suite 210
2011 Crystal Drive
Arlington, VA 22202-3709

Examiner

Tran, Phuoc

Art Unit	Paper Number
2621	#11

Date Mailed:
July 8, 2002

Notice Regarding Drawings

Corrected drawings for the above-identified application, received in the USPTO on 03/18/2002, are still not acceptable for the reason(s) identified on the attached PTO-948. Applicant is given one opportunity to correct the informalities within a two month time period from the mailing date of this Notice. **THIS TIME PERIOD IS NOT EXTENDABLE UNDER EITHER 37 CFR 1.136(a) OR 1.136(b).** Failure to take corrective action within the set period will result in abandonment of the application.

ATTACHMENT: PTO-948 Notice of Draftsperson's Patent Drawing Review

Draftsman: William J Sale
Office of Patent Publication,
Publishing Division
703-308-3573 (To contact me, first call the relay service at 711)

Rev. 10/01

9/475350

NOTICE OF DRAFTSPERSON'S
PATENT DRAWING REVIEWThe drawing(s) filed (insert date) 3/18/02:A. ☐ approved by the Draftsperson under 37 CFR 1.84 or 1.152.B. ☒ objected to by the Draftsperson under 37 CFR 1.84 or 1.152 for the reasons indicated below. The Examiner will require submission of new, corrected drawings when necessary. Corrected drawing must be submitted according to the instructions on the back of this notice.

1. DRAWINGS. 37 CFR 1.84(a): Acceptable categories of drawings: Black ink. Color. Color drawings are not acceptable until petition is granted. Fig(s) _____ Pencil and non black ink not permitted. Fig(s) _____	8. ARRANGEMENT OF VIEWS. 37 CFR 1.84(i) Words do not appear on a horizontal, left-to-right fashion when page is either upright or turned so that the top becomes the right side, except for graphs. Fig(s) _____
2. PHOTOGRAPHS. 37 CFR 1.84(b) 1 full-tone set is required. Fig(s) _____ Photographs may not be mounted. 37 CFR 1.84(c) Poor quality (half-tone). Fig(s) _____	9. SCALE. 37 CFR 1.84(k) Scale not large enough to show mechanism without crowding when drawing is reduced in size to two-thirds in reproduction. Fig(s) _____
3. TYPE OF PAPER. 37 CFR 1.84(e) Paper not flexible, strong, white, and durable. Fig(s) _____ Erasures, alterations, overwritings, interlineations, folds, copy machine marks not accepted. Fig(s) _____ Mylar, velum paper is not acceptable (too thin). Fig(s) _____	10. CHARACTER OF LINES, NUMBERS, & LETTERS. 37 CFR 1.84(l) Lines, numbers & letters not uniformly thick and well defined, clean, durable, and black (poor line quality). Fig(s) <u>Fig 1 #14</u>
4. SIZE OF PAPER. 37 CFR 1.84(f): Acceptable sizes: 21.0 cm by 29.7 cm (DIN size A4) 21.6 cm by 27.9 cm (8 1/2 x 11 inches) All drawing sheets not the same size. Sheet(s) _____ Drawings sheets not an acceptable size. Fig(s) _____	11. SHADING. 37 CFR 1.84(m) Solid black areas pale. Fig(s) _____ Solid black shading not permitted. Fig(s) _____ Shade lines, pale, rough and blurred. Fig(s) _____
5. MARGINS. 37 CFR 1.84(g): Acceptable margins: Top 2.5 cm Left 2.5 cm Right 1.5 cm Bottom 1.0 cm SIZE: A4 Size Top 2.5 cm Left 2.5 cm Right 1.5 cm Bottom 1.0 cm SIZE: 8 1/2 x 11 Margins not acceptable. Fig(s) <u>1</u> Top (T) _____ Left (L) _____ Right (R) _____ Bottom (B) _____	12. NUMBERS, LETTERS, & REFERENCE CHARACTERS. 37 CFR 1.84(p) Numbers and reference characters not plain and legible. Fig(s) <u>1</u> Figure legends are poor. Fig(s) <u>1</u> Numbers and reference characters not oriented in the same direction as the view. 37 CFR 1.84(p)(1) Fig(s) _____ English alphabet not used. 37 CFR 1.84(p)(2) Figs _____ Numbers, letters and reference characters must be at least .32 cm (1/8 inch) in height. 37 CFR 1.84(p)(3) Fig(s) _____
6. VIEWS. 37 CFR 1.84(h) REMINDER: Specification may require revision to correspond to drawing changes. Partial views. 37 CFR 1.84(h)(2) Brackets needed to show figure as one entity. Fig(s) _____ Views not labeled separately or properly. Fig(s) _____ Enlarged view not labeled separately or properly. Fig(s) _____	13. LEAD LINES. 37 CFR 1.84(q) Lead lines cross each other. Fig(s) _____ Lead lines missing. Fig(s) _____
7. SECTIONAL VIEWS. 37 CFR 1.84 (h)(3) Hatching not indicated for sectional portions of an object. Fig(s) _____ Sectional designation should be noted with Arabic or Roman numbers. Fig(s) _____	14. NUMBERING OF SHEETS OF DRAWINGS. 37 CFR 1.84(i) Sheets not numbered consecutively, and in Arabic numerals beginning with number 1. Sheet(s) _____
	15. NUMBERING OF VIEWS. 37 CFR 1.84(u) Views not numbered consecutively, and in Arabic numerals, beginning with number 1. Fig(s) _____
	16. CORRECTIONS. 37 CFR 1.84(w) Corrections not made from prior PTO-948 dated _____
	17. DESIGN DRAWINGS. 37 CFR 1.152 Surface shading shown not appropriate. Fig(s) _____ Solid black shading not used for color contrast. Fig(s) _____

COMMENTS

- characters in shaded area must be shown clear (Fig 1)
- Long lead line of (Fig 1 #14)
Arrowhead - lead line of (Fig 1 #14)

REVIEWER

WY

DATE

7/8/02

TELEPHONE NO.

7033058404

ATTACHMENT TO PAPER NO.

#11



#13 93
612
CV
PATENT
P-2883-US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT(S): SHERAIZIN, Semion M. et al.

SERIAL NO: 09/524,618 Examiner: TRAN, Phuoc

FILED: March 14, 2000 GROUP ART UNIT: 2621

FOR: METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE
SYNTACTIC ENCODING

ASSISTANT COMMISSIONER FOR PATENTS
WASHINGTON, DC 20231


BOX ISSUE FEE
Attn: Official Draftsperson

SUBMISSION OF FORMAL DRAWINGS

Sir:

In response to a Notice Regarding Drawings dated July 8, 2002, Applicant(s) submit(s) new page of corrected Figure 1 for the above-identified Application. If there are any questions regarding these drawings, please call the undersigned attorney of record at (703) 486-0600.

Respectfully submitted,


Vladimir Sherman
Attorney for Applicant(s)
Registration No. 43,116

Dated: August 6, 2002

Eitan, Pearl, Latzer & Cohen-Zedek
One Crystal Park, Suite 210
2011 Crystal Drive
Arlington, VA 22202-3709
U.S.A.

Tel: (703) 486-0600
Fax: (703) 486-0800

12/01

6473532

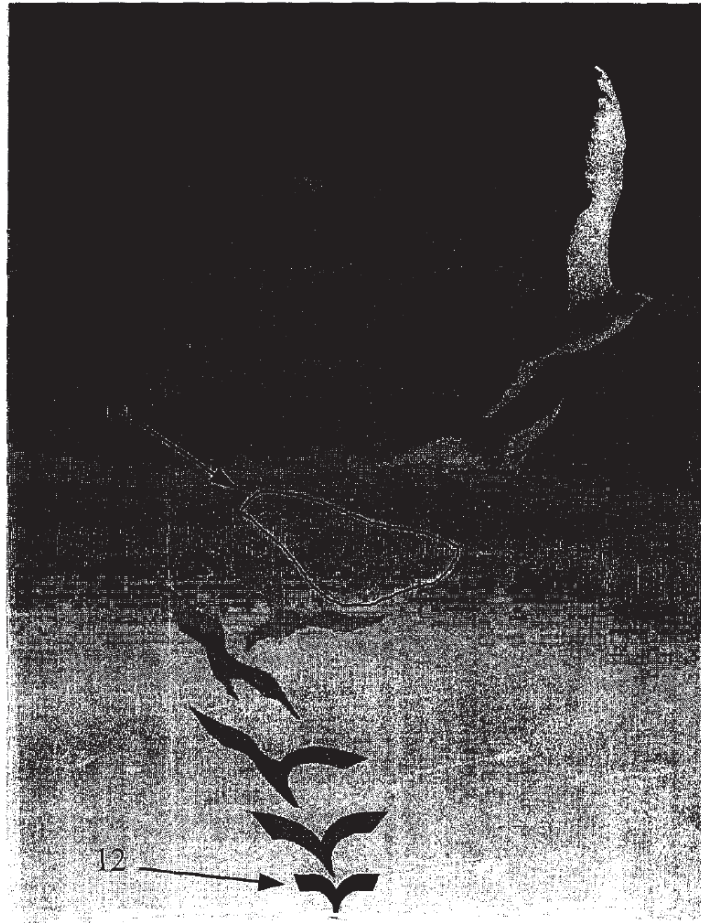


Fig. 1

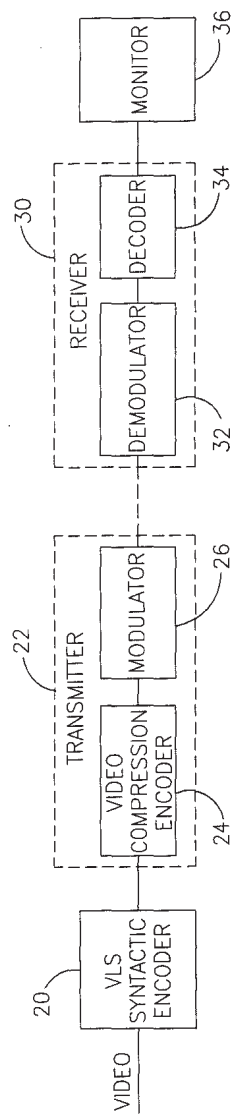


FIG. 2

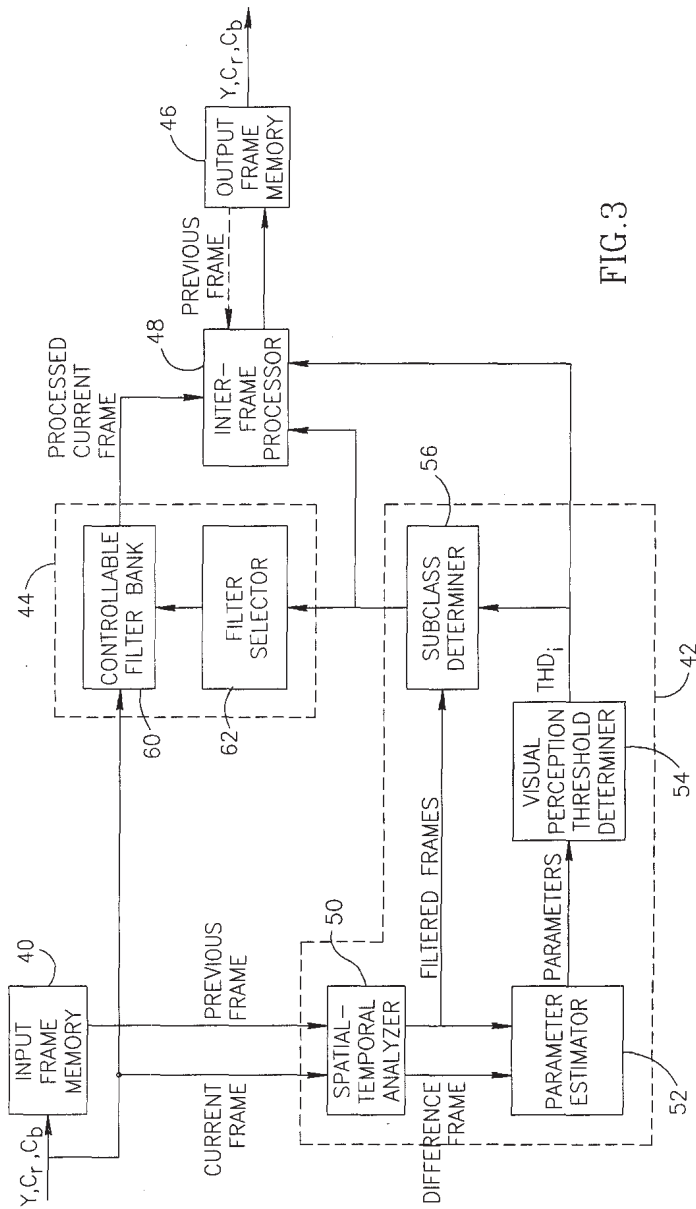


FIG. 3

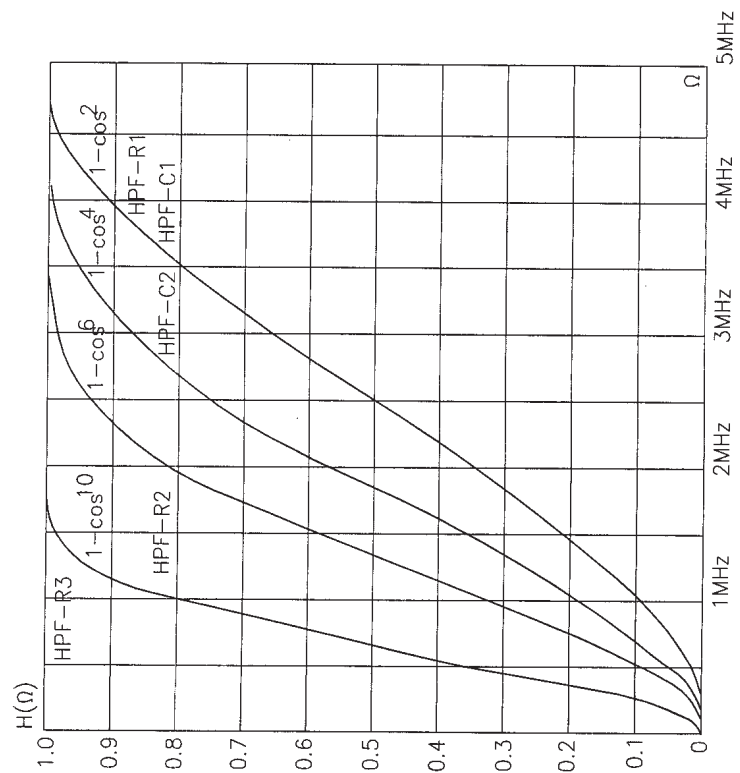


FIG. 4

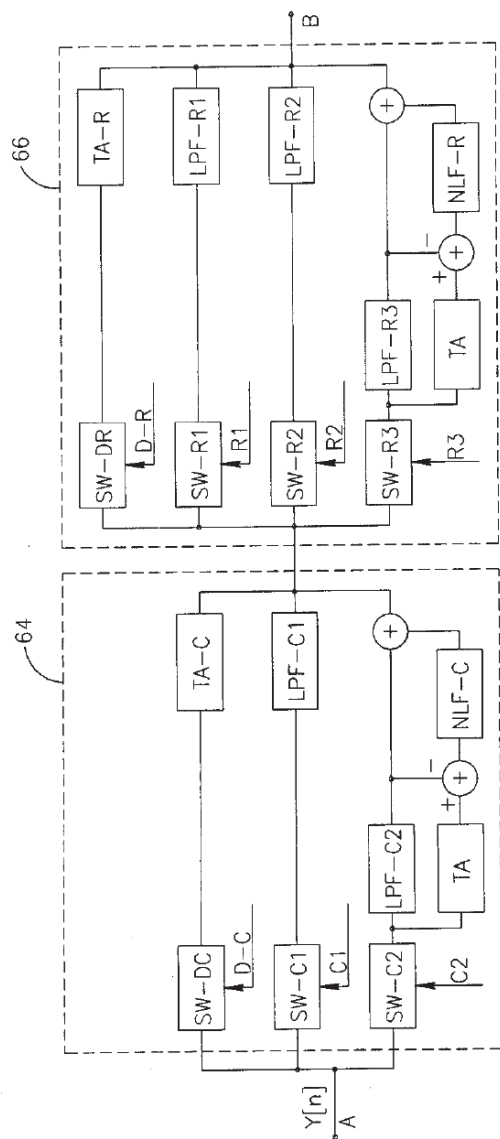


FIG.5A

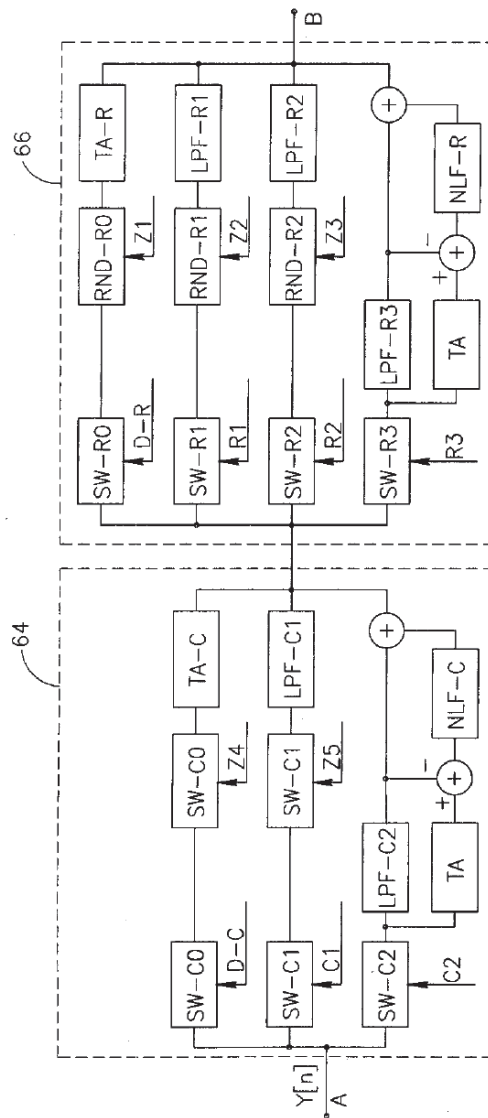


FIG.5B

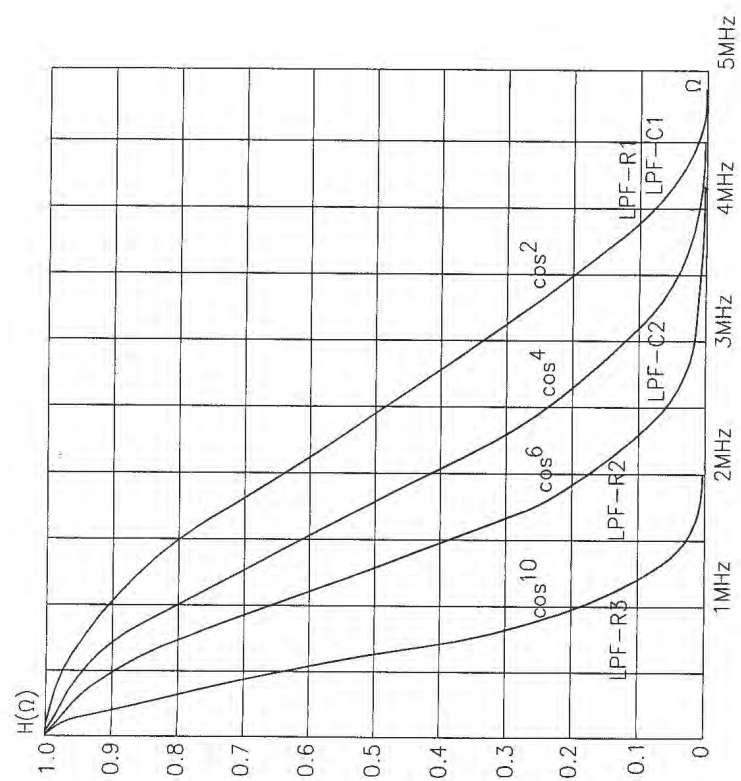


FIG. 6

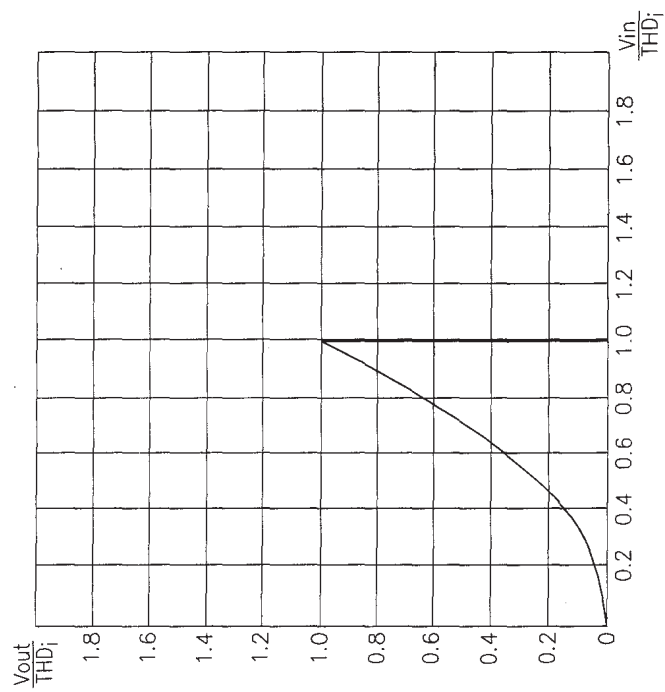
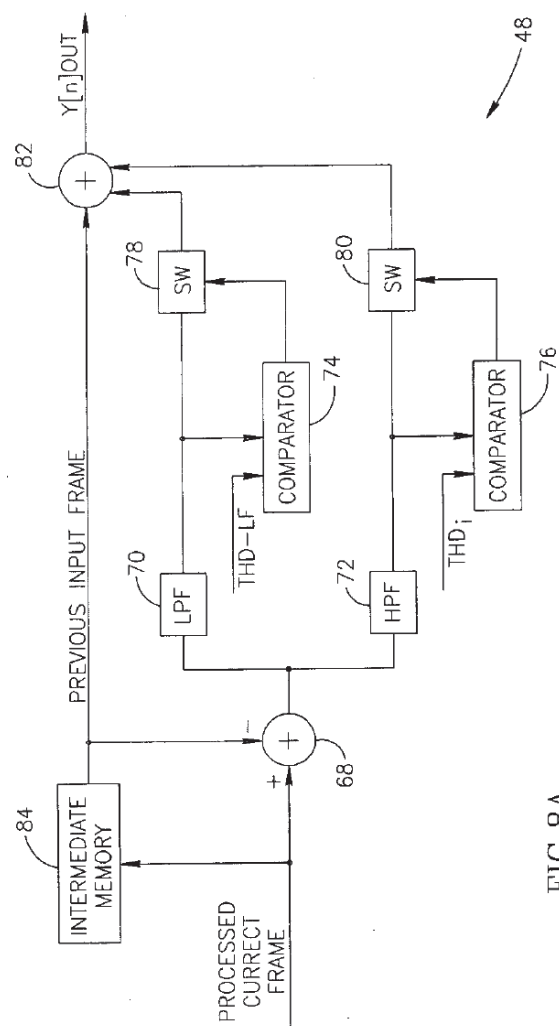
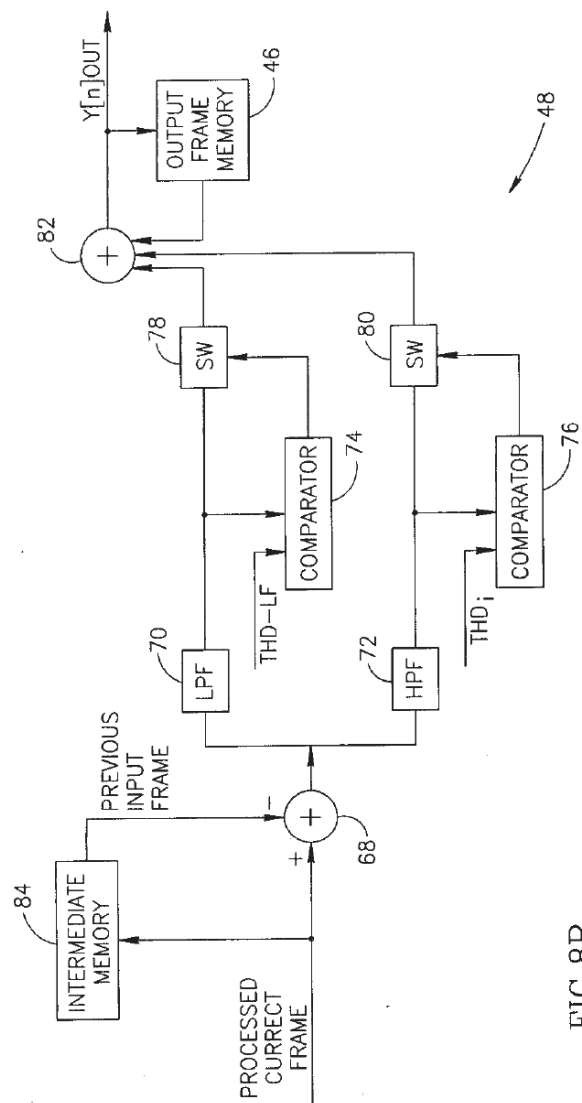


FIG. 7

8





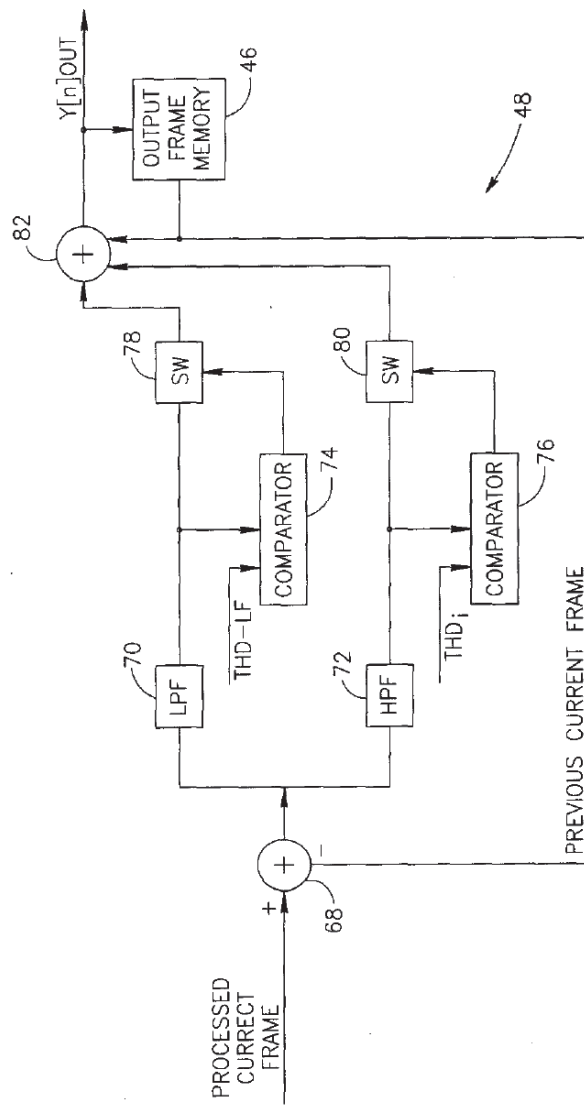


FIG. 8C

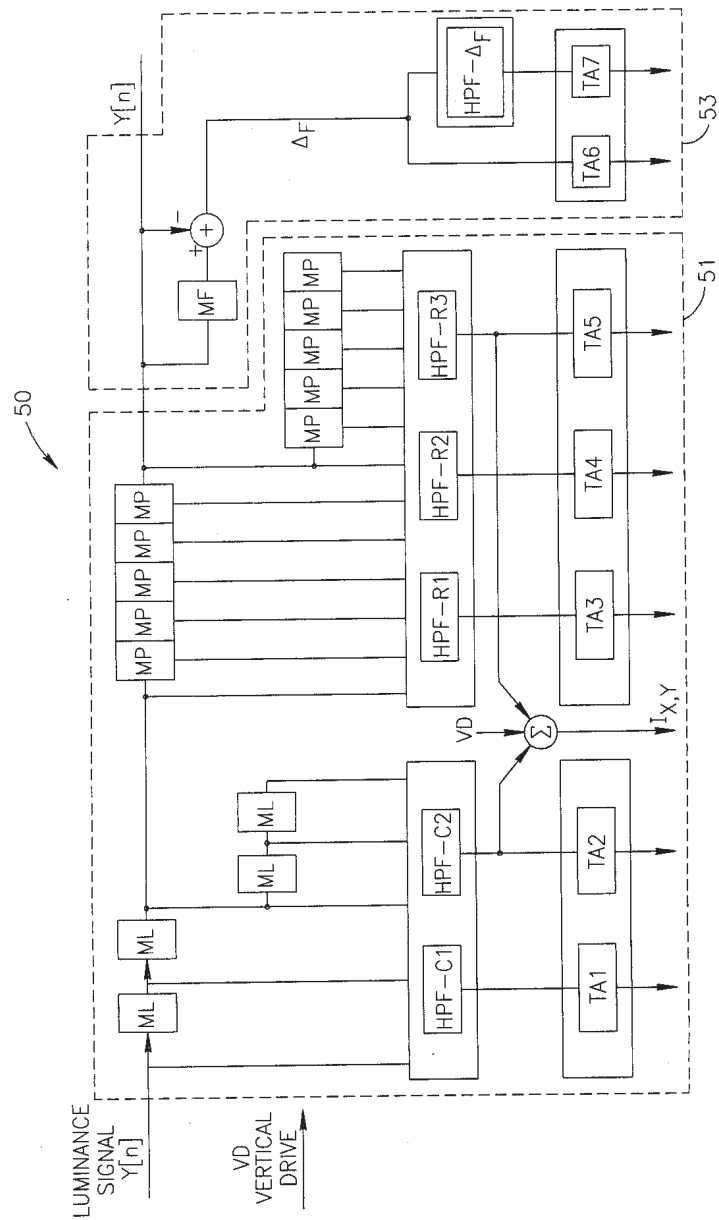


FIG. 9

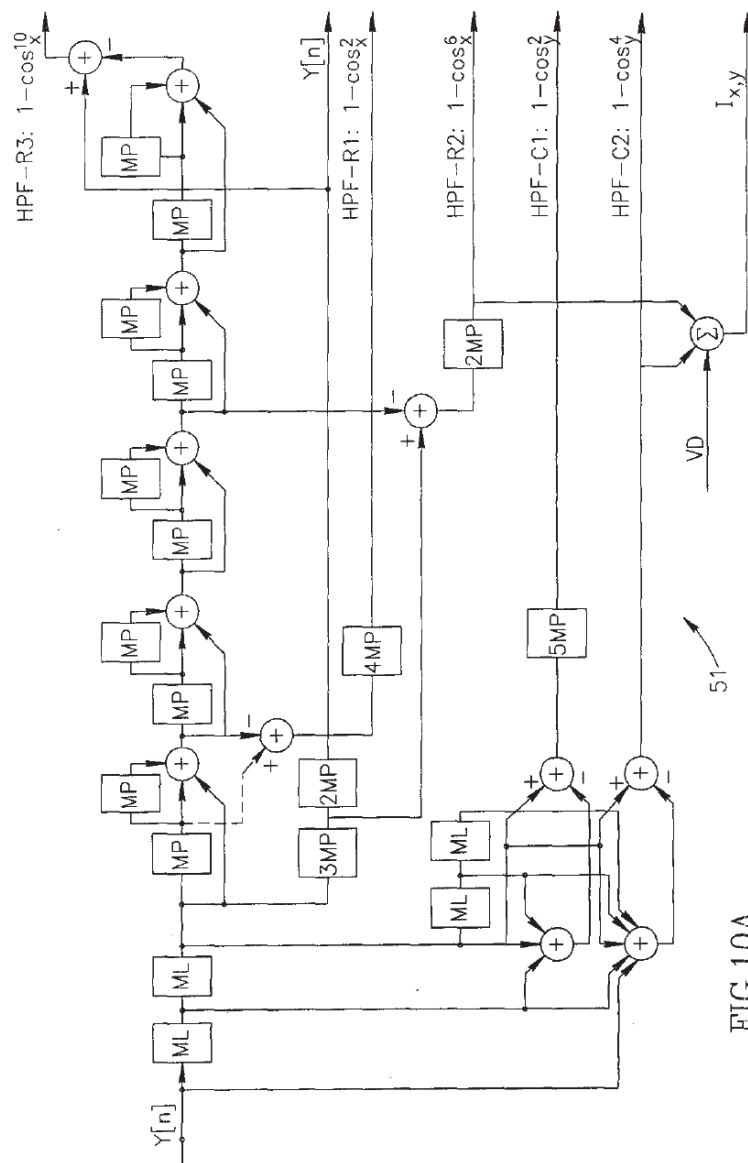


FIG. 10A

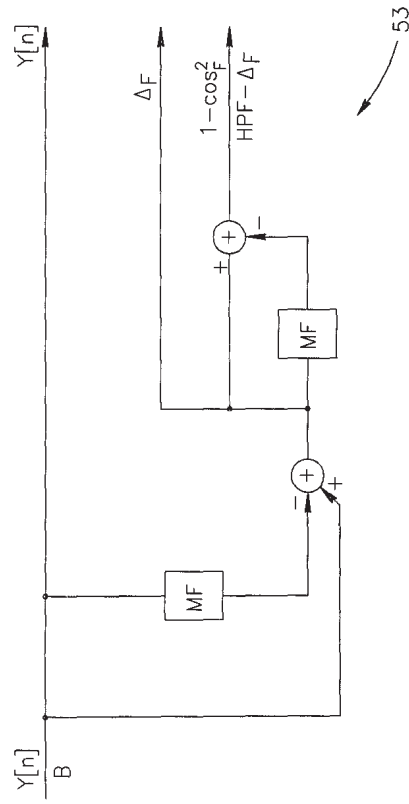
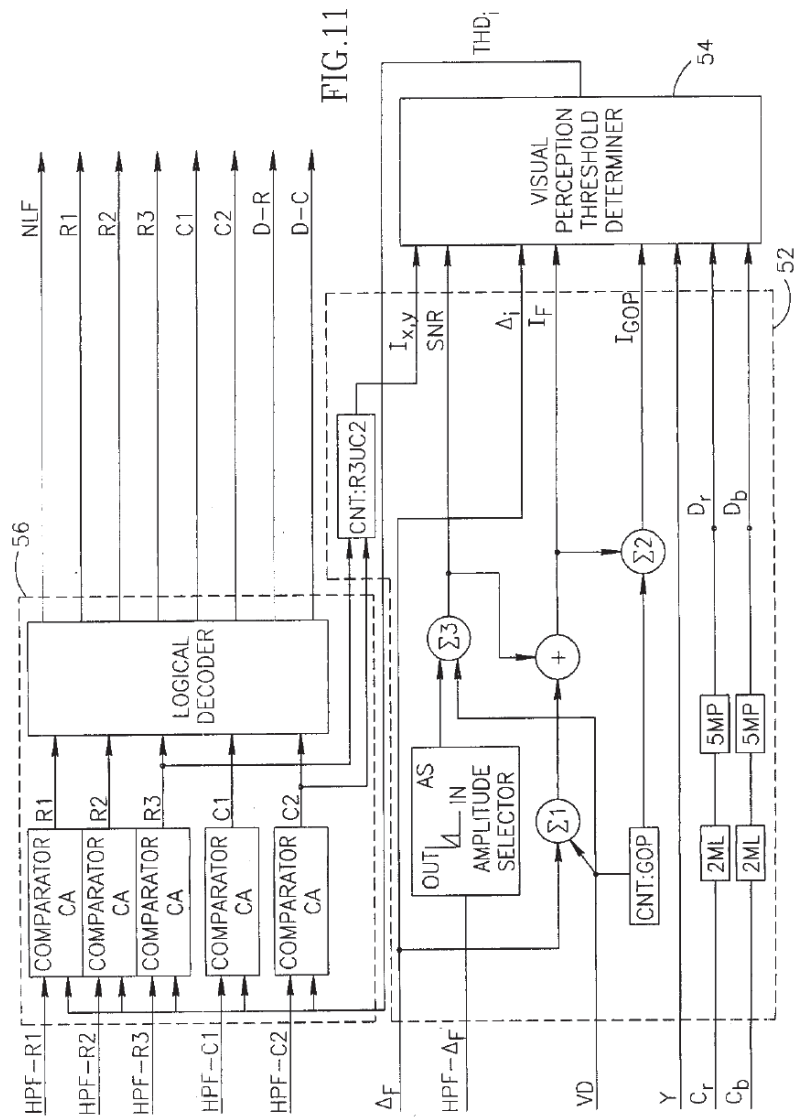


FIG.10B



3015Y

PATENT APPLICATION FEE DETERMINATION RECORD				Application or Docket Number	
Effective December 29, 1999					
CLAIMS AS FILED - PART I					
	(Column 1)	(Column 2)			
FOR	NUMBER FILED	NUMBER EXTRA			
BASIC FEE					
TOTAL CLAIMS	17	minus 20 = *			
INDEPENDENT CLAIMS	2	minus 3 = *			
MULTIPLE DEPENDENT CLAIM PRESENT					
* If the difference in column 1 is less than zero, enter "0" in column 2					
CLAIMS AS AMENDED - PART II					
	(Column 1)	(Column 2)	(Column 3)		
AMENDMENT A	CLAIMS REMAINING AFTER AMENDMENT	HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		
	Total	*	Minus	**	=
	Independent	*	Minus	***	=
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM				
AMENDMENT B	CLAIMS REMAINING AFTER AMENDMENT	HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		
	Total	*	Minus	**	=
	Independent	*	Minus	***	=
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM				
AMENDMENT C	CLAIMS REMAINING AFTER AMENDMENT	HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		
	Total	*	Minus	**	=
	Independent	*	Minus	***	=
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM				
* If the entry in column 1 is less than the entry in column 2, write "0" in column 3. ** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20." *** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3." The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.					

SMALL ENTITY TYPE <input type="checkbox"/>		OR		OTHER THAN SMALL ENTITY	
RATE	FEE			RATE	FEE
	345.00	OR			690.00
X\$ 9=		OR		X\$18=	
X39=		OR		X78=	
+130=		OR		+260=	
TOTAL		OR		TOTAL	690

SMALL ENTITY		OR		OTHER THAN SMALL ENTITY	
RATE	ADDITIONAL FEE			RATE	ADDITIONAL FEE
X\$ 9=		OR		X\$18=	
X39=		OR		X78=	
+130=		OR		+260=	
TOTAL ADDIT. FEE		OR		TOTAL ADDIT. FEE	

SMALL ENTITY		OR		OTHER THAN SMALL ENTITY	
RATE	ADDITIONAL FEE			RATE	ADDITIONAL FEE
X\$ 9=		OR		X\$18=	
X39=		OR		X78=	
+130=		OR		+260=	
TOTAL ADDIT. FEE		OR		TOTAL ADDIT. FEE	

SMALL ENTITY		OR		OTHER THAN SMALL ENTITY	
RATE	ADDITIONAL FEE			RATE	ADDITIONAL FEE
X\$ 9=		OR		X\$18=	
X39=		OR		X78=	
+130=		OR		+260=	
TOTAL ADDIT. FEE		OR		TOTAL ADDIT. FEE	

MULTIPLE DEPENDENT CLAIM FEE CALCULATION SHEET (FOR USE WITH FORM PTO-875)							APPLICANT(S) 52461P						
CLAIMS													
	AS FILED		AFTER 1st AMENDMENT		AFTER 2nd AMENDMENT			•		•		•	
	IND.	DEP.	IND.	DEP.	IND.	DEP.		IND.	DEP.	IND.	DEP.	IND.	DEP.
1							51						
2							52						
3							53						
4							54						
5							55						
6							56						
7							57						
8							58						
9							59						
10							60						
11							61						
12							62						
13							63						
14							64						
15							65						
16							66						
17							67						
18							68						
19							69						
20							70						
21							71						
22							72						
23							73						
24							74						
25							75						
26							76						
27							77						
28							78						
29							79						
30							80						
31							81						
32							82						
33							83						
34							84						
35							85						
36							86						
37							87						
38							88						
39							89						
40							90						
41							91						
42							92						
43							93						
44							94						
45							95						
46							96						
47							97						
48							98						
49							99						
50							100						
TOTAL IND.	2						TOTAL IND.						
TOTAL DEP.	15						TOTAL DEP.						
TOTAL CLAIMS	17						TOTAL CLAIMS						

PTO-1360 (3-78) *MAY BE USED FOR ADDITIONAL CLAIMS OR AMENDMENTS U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

AN ADAPTIVE HYBRID CODING SCHEME FOR HDTV AND DIGITAL VIDEO SEQUENCES

Awad Kh. Al-Asmari

Member IEEE

Department of Electrical Engineering
King Saud University, P.O. Box 800
Riyadh 11421, Saudi Arabia

Abstract

In this paper, the image sequences are decomposed into seven sub-bands in the spatial domain using GQMF filters. Then, the base-band is interframe encoded using a simple 3-D block-matching method and LADVQ search algorithm. This band is used to decide the codebooks and the blocks sizes for encoding the high frequency bands. To reduce the complexity, the high-bands are encoded using the LAVQ.

The simulation results show that good quality pictures can be obtained at a bit rate as low as 0.290 bpp (about 15 Mbits/sec). This rate will facilitate the transmission of three HDTV channels via a DS3 rated 45 Mbits/sec channel.

This algorithm is tested on digital video sequences. We obtain a high quality reconstructed images at an average bit rate of 0.460 bpp with PSNR of 35.04 dB.

1 Introduction

In a previous work [1-3], an HDTV codec operating at data rates between 27 and 55 Mbits/sec, including the DS3 rate of 45 Mbits/sec, was proposed. These algorithms employ an intraframe sub-band coding technique for the HDTV images. However, the frame to frame correlation was not taken into consideration. The HDTV picture is trans-

mitted at 30 frames per second, consequently a high degree of redundancy exists between adjacent frames. Therefore, more compression can be achieved by taking advantage of interframe and intraframe data correlation.

The motion compensated image coding technique is one of the most popular interframe coding schemes which takes advantage of the frame to frame redundancy to achieve high data rate compression [4]-[5]. Motion processing is used in all well-known and efficient coding algorithms like MPEG1, MPEG2, H.261 and other coding techniques [6]. It is also the basis of several frame rate standard conversion techniques proposed in the literature [6]. It has been found that the HDTV image sequences can be encoded at a rate of 20Mbits/sec [7]. However, the same sequences can be encoded at a lower rate with comparable complexity. The goal of this research effort is to develop a data compression scheme for the HDTV signals with data rate reduction from about 500 Mbits/sec to around 15 Mbits/sec while still retaining the signal fidelity. Furthermore, the compression scheme is made simple and easy to implement in hardware.

Vector quantization (VQ) is another coding technique for image sequences. Many adaptations and improvements have been made to the basic concept of vector quantization to remove the difficulties of codebook design, the codeword which gives the least

mean squared error (LMSE), and the effect of blockiness when the LMSE is large and the rate is low. A number of reviewed papers which summarize these improvements have been reported in the literature [8]-[9]. The VQ algorithms, designed for coding video images, rely on shared codebooks which are either designed for the entire sequence [10]-[11] or are updated for each image [12]-[13]. These algorithms are adaptive for interframe coding into the source statistics. Other methods for adaptive VQ for image coding based on DPCM have been presented in [14].

This paper introduces a new combined coding technique based on sub-band coding (SBC) for the intraframe coding and vector quantization for the sub-bands. A search region for block matching is selected from the middle frame of the low frequency band. Then, the locally adaptive differential vector quantization (LADVQ) is used to encode this block. The result of the matching will decide the codebook size of the high-frequency data. The details of this coding technique is described in the following sections. The sub-band filtering using the four-band filters is discussed in section 2. The LAVQ algorithm is given in section 3. Whereas the LADVQ algorithm for interframe coding is explored in section 4. While the proposed coding algorithm is presented in section 5. Simulation results are given in section 6. Some conclusions are drawn in section 7.

2 Sub-Band Filtering by GQMF

The image sequences are split into seven bands, in the spatial domain, using four-band generalized quadrature mirror filters (GQMF) given in [2]. First, the image is filtered horizontally into four sub-bands using one-dimensional GQMF followed by 4:1 decimation. Subsequently, the base-band is filtered vertically into four more sub-bands.

The octave splitting of the image for this case is shown in Figure 1.

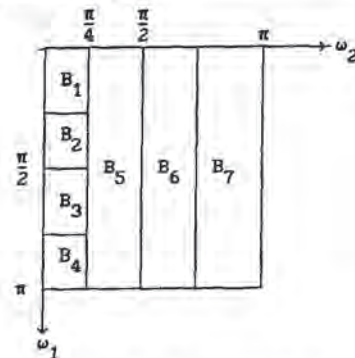


Figure 1. The Octave splitting of the image (seven bands)

The GQMF filters have been used for sub-band coding of images and achieve high compression rates [1]-[2]. The advantage of using GQMF filters is the saving in computational complexity over the conventional QMF. It has been found that four-band GQMF filters scheme will give a 33.33% saving in FFT operation over the QMF [1]-[2]. A number of low-pass prototype filters have been designed [2] using a nominal bandwidth $\pi/2M$, where M is the decimation factor. The filter used are 2D separable GQMF derived from the 1D filter named, 24 tap 4 band.

3 The LAVQ Algorithm

The basic idea of the locally adaptive vector quantization (LAVQ) algorithm is to take a block from the input image sequences and compare it with the stored codewords in the codebook. If there exists a codeword sufficiently close to the image block, within the error allowance, the index itself is sent and the codeword is moved to the top of the codebook. If the codeword does not exist, a special index is sent and followed by the block itself. This block becomes a new codeword

and is placed at the top of the codebook. As a result, all the codewords of the codebook are pushed down. If the number of codewords exceeds the maximum allowed in the codebook, the last codeword is lost [15]-[16].

At the decoder side, if the special index is sent, the receiver expects a block to be received. This block is placed at the top of the receiver codebook and all other codewords are pushed down. If the codebook is already full, the last codeword will be discarded. Therefore, the encoder and decoder will have the same codebook at each step [16].

The basic LAVQ generates a suboptimal codebook with an inferior rate-distortion performance as compare to that of the LBG algorithm [17] which spends more time optimizing the codebook for lower distortion. It has been found that LAVQ algorithm represents detailed areas and edges accurately, but it renders smooth, low detail areas with visible distortion. On the other hand, the LBG algorithm represents the smooth regions without much distortion, and the detail areas and edges with poor quality [16]. The most significant advantage of LAVQ algorithm over other VQ-based image coding algorithm is the high speed in generating and encoding the codebook.

4 The Codebook Design for LADVQ

In this section, the (spatial/temporal) the locally adaptive differential vector quantization (LADVQ) to encode the HDTV signal is introduced. The LADVQ combines the method of LAVQ and adaptive differential pulse code modulation (ADPCM). Hence, LADVQ has many of the compression advantages of both LAVQ and ADPCM. In this paper, a hybrid LADVQ/LAVQ scheme is used. The LADVQ is used for the lowest frequency sub-band (B_1) whereas LAVQ is used to encode the high frequency bands (B_2 - B_7).

The bit rate required to encode the band B_1 using basic LAVQ is found to be twice that needed to encode the same band using LBG algorithm at the same mean square error (MSE). Therefore, improved techniques to encode the image at low bit rate with an acceptable image quality are developed. These improvements are the use of ADPCM technique for the codebook design and the use of the lossless data compression techniques for the encoded information. Moreover, the computational complexity can be reduced by searching for the first occurrence of the codeword that satisfies the error allowance.

4.1 The Base-Band Coding using LADVQ

The base-band B_1 has a histogram that is similar to the original image with high pixel-to-pixel correlation. Due to this high correlation, the ADPCM with non-linear quantization approach is chosen to encode this band. In the ADPCM encoder used, a linear predictor constructs the predicted pixel as weighted summation of previous pixels in both intraframe and interframe. A very effective ADPCM sub-band encoder has been developed in [18]. Five different methods for linear predictive coding are tested on the sub-band data for several different image sequences. It has been found that method 5 yields the best prediction gains but is based on a seventh-order predictor [18]. In this paper, we adapt method 2 because it provides a significant prediction gain with a third-order predictor. The predicted value is given by

$$\hat{x}(i, j, n) = ax(i, j-1, n) + bx(i-1, j, n) + cx(i, j, n-1) \quad (1)$$

where $x(i, j-1, n)$, $x(i-1, j, n)$ and $x(i, j, n-1)$ are previously reconstructed pixels at location (i, j) in the spatial coordinate at time n . The weighting factors a , b and c are adapted for

each frame and optimized to give the required minimum mean square error.

The signal $D(i,j,n)$ is the difference between pixel $x(i,j,n)$ to be coded and corresponding predicted pixel $\hat{x}(i,j,n)$

$$D(i,j,n) = x(i,j,n) - \hat{x}(i,j,n) \quad (2)$$

This error signal is used to design the codebook of the LADVQ. Hence, a block of $N \times N$ is stored in the codebook and used for encoding. The elements of this block are $D(i,j,n)$ at location (i,j) ($i=1,\dots,N$ and $j=1,\dots,N$). Since the difference values for the vectors tend to be small, non-linear quantization and entropy coding of codeword values are used to improve encoding performance. Therefore, a 29-level symmetric, non-uniform Laplacian quantizer is used to quantize the difference values.

The use of lossless data compression algorithms for the quantized values and the indices of the codebook can yield better compression without image quality degradation. The variable-length codes (VLC's) are used to encode the quantized difference vectors. An average probability of each reconstructed level is computed, and the Huffman codes are designed accordingly. The one-dimensional B_1 -code [19] is used to encode the indices.

4.2 Interframe coding

Figure 2 shows the overall hybrid coding system. For the best performance, a combination of hybrid LADVQ/VLC's for B_1 and LAVQ/VLC's for bands $B_2 - B_7$ are considered. The VLC's is a combination of Huffman code for the quantized values and B_1 -code for the indices of the codebook. Moreover, the LADVQ and LAVQ are designed such that the correlation in the spatial/temporal domains are considered. The bands QB_1, QB_2, \dots, QB_7 , shown in Figure 2, are quantized/encoded versions of the bands B_1, B_2, \dots, B_7 .

To preserve the Hierarchical encoding of the sequences, three frames are taken at a

time and encoded independent of the other frames, but with a possibility of sharing the same codebook. Each group of thirty frames in the HDTV sequence are divided into ten groups, three frames each as shown in Figure 3 (a). The search regions in the three frames can be represented as shown in Figure 3(b). The $N \times N$ block size is selected from the middle frame (current frame) and compared with the neighbors of the same size in the (spatial/temporal) domains.

Since the prediction error vector information has to be transmitted for every block, a larger block size will reduce the overall bit rate. However, a small block size, in general, gives a better performance than a larger block size because the object motion can be tracked closely. Therefore, the block size is selected to compromise between the desired bit rate and image quality. Small block size for regions with more detail and larger block size for low detail regions are needed to reduce the distortion rate.

In this paper, the block size of 4×4 is found to be an appropriate choice for B_1 because it contains most of the information about the original image. Large changes of brightness in a picture occur at the edges and those sharp contours are contained in the high-frequency bands (B_2-B_7). The linear prediction for sub-band (B_1) is locally adapted depending on the local activity of the image. This band is used to decide the codebooks and the blocks sizes for encoding the high-bands. Two different LAVQ codebooks are designed for these bands. The first codebook is designed with 8×8 pixel blocks and 255 entry. The second codebook is applied with 128 entry each of 16×16 pixel blocks.

The codebook selection depends on the results of the block matching method used in encoding the base-band. If the matching exists, the block is classified as low activity block. Accordingly, the corresponding blocks of the high-frequency bands will have low information. Therefore, these bands are encoded using the second codebook as of its low bit rate requirements. If the matching

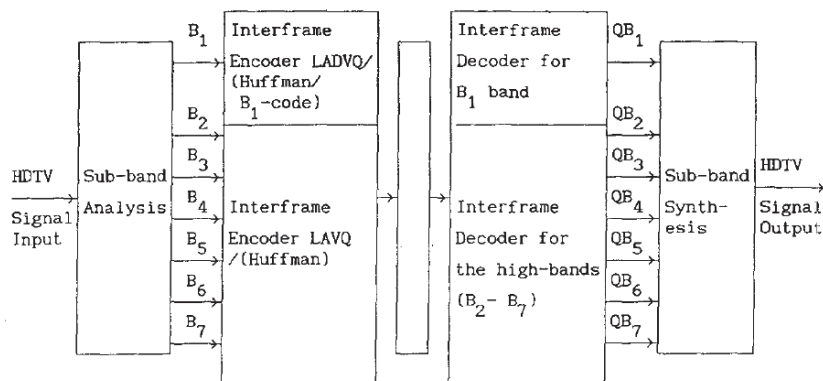
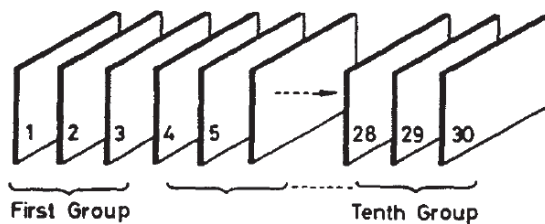
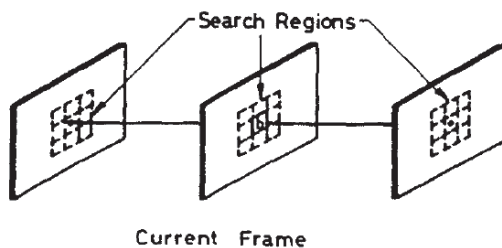


Figure 2: Overall hybrid coding system.



(a)



(b)

Figure 3: Hierarchical encoding of the HDTV frames. (a) Ten groups, three frames each. (b) The 3-D search regions in the group. A block "b" is selected from the current frame.

does not exist, the block of the lowest frequency band (B_1) is classified as an active block. Hence, the corresponding blocks of the high-frequency bands are moving edges or highly textured regions. Therefore, the bit allocation for these bands are increased by using the first codebook.

5 Description of the Algorithm

In this section, the proposed coding system is described for both LAVQ and LADVQ algorithms. The block b of size $N \times N$ is selected from the middle frame. If LADVQ is used to encode the base-band (B_1), the entry of the block b would be the difference between the original pixels of the band and the predicted pixels. The block b is sent first by searching the codebook. The searching method is performed by calculating the difference (d_m) as follow:

$$d_m = \sum_{i=1}^{N^2} \left| \frac{b(i) - v_m(i)}{N^2} \right|, \quad m = 1, 2, \dots, 255 \quad (3)$$

where $b(i)$ is the pixel value of the block b , $v_m(i)$ is the element value of the codebook vector v_m , and d_m is the difference between the block b and the codebook vector v_m . Then, d_m is compared with a pre-decided threshold error (e_{th}). The choice of e_{th} is based on the compromise between the required image quality and the computational complexity. If $d_m \leq e_{th}$, then the index m is sent and the vector v_m is moved into the top of the codebook. Otherwise, a special codeword (SCW) is sent and followed by the block b . The Huffman codes are used to encode this block. Then, block b becomes a new codeword and is placed at the top of the codebook.

The next step is to test the neighbor's blocks for matching with the block b . The search regions in each frame will be a window of size $3N \times 3N$ as shown in Figure 3. Therefore, the number of blocks to be tested

for matching are 26 each of size $N \times N$. The matching method is performed by calculating the difference (d_k) as

$$d_k = \sum_{i=1}^{N^2} \left| \frac{b(i) - b_k(i)}{N^2} \right|, \quad k = 1, 2, \dots, 26 \quad (4)$$

where $b_k(i)$ is the pixel value of the block b_k at location k and d_k is the difference between the block b and the block b_k . Then, d_k is compared with the threshold error (e_{th}). If $d_k \leq e_{th}$, then the matching exists and the block location (index "k") is sent. The modified B_1 -code shown in Table 1 is used to encode the index "k". The C bit in each codeword is a "continuous bit" which can be either "0" or "1". For this application, C is chosen to be "0" for odd locations and "1" for even locations.

All of the 26 blocks will be tested for matching with block b , whenever the matching exists, the index "k" is sent by using modified B_1 -codes. After sending all of the matched blocks, a codeword end of match (EOM) will be sent. Then, the non-matched blocks will be sent by searching the codebooks. The receiver will expect the non-matched block with the least index "k" to be received after the codeword (EOM). For non-matched blocks, the difference d_{km} is given by

$$d_{km} = \sum_{i=1}^{N^2} \left| \frac{b_k(i) - v_m(i)}{N^2} \right|, \quad m = 1, 2, \dots, 255 \quad (5)$$

where $b_k(i)$ is the pixel values of the non-matched block b_k . If $d_{km} \leq e_{th}$, then the index m is sent and the vector v_m is moved into the top of the codebook. Otherwise, the codeword (SCW) is sent and followed by the non-matched block b_k . After sending all of the 26 blocks of the search regions, the codeword end of search (EOS) is sent.

The choices of the codewords SCW, EOM and EOS will give a low bit rate and guarantee that the encoded data will be uniquely decodable at the receiver. The codeword SCW is found to be the one that comes more

Table 1: The modified B_1 -code codewords for the blocks locations and (indices).

b_k	Codeword	b_k	Codeword
1	1C0	2	1C0
3	1C1	4	1C1
5	1C0C0	6	1C0C0
7	1C0C1	8	1C0C1
9	1C1C0	10	1C1C0
11	1C1C1	12	1C1C1
13	1C0C0C0	14	1C0C0C0
15	1C0C0C1	16	1C0C0C1
17	1C0C1C0	18	1C0C1C0
19	1C0C1C1	20	1C0C1C1
21	1C1C0C0	22	1C1C0C0
23	1C1C0C1	24	1C1C0C1
25	1C1C1C0	26	1C1C1C0
	C=0		C=1

frequent, specially, for lower e_{th} . Therefore, SCW should be represented with fewer bits.

In this research, the codebook indices (m), Huffman codes and block locations (k) are sent serially one after another. For this to be practicable, it is obvious that the codes used must be uniquely decodable and distinguishable from each other. So the modified B_1 -codes are designed by adding "1" to the original B_1 -codes at the beginning as shown in Table 1.

The least bits for the codewords SCW, EOM and EOS are selected to be uniquely decodable. Therefore, SCW will be represented by the codeword "111" and each of EOM and EOS will be represented by the codeword "0111". Also, Huffman codes should not have codeword of one bit length. Moreover, the codeword "0111111" should not be in the Huffman codes.

6 Simulation Results

In order to evaluate the hybrid (interframe three codebooks/search regions) algorithm, a simulated coding system is developed. The algorithm is tested on the HDTV test se-

quence MIT to examine the performance. The results presented here represent the computer simulation on 15 frames (MIT01, MIT02,...,MIT15) of the MIT sequences. They are windows of 768×416 pixels in size form the 1125×2000 pixels original images. These HDTV image sequences are divided into five groups three frames each.

Although the final measure of the performance of any compression algorithm is a subjective quality depending on the processed picture, it is convenient to use a quantitative measure during the development phase of the algorithm. This is especially true when the facility to subjectively evaluate the pictures is not available at the same location where the algorithm is being developed. In our laboratory, we do not have the ability of viewing HDTV sequences. Therefore, we use the objective quality measure. A commonly used quantitative measure for processed images is the peak signal to noise ratio (PSNR) being defined as

$$PSNR = 20 \log_{10}(255/\sqrt{MSE}) \quad (6)$$

where MSE is the mean square error. From experience, an SNR of 30 to 32 dB gives good picture quality. The average number of bits per pixel (bpp) is another quantitative measure to evaluate the performance of the compression algorithm. In this paper, the average bit rate is calculated as

$$Average(bpp) = R_q + R_m + R_f \quad (7)$$

where R_q is the number of bits due to quantization, R_m is the number of bits due to block matching, and R_f is the number of bits for sending the flags and the overhead bits.

Table 2 shows the simulation results for the basic LAVQ, the LADVQ and the LBG algorithms. These results are calculated for the base-band of the first group. Since the band B_1 has most of the information about the original image, this band should be encoded with high performance. Hence, e_{th} is

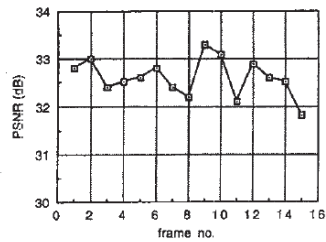


Figure 4. The simulation results for the MIT sequences.

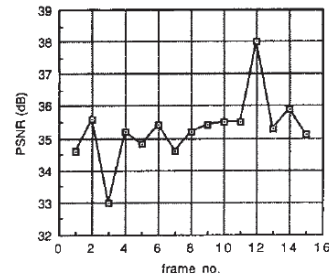


Figure 5. The simulation results for the WALTER sequences.

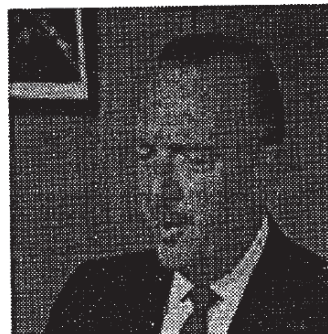
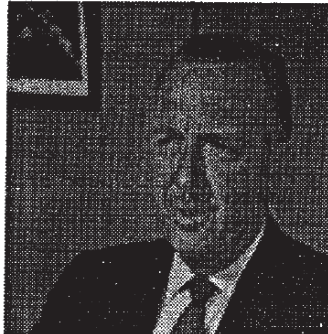


Figure 6. The Original images of WALTER sequences(Frame No. 1 Left and Frame No. 16 Right).

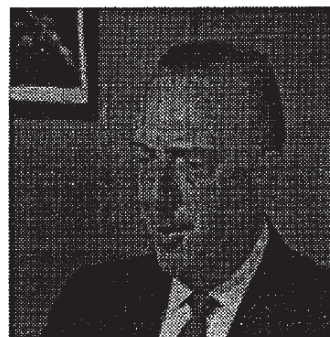
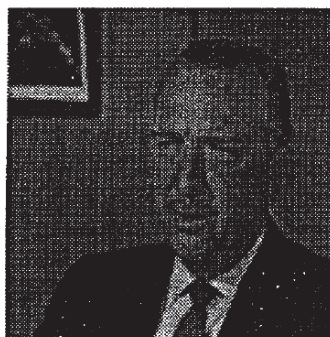


Figure 7. The Reconstructed images of WALTER sequences(Frame No. 1 Left and Frame No. 16 Right).

Table 2: The bit rate and SNR for B_1 of the first group at and MSE = 5.

Algorithm	SNR (dB)	Bit Rate (bpp)
LBG	41.141	2.012
LAVQ	41.141	4.050
ILAVQ	41.141	1.931

selected to be 2.5 for the base-band. On the average, MSE is found to be 5. For fixed error, it is clear that the bit rate required to encode B_1 using LAVQ is more than twice as compared to that required to encode the same band using LBG algorithm. On the other hand, the performance of the LADVQ algorithm is found to be better than that of LBG algorithm. The encoding procedure of LADVQ is also computationally attractive. Although the scheme provides a 3-D hybrid coding, the computational complexity is less than that needed for LBG algorithm. In this study, the LBG algorithm is applied with 256 vectors, each of a block 4×4 pixels.

The simulation results for the proposed technique tested on the seven bands are shown in Table 3. The LADVQ with a block 4×4 pixels is tested on B_1 for the MIT test sequences. The band B_1 is encoded for the fifteen frames using LADVQ with an average MSE = 5.317. The threshold error (e_{th}) is selected to be 6 for bands B_2 , B_3 , B_5 and B_6 . For these bands, the average MSE for the fifteen frames is found to be 35.067. Since there is less information in bands B_4 , B_7 , e_{th} is chosen to be 10. Based on the computer simulations, these thresholds give an acceptable image quality at low bit rate.

The interframe coding algorithm developed reconstructs the base-band (B_1) with an average SNR of 40.874 dB at about 1.918 bpp. Figure 4 shows the simulation results for the first fifteen frames of the MIT sequences. It is found that this scheme gives good quality images at an average bit rate of 0.290 bpp for the test sequences with an average SNR of about 32.570 dB.

In a previous study [7], the first group

Table 3: The threshold error and MSE for the seven bands of the and MIT test sequences.

Band(s)	Codebook	e_{th}	MSE
B_1	LADVQ	2.5	5.317
B_2, B_3, B_5, B_6	LAVQ	6	35.067
B_4 and B_7	LAVQ	10	89.13

has been encoded with a bit rate around 20 Mbits/sec with a SNR of about 33 dB. In that study, all reconstructed images are found to have very good subjective quality. From Figure 4, it can be seen that the same images can be reconstructed with 0.287 bpp (about 14.82 Mbits/sec) with a SNR of about 32.737 dB. These results show that an improvement, in the bit rate, of about 5 Mbits/sec can be obtained with a comparable SNR (i.e. only 0.263 dB less).

To test the performance of this algorithm on a regular digital video images, the WALTER image sequences of 256×256 pixels are tested. The result of this simulation is shown in Figure 6 for the first fifteen frames. It is found that these images can be encoded at an average of 35.04 dB with bit rate of 0.460 bpp. Figure 6 shows the original images for the first and the fifteenth frames. The reconstructed for these frames are shown in Figure 7. Therefore, it can be concluded that the digital video images used in this research can be reconstructed with high subjective quality as shown in Figure 7.

7 Conclusions

An efficient hybrid coding algorithm which compresses digitized HDTV signal to 15 Mbits/sec is developed. This rate will facilitate the transmission of three HDTV channels on the DS3 rated 45 Mbits/sec channel. Moreover, the computational complexity is reduced by using the GQMF filters. The overall delay is found to be small using such filters; which is of great importance in high

data rate applications such as HDTV.

The base-band is interframe encoded using a simple 3-D block-matching method and LADVQ search algorithm. To reduce the complexity, the high-bands are encoded using the basic LAVQ.

The performance of the overall algorithm is improved by using lossless compression technique. The simulation results show that good quality pictures can be obtained at a bit rate as low as 0.290 bpp (15 Mbits/sec) for the HDTV images. The results show that an improvement of about 5 Mbits/sec can be obtained when compared with the higher complexity algorithm reported in literature for the same test sequences.

For the digital video images, this algorithm shows a good performance. The visual quality of the reconstructed images is found to be the same as the original images at an average low bit rate of 0.460 bpp. Therefore, it can be concluded that an adaptive algorithm is developed for different images resolution.

Acknowledgement

The author would like to thank Abobakr Sultan Ahmed of King Saud University, College of Engineering, Riyadh for his help in developing the processed images.

References

- [1] A. Kh. Al-Asmari, "Sub-band coding of HDTV images using GQMF," *IEE Electron. Lett.*, vol. 28, no. 14, pp. 1335-1337, July 1992.
- [2] N. Coppisetti, S. C. Kwatra, and A. Kh. Al-Asmari, "Low complexity subband encoding for HDTV images" *IEEE J. Select. Areas Commun.*, vol 11, no. 1, pp. 77-87, Jan. 1993.
- [3] A. Kh. Al-Asmari and R. E. Ahmed, "VLSI Architecture for HDTV Sub-Band Coding Using GQMFs Filterbanks" *IEEE Trans. Consumer Elec.*, vol 41, no. 1, pp. 1-11, Feb.1995.
- [4] A. N. Netravali and J. D. Rabbits, "motion-compensated television coding: part I," *Bell Syst. Techn. J.*, vol. 58, no. 3, pp. 631-670, Mar. 1979.
- [5] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice Hall, 1984.
- [6] H. Li, A. Lundmark and R. Forchheimer, "Image Sequence Coding at Very Low Bit Rates: A Review," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 589-609, Sept. 1994.
- [7] N. Coppisetti and S. C. Kwatra, "A low complexity 20 Mbits/sec. CODEC for HDTV sequences," in *picture coding symposium '93*, Switzerland, Mar. 17-21, 1993.
- [8] N. M. nasrabadi and R. A. King, "Image coding using vector quantization: A review," *IEEE Trans. Commun.*, vol. 36, pp. 957-971, no. 8, Aug. 1988.
- [9] R. M. Gray, "Vector quantization," *IEEE ASSP Magazine*, vol. 1, no. 2, pp. 4-29, Apr. 1984.
- [10] H. Sun and M. Goldberg, "Adaptive vector quantization for image sequence coding," in *proc. IEEE International Conf. on ASSP*, pp. 339-342, 1985.
- [11] M. Goldberg and H. Sun, "Frame adaptive vector quantization for image sequence coding," *IEEE Trans. Commun.*, vol. 36, pp. 629-635, 1988.
- [12] W. Y. Chan and A. Gersho, "Constrained-storage quantization of multiple vector sources by codebook sharing," *IEEE Trans. Commun.*, vol. 39, no. 1, pp. 11-13, Jan. 1991.
- [13] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. Kluwer Academic Publishers, 1992.

- [14] H. Hang and J. W. Wood, "Predictive vector quantization of image," *IEEE Trans. Commun.*, vol. 33, no. 11, pp. 1208-1219, Nov. 1985.
- [15] J. L. Bentley, D. D. Sleator, R. E. Tarjan, and V. K. Wei, "A Locally adaptive data compression scheme," *ACM Commun.*, vol. 29, pp. 320-330, 1986.
- [16] M. Sayano, *Analysis of Coding and Compression Strategies for Data Storage and Transmission*, Ph.D. thesis, California Institute of Technology, 1992.
- [17] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, no. 1, 84-95, Jan. 1980.
- [18] C. I. Podilchuk, N. S. Jayant, and N. Farvardin, "Three-Dimensional Subband Coding of Video," *IEEE Trans. Image Processing*, vol. 4, no. 2, pp. 125-139, Feb. 1995.
- [19] H. Mery, H. G. Rosdolsky, and T. S. Huang, "Optimum run length codes," *IEEE Trans. Commun.*, vol. 22, no. 6, pp. 826-835, June 1974.

From 1982 to 1983, he was with the Department of Electrical Engineering, King Saud University. From 1984 to 1990, he was on scholarship by the same University for his graduate study. He is currently working as an Assistant Professor in the Department of Electrical Engineering, King Saud University, Riyadh. His current research activities are in signal and image processing including image coding, multi-rate signal processing for digital video processing, and HDTV.

Biography



Awad K. Al-Asmari (S'87-M'91) was born in Balasmar, Saudi Arabia, on Jan. 1, 1958. He received the B.S. degree from King Saud University, Riyadh, Saudi Arabia, in 1982, the M.S. degree from Ohio State University, Columbus, OH, in 1986, and Ph.D. degree in electrical engineering from the University of Toledo, Toledo, OH, in 1991. All in electrical engineering.

A Robust, Scalable, Object-Based Video Compression Technique for Very Low Bit-Rate Coding

Raj Talluri, *Member, IEEE*, Karen Oehler, *Member, IEEE*, Thomas Bannon, Jonathan D. Courtney, *Member, IEEE*, Anub Das, and Judy Liao

Abstract—This paper describes an object-based video coding scheme (OBVC) that was proposed by Texas Instruments to the emerging ISO MPEG-4 video compression standardization effort. This technique achieves efficient compression by separating coherently moving objects from stationary background and compactly representing their shape, motion, and the content. In addition to providing improved coding efficiency at very low bit rates, the technique provides the ability to selectively encode, decode, and manipulate individual objects in a video stream. This technique supports all three MPEG-4 functionalities tested in the November 1995 tests, namely, improved coding efficiency, error resilience, and content scalability. This paper also describes the error protection and concealment schemes that enable robust transmission of compressed video over noisy communication channels such as analog phone lines and wireless links. The noise introduced by the communication channel is characterized by both burst errors and random bit errors. Applications of this object-based video coding technology include videoconferencing, video telephony, desktop multimedia, and surveillance video.

Index Terms—Error Resilience, MPEG, MPEG-4, standards, video compression, videoconferencing, wavelets, wireless video.

I. INTRODUCTION

VIDEO is an integral part of multimedia; video conferencing, video-on-demand, broadcast digital video, and high-definition television (HDTV) are expected to have a substantial consumer market. These applications of digital video use video compression techniques to make efficient use of the available bandwidth. Existing video coding techniques such as those used in MPEG-1 (ISO), MPEG-2 (ISO), H.261 (ITU), and H.263 (ITU) standards employ schemes that use intraframe coding or interframe coding techniques to achieve high compression rates. Intraframe coding operates on each frame of the sequence separately, achieving compression by utilizing the spatial correlation within the frame. In contrast, interframe coding exploits temporal correlation between frames, thereby achieving even higher compression ratios.

Manuscript received April 20, 1996; revised August 20, 1996. This paper was recommended by Guest Editors Y.-Q. Zhang, F. Pereira, T. Sikora, and C. Reader.

R. Talluri, K. Oehler, T. Bannon, and J. D. Courtney are with Corporate Research, Texas Instruments, Dallas, TX 75265 USA.

A. Das was with Corporate Research, Texas Instruments, Dallas, TX 75265 USA. He is now with the University of Maryland, College Park, MD USA.

J. Liao was with Corporate Research, Texas Instruments, Dallas, TX 75265 USA. She is now with Rockwell Semiconductor Systems, Newport Beach, CA USA.

Publisher Item Identifier S 1051-8215/97/00879-3.

The existing video coding standards result in a number of unacceptable artifacts such as blockiness and unnatural object motion when operated at very low bit rates. Since these techniques use only the statistical dependencies in the signal at a block level and do not consider the semantic content of the video, at very low bit rates (high quantization factors) artifacts are introduced at the block boundaries. Usually these block boundaries do not correspond to physical boundaries of the moving objects and hence, visually annoying artifacts are introduced. Unnatural motion arises when the limited bandwidth forces the frame rate to fall below that required for smooth motion. Hence, there is a need for new standards that have improved coding efficiency and produce acceptable quality video at very low bit rates.

Most very low bit rate video coding standards use predictive coding techniques such as block-motion compensation, which is used to remove the temporal redundancies in the video signal. In addition, these techniques also use entropy coding techniques such as Huffman codes to achieve high compression. These entropy coding techniques typically use variable length codes. When the compressed video data is transmitted over existing communication channels such as analog phone lines or wireless channels, it is susceptible to the errors introduced by the channel degradations. Unless suitably dealt with, these errors result in loss of synchronization between the encoder and decoder and also result in a severe degradation in picture quality. The variable length codes and the predictive coding techniques used in the video compression schemes are particularly vulnerable to bit errors. Hence, the errors quickly propagate across the video sequence, rapidly rendering the video data unusable. Fig. 1(b) shows one frame of a video sequence decoded by the H.263 in the presence of random bit error (BER 10^{-3}) and burst error patterns used in the MPEG-4 robustness tests. Fig. 1(a) shows the same frame without the errors. As this figure illustrates, the quality of the sequence degrades rapidly to completely unusable levels. In most real-time video communication applications, retransmission is not possible due to the delay introduced by the retransmission mechanism. Hence, there is a definite need for new video coding standards that are much more robust to channel degradations and provide acceptable quality video even under adverse channel conditions.

Another drawback of the existing video coding standards is that they have very poor scalability properties. Scalability can

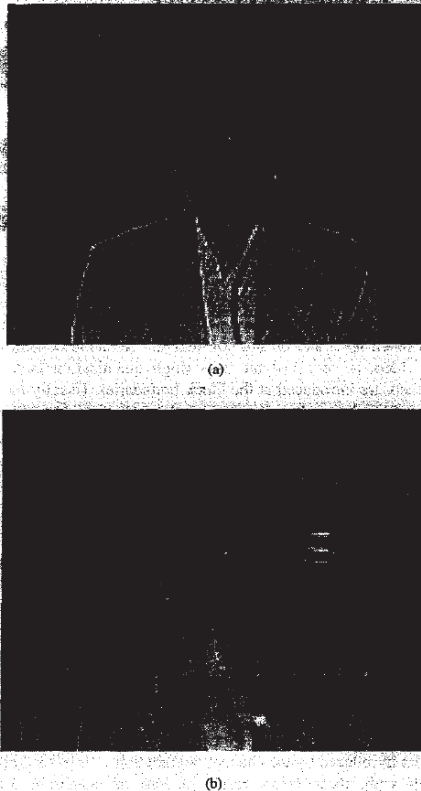


Fig. 1. The quality of a decoded H.263 video sequence degrades substantially in the presence of channel errors (a) decoded without errors and (b) decoded with random bit errors and burst errors.

be defined as one embedded bit stream with many available rates. One of the advantages of a scalably encoded bit stream is that the encoder can make optimal use of the varying channel bandwidth by extracting a subset of the bitstream to match the available channel bandwidth. In a multicast environment, scalability is a particularly useful property. If the different receivers are connected by communication channels of widely varying channel bandwidths, the encoder can achieve efficient transmission from the same coded bit stream without the need for re-encoding at different rates. Hence, another functionality that the next generation video coding standards need to support is scalability.

Current video coding standards such as MPEG-1, MPEG-2, H.261, and H.263 deal with video exclusively at a frame level and therefore do not have the ability to encode, decode, and manipulate individual objects within a video bitstream.

For example, in a videoconferencing system operating over a limited bandwidth channel, the ability to spend more bits on a particular object of interest such as person talking and less on the other parts of the scene can optimize the utility of the system. This capability can provide important functionalities in many interactive applications including multimedia databases, videoconferencing, and video authoring tools. Consider a video sequence that is coded as individual objects based on content and stored in a multimedia database. The user can then interactively query and retrieve individual objects from this compressed scheme instead of retrieving the entire frames of the video. This content manipulation is another functionality that the future video coding standards need to support.

A. MPEG-4

The Moving Pictures Expert Group (MPEG), which is a part of the International Organization for Standardization (ISO), is in the process of creating a new audio-visual coding standard called MPEG-4. MPEG-4 supports new ways (notably content-based) for communication, access, and manipulation of digital audio-visual data [1]. MPEG-4 will offer a flexible framework and an open set of tools supporting a range of both novel and conventional functionalities [2]. In addition to providing improved coding efficiency at lower bit rates, MPEG-4 is also expected to provide a number of new and useful functionalities that are not easily supported by the existing compression standards. These functionalities include improved coding efficiency, scalability, robustness in error prone environments, content-based manipulation, multimedia data access tools, ability to encode both graphics and video, and improved random access [3].

The structure of MPEG-4 will be composed of four different elements: syntax, tools, algorithms, and profiles. The syntax is an extensible description language that allows for selection, description, and downloading of tools, algorithms and profiles [4], [5]. A tool is a technique (e.g., contour representation, motion compensation, etc.) that is accessible via the syntax. A algorithm is an organized collection of tools that provides one or more functionalities. A profile is a collection of algorithms constrained in a specific way to address a specific application. Hence, MPEG-4 is now defined to offer a flexible syntax and an open set of tools supporting a range of both novel and conventional functionalities that extend beyond just better coding efficiency at very low bit rates. MPEG-4 is also developing a flexible and extensible syntactic object-oriented description language called MPEG-4 Systems and Description Language (MSDL) which will not only allow for the description of the bitstream structure but also configuration and programming of the decoder. MSDL will support the varied set of coding techniques that realize the new functionalities MPEG-4 is planning to support.

MPEG-4 issued a first call for proposals for tools and algorithms addressing a subset of these functionalities in November 1994. The proposals were due in October 1995 and the subjective evaluation of these proposals took place in the first week of November 1995. In the first round of tests, proposals were solicited for three representative functionalities: 1)

improved coding efficiency, 2) content-based scalability, and 3) robustness to error-prone environments. A number of audio and video test sequences were selected for testing the proposed techniques [6]. The test sequences were classified with respect to the complexity of the scene content, and a range of bit rates from 10 Kb/s to 1 Mb/s was used to code the sequences. The standard is expected to reach the Committee Draft stage by November 1997 and become an International Standard by November 1998.

B. Object-Based Coding Versus Block-Based Coding

There has been considerable research in object-based coding for very low bit rate video compression [7]–[10]. These techniques achieve efficient compression by separating coherently moving objects from stationary background and compactly representing their shape, motion, and the content. Some of the advantages of these approaches include a more natural and accurate rendition of the moving object motion and efficient utilization of the available bit rate by focusing on the moving objects. In addition to these compression advantages, if suitably constructed, the object-based coding techniques can also support content-based functionalities such as the ability to selectively code, decode, and manipulate specific objects in a video stream. The ability to scale the bit stream at an object level can also be supported by these techniques. To elaborate, the decoder can use only part of the original coded bit stream to selectively decode a subset of the coded objects at varying quality and resolution for each individual object. These functionalities can prove to be useful in some PC multimedia authoring tools and real-time video conferencing systems operating over low bandwidth channels. In light of these advantages, it is expected that the object-based coders will contribute to the MPEG-4 standard.

However, traditionally most of the object-based coding techniques suffer from: 1) the considerable overhead for accurately representing the shape of the moving objects results in insufficient number of bits for coding the content, and 2) the segmentation and motion estimation techniques can prove to be computationally very expensive. The more recent block-based video compression schemes such as the ITU H.263 [11] do not suffer from the same drawbacks as the object-based coders. By adopting a fixed block-based partitioning of the image, they avoid the need for shape coding. Some of the newer block motion compensation techniques such as overlapped block-motion compensation (OBMC) also help reduce the prediction error significantly. These techniques provide computationally inexpensive and efficient techniques for very low bit rate video communications. However, these new generation block-based coders suffer from: 1) the use of fixed block partitioning on a fixed grid, resulting in blocking artifacts and unnatural object motion from at very low bit rates, 2) inability to support any of the content-based functionalities such as object scalability, and 3) the optimized syntax of these coders for improved coding efficiency, making them highly susceptible to channel noise.

In this paper we present a hybrid object-based video coding approach that retains the relative advantages of both the object-based and block-based coders while minimizing the drawbacks

of both. By employing motion segmentation techniques to separate moving objects from stationary backgrounds, the coder optimizes the bit allocation to those areas that are changing most frequently. By employing an efficient and modular syntax to represent the shape, motion, and the residuals for each segmented object, the coding scheme enables us to maintain the high coding efficiency at very low bit rates with reduced artifacts. This technique also provides the ability to selectively encode, decode, and manipulate individual objects in a video stream and, hence, supports content-based functionalities such as object scalability and object manipulation easily. By using an efficient shape representation that is well matched with the residual coding stage, the coding scheme alleviates the drawbacks of traditional object-based coders thus offering the best of both the object-based coders and the block-based coders.

We also incorporated efficient error concealment and error correction strategies into the overall coding scheme to enable robust transmission of the compressed video data over noisy communication channels, e.g., wireless channels [12]. The error correction techniques add a controlled amount of redundancy to the data to enable detection and correction of the bit errors. The error concealment techniques minimize the degradation of the picture quality by properly utilizing the correctly decoded data in masking the effects of residual errors (errors remaining after correction). The object-oriented structure of the coder and the associated syntax enables us to contain the effects of errors within the boundaries of each coded object and hence stop the spreading of the errors across the image sequence.

This hybrid coding scheme was proposed by Texas Instruments as a candidate for the evolving ISO MPEG-4 audio and video coding standard. This proposal performed impressively in all the three categories of MPEG-4 subjective tests held in November 1995; namely 1) compression, 2) error robustness, and 3) scalability. Our coder competed in tests using two classes of sequences—Class A and Class B. Class A sequences are characterized by those having low motion and low spatial detail—typical of a videophone-type application. Class B sequences are characterized by those having either medium motion or medium spatial detail. Each class of sequences is coded at three different bit rates. The Class A sequences are coded at 10 Kb/s, 24 Kb/s, and 48 Kb/s. The Class B sequences are coded at 24 Kb/s, 48 Kb/s, and 112 Kb/s. There are four sequences in each class. Akiyo, Hall Monitor, Container, and Mother Daughter in Class A and Coastguard, News, Foreman, and Silent in Class B. Each sequence is of 10 s duration (300 frames).

In the compression tests our coder was rated to be in fifth place (out of 18 proposals) for Class A sequences and sixth place (out of 13) for Class B sequences. In the error resilience and recovery tests, our coder is rated to be in second place (out of nine proposals) for Class A sequences and first place (out of seven proposals) in for Class B sequences. In the object scalability tests, our coder is rated to be in fourth place (out of nine proposals) [13]. It outperformed all other object-based coders in the compression and robustness tests and was the only coder that supported all three functionalities.

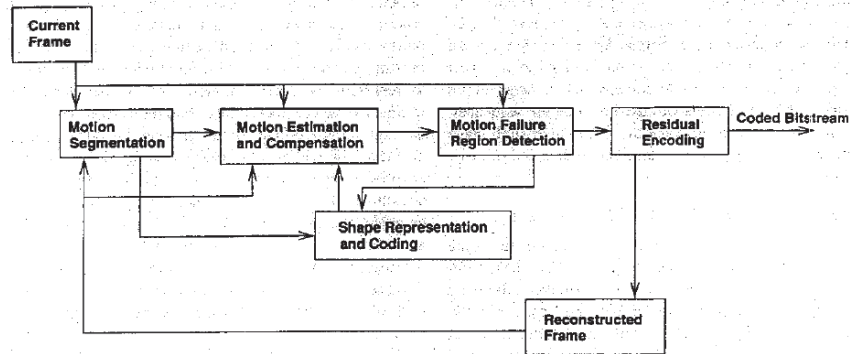


Fig. 2. The overall block diagram of the hybrid object-based coder.

The rest of the paper is organized as follows: Section II describes the overall codec architecture and describes the various stages of processing in coding a video sequence. Section III describes the error correction and concealment strategies developed that enable robust transmission of the compressed bit stream over noisy channels. Section IV describes the techniques employed in building a content scalable bit stream that enables object manipulation. Section V summarizes the techniques presented in this paper and points out future directions of research.

II. CODEC ARCHITECTURE

The object-based approach to video compression presented in this paper operates by estimating moving objects and coding compact representations of the object's shape, motion, and the content [14], [15]. This approach has the advantages of enabling content scalability and error robustness while providing efficient video compression for low bit rates (e.g., 10 Kb/s to 112 Kb/s) and potential quality enhancements at higher bit rates. Fig. 2 shows the overall block diagram of the object-based video coding (OBVC) scheme. The first frame of the video sequence is coded as an intraframe using the block-based discrete cosine transform (DCT) coding scheme. Subsequent frames are partially coded in the following modular five-step process:

- 1) moving object detection;
- 2) shape representation;
- 3) motion estimation;
- 4) motion failure region detection;
- 5) residual signal encoding.

The modularity of this framework facilitates future enhancements as improved techniques are devised. Our methods for error robustness and content scalability were implemented within this framework and are described below following a characterization of the five main steps. We illustrate these steps with examples for the "Akiyo" video data used in MPEG testing, from which two frames are reproduced in Fig. 3.

1) *Moving Object Detection*: The goal of the first step is to segment the images in the video sequence into moving objects and stationary background. Whereas block-based coding approaches arbitrarily divide the image into a regular set of square regions, object-based schemes ideally divide the image along object boundaries. However, it is not necessary to produce a precise segmentation of the objects to achieve gains in coding efficiency. Many techniques are relevant to moving object detection and there are many tradeoffs to consider.

In our scheme, a segmentation mask is generated for the current image using image differencing, morphology, and connected components analysis. Thus, the segmentation mask C_n is computed as

$$C_n = \text{ccomps}((T_h \bullet k) \circ k).$$

Here, T_h is the binary image resulting from thresholding the absolute difference of the current and previous images at h , \bullet and \circ are the morphological close and open operators, respectively, with structuring element k , and $\text{ccomps}(\cdot)$ is the connected components analysis function [16].

The image T_h is defined as

$$T_h(i, j) = \begin{cases} 1, & \text{if } |I_n(i, j) - I_{n-1}(i, j)| \geq h \\ 0, & \text{otherwise} \end{cases}$$

for all pixels (i, j) in T_h . The operation $a \bullet k$ is defined as

$$a \bullet k = (a \oplus k) \ominus k$$

where \oplus is the morphological dilation operator and \ominus is the morphological erosion operator [17]. Similarly, the operation $a \circ k$ is defined as

$$a \circ k = (a \ominus k) \oplus k.$$

In our scheme, the structuring element k is a circular kernel, 11 pixels in diameter.

Fig. 4(a) shows examples of image regions detected as moving objects for two frames of the "Akiyo" video. This segmentation does not exactly follow the boundary between

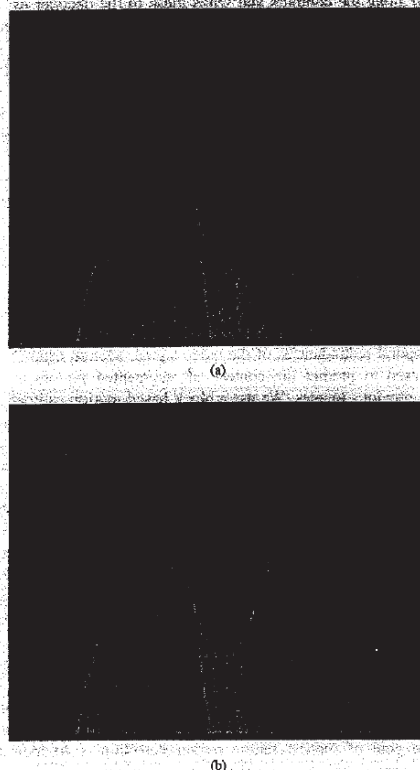


Fig. 3. (a) First frame and (b) a subsequent frame of "Akiyo" MPEG video test data.

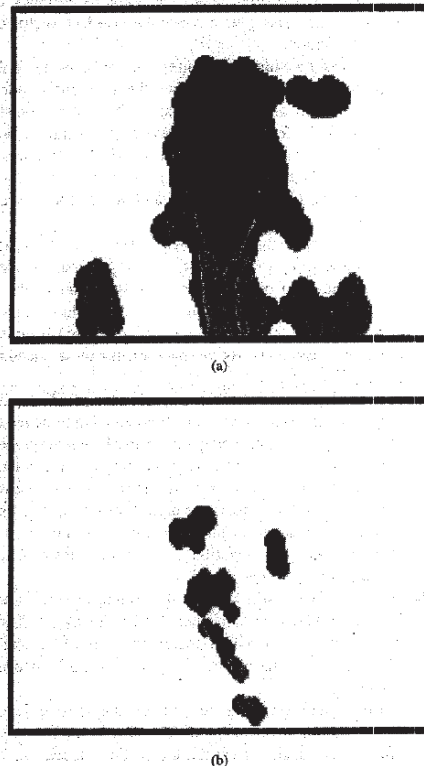


Fig. 4. (a) Image regions produced by moving object detection and (b) motion failure regions remaining after motion compensation.

the foreground and background, but it is still a useful approximation. If the entire scene is moving or if the camera is panning, a single large region may be detected; in this case compression will still succeed as a result of subsequent steps. Alternatively, other moving object detection techniques (e.g., [18]) and global motion compensation techniques (e.g., [19]) may be incorporated to enable better performance for moving camera situations, but the advantages might be offset for some applications by considerations such as computational expense.

2) *Shape Representation*: The task of shape representation is to efficiently code the bounding contours of image regions that were produced in the moving object detection step. We experimented with two shape representation strategies.

The first strategy calculates a bounding rectangle for each of the moving object regions and tiles the rectangles with 16×16 -pixel macroblocks. A bitmap mask of the region is formed which represents macroblocks that lie inside the region with a 1 b and blocks outside the region with a 0 b. The algorithm then finds the most efficient representation for

the mask from five possible coding formats. Such an adaptive approach is more efficient in terms of bits-per-object than using an identical shape coding format for every object.

This block-based approach is based on the observation that the shape masks typically exhibit significant similarity between successive coded video frames. By keeping a record of the shape masks in the previous coded frame, the algorithm exploits the temporal similarity of the shape masks to reduce the number of bits needed to represent a given shape in the current frame.

To efficiently code the shape mask, the algorithm performs the following steps.

- 1) It first searches the record of previous masks for a mask of identical shape and position. If such a mask is found, the algorithm codes the shape as simply an index to the matching mask.
- 2) If no such mask is found, the algorithm then searches for a previously-coded mask of identical shape but different position than the current mask. If such a mask

is found, the algorithm codes the shape as an index to the matching mask plus a row/column offset vector of pixel resolution.

- 3) If no mask meeting the requirements of steps 1) or 2) is found, the algorithm searches the previously-coded masks for the one most similar to the current mask. This mask is found by correlating the current mask with each previously-coded mask. The most similar mask will exhibit the fewest number of nonoverlapping macroblocks (either missing or additional) for a given row/column offset.

The algorithm then codes the current mask as a series of modifications to the most similar previously-coded mask. Thus, the object shape is coded as an index to the most similar previous mask, a row/column offset vector, a list of the macroblocks to be dropped from the previous mask, and a list of macroblocks to be added to the previous mask.

In the best case, this shape representation will require only a single macroblock to be dropped from or added to the most similar previously-coded mask. However, the worst case may require a large number of macroblock changes and this will thus be an inefficient shape representation. For this reason, the algorithm attempts to find a more efficient representation via steps 4) and 5). If a more efficient representation is found, the current shape coding format is abandoned.

- 4) If no previous masks exist or the coding of step 3) has been abandoned, the algorithm codes the object shape as a row/column position vector, the width and height of the shape in macroblocks, and a run-length encoded shape mask.

Run-length encoding will not be the most efficient representation if the shape to be coded is small or contains many transitions per mask row. For this reason, the algorithm attempts to find a more efficient representation using step 5).

- 5) If step 4) has been abandoned, the algorithm codes the object shape as a row/column position vector, the width and height of the shape in macroblocks, and the bitmap of the shape mask.

Thus, this algorithm chooses the most efficient of the following five coded-shape formats:

- 1) previous mask index;
- 2) previous mask index and offset vector;
- 3) previous mask index, offset vector, and macroblock changes (dropped and added);
- 4) position vector and run-length encoded shape mask;
- 5) position vector and shape mask bitmap.

In transmitting a coded shape, the encoder first sends a 3-b tag identifying the format of the shape data to follow. The decoder reads this tag and decodes the shape data appropriately. Furthermore, the encoder and decoder both keep a record of the shape masks in the previous frame in order to decode shape formats 1, 2, and 3.

We also experimented with a spline-based shape representation strategy. In this approach the encoder detects corner points

[20] along the contours, and sends these to the decoder using a differential coding technique. The coder and decoder both approximate the contour by fitting Catmull-Rom splines to the corner points, so that the coder and decoder agree precisely.

We found that the spline-based shape representation is more accurate but consumes more bandwidth than the block-based representation. Hence, a tradeoff can be made between accuracy of representation and bandwidth, depending on the application. In general, if the automatic motion segmentation strategy does not result in pixel-accurate boundaries of the moving object, there is no need to apply an accurate shape representation such as Catmull-Rom splines and a block-level representation will suffice. However, if the segmentation is performed off-line using more precise techniques such as blue-screening, this representation can prove to be very useful.

3) *Motion Estimation*: At this point, the goal is to remove temporal redundancies of the video signal. Motion estimation is used to predict the contents of segmented regions from the previous images. We use a block-based motion estimation technique. For each 16×16 block that lies inside a moving object, the algorithm searches to find a corresponding location in the previous frame such that the sum of absolute differences between the current and previous blocks is a minimum. To obtain a good match, the search is conducted at half-pixel resolution, similar to the motion estimation strategy of H.263. The motion vector corresponding to the location that produces the minimum error is then efficiently encoded. We also incorporated the advanced prediction mode using optional 8×8 or 16×16 motion vectors and the overlapped motion compensation technique from H.263 [21].

4) *Motion Failure Region Detection*: After motion estimation, a significant amount of residual information remains in some areas of the image where motion compensation alone was inadequate, e.g., as may occur in areas of rapid change. The goal of motion failure region detection is to identify these areas. This is accomplished by performing a second-pass motion segmentation between the compensated image and the current image, which is essentially the same procedure used in the moving object detection step. Examples of motion failure regions detected in the "Akiyo" video are shown in Fig. 4(b). The regions isolated in this step are marked as areas of motion failure, and their shapes are encoded in combination with the previous shape representations.

5) *Residual Signal Encoding*: The goal of this final step is to encode the residual information in the motion failure regions. The decoder uses this information (i.e., differences between the current and the compensated images) to correct for errors of motion estimation. Our coder has the ability to integrate two different kinds of waveform coding techniques: 1) DCT and 2) wavelets. Either technique can be selected during encoding. This demonstrates the modular nature of the object-based video coding scheme. The DCT and wavelet approaches described here are also used for INTRA coding, where the entire frame was coded rather than the motion failure region.

In the DCT approach, the pixels in the motion failure region are first transformed using 8×8 DCT. The DCT coefficients are scalar quantized, run-length encoded in a zigzag fashion, and then Huffman coded (similar to H.263).

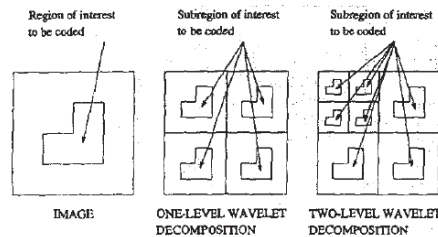


Fig. 5. Subregions of wavelet coefficient to be coded corresponding to the motion failure region in motion-compensated residual image.

To take advantage of only coding the segmented regions in the OBVC system, a region-based wavelet coding technique is proposed [22]. The wavelet coding approach consists of a wavelet decomposition of the entire error image (three-level decomposition for the chrominance, four-level decomposition for luminance) followed by zerotree quantization and arithmetic coding. Instead of sending wavelet information for the entire frame, we code only those wavelet coefficients which correspond to selected regions (i.e., motion failure regions). A regions-of-interest map specifies which 16×16 macroblocks contain pixels to be residual encoded. Due to the multiresolution nature of the wavelet transform, these subregions of coefficients can be obtained by downsampling the map at each decomposition level, as shown in Fig. 5. This eliminates the bits needed to encode wavelet information corresponding to pixel locations outside of the regions of interest (where the coefficients are generally insignificant).

In our implementation, a (6,2) tap biorthogonal filter pair is used for the decomposition. This filter (referred to as filter #5 in [23]) can be implemented with integer coefficients using a scaling factor. The embedded zerotree quantization and arithmetic coding introduced by Shapiro [24] is adapted to code the coefficients within the regions of interest. The zerotree procedure consists of repeated dominant and subordinate passes, each relative to a particular quantization threshold which decreases for each dominant pass. This repeated encoding process will be terminated when either 1) a predetermined bit budget for the frame has been reached or 2) the stopping quantization threshold has been reached. We obtained the best results by keeping the stopping quantization threshold (threshold of the last dominant pass) constant for all residual coding. This keeps the encoding quality fairly constant between frames. This is similar to using a fixed Q or quantization factor in DCT-based schemes. Image areas outside of the regions of interest are set to zero before decomposition and after reconstruction to reduce any smearing effects from the noncoded regions. Informal subjective tests have shown that the wavelet coding approach is preferred for the lower range of bit rates (e.g., 10 kb/s) while DCT is preferred for the higher bit rate range.

Tables I and II illustrate the coding efficiency of our coder. These are the results as submitted to the first round of MPEG-4 tests. As mentioned before, the coder is tested on two kinds of

TABLE I
CODING RESULTS FOR CLASS A SEQUENCES

Sequence Name	Bitrate Kbps	Frame Rate (fps)	Resolution	PSNR dB			Type of Coding
				Y	Cb	Cr	
Akiyo	10	5	QCIF	32.3	35.4	38.4	WAV
Akiyo	24	7.5	CIF	33.2	37.7	40.0	WAV
Akiyo	48	10	CIF	34.0	40.9	43.4	DCT
Container	10	4.3	QCIF	29.5	35.9	35.7	WAV
Container	24	7.5	QCIF	31.6	37.2	37.2	WAV
Container	48	10	CIF	29.7	37.4	37.7	DCT
Hall Mon.	10	4.3	QCIF	29.7	36.5	39.0	WAV
Hall Mon.	24	10	QCIF	32.0	36.5	39.4	WAV
Hall Mon.	48	10	CIF	32.5	37.9	40.3	DCT
Moth. Dtr.	10	3.8	QCIF	31.7	36.9	38.5	WAV
Moth. Dtr.	24	10	CIF	33.8	39.2	40.5	DCT
Moth. Dtr.	48	10	CIF	34.9	39.4	41.2	DCT

TABLE II
CODING RESULTS FOR CLASS B SEQUENCES

Sequence Name	Bitrate Kbps	Frame Rate (fps)	Resolution	PSNR dB			Type of Coding
				Y	Cb	Cr	
Coastguard	24	6	QCIF	26.3	36.4	40.1	WAV
Coastguard	48	10	QCIF	27.9	38.2	40.7	DCT
Coastguard	112	7.5	CIF	27.4	37.2	41.1	DCT
News	24	5	QCIF	29.5	35.9	35.2	WAV
News	48	10	QCIF	31.7	35.9	35.8	WAV
News	112	10	CIF	33.7	35.3	33.5	DCT
Silent	24	7.5	QCIF	29.2	36.0	37.5	DCT
Silent	48	7.5	QCIF	32.7	38.0	39.2	WAV
Silent	112	10	CIF	32.3	38.7	39.4	WAV
Foreman	24	4.3	QCIF	26.3	34.9	34.7	DCT
Foreman	48	6	QCIF	28.5	36.4	36.5	DCT
Foreman	112	10	QCIF	30.8	38.2	38.3	DCT

sequences—Class A and Class B and each class of sequences is coded at three different bit rates. The table also illustrates which sequences are coded using wavelets (WAV) and which are coded using DCT's.

III. ERROR RESILIENCE

This section describes the error protection and concealment techniques that enable robust transmission of the compressed bit stream over noisy communication channels such as analog phone lines and wireless links. The noise introduced by the communication channel is characterized by both burst errors and random bit errors. In the past few years, various approaches have been proposed to combat the cell-loss problem in DCT-based as well as other coding systems [25]. One popular method is to exploit the insensitivity of the human visual system to the high-frequency component of video signals by using layered coding [26]–[28]. With such schemes, the low-frequency coefficients and other more important information (such as coding modes and motion vectors) are transmitted with a higher level of priority and protection than for the less important high-frequency components. To limit error propagation in coders using motion-compensated prediction, studies have also been made on how to perform motion compensation only over the low-frequency components, such that discarding the high-frequency components does not cause error propagation [29]. Leaky prediction which reduces the

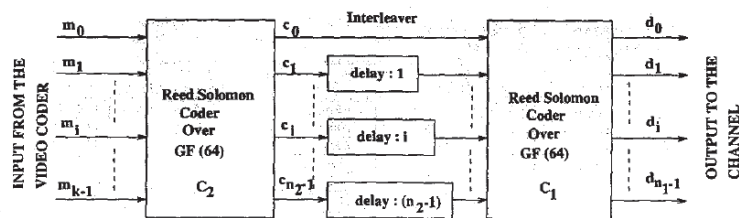


Fig. 6. The coder architecture for error protection.

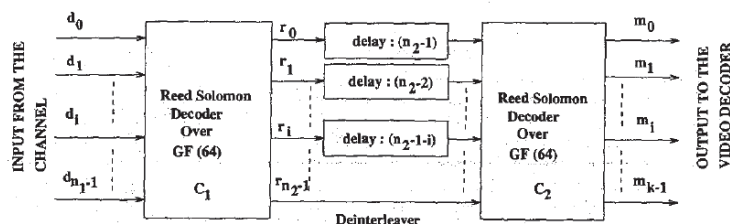


Fig. 7. The decoder architecture for error protection.

effect of previously corrupted frames to the current frame by attenuating the prediction results by a leakage factor has also been investigated [28]. With these hierarchical transmission schemes, the low-priority, high-frequency components can be simply replaced by zeros if they are lost. The resulting image, though blurred, is still acceptable. These techniques are well suited in asynchronous transfer mode (ATM) environments, where it is possible to have different levels of priority for the packets. However, this may not be possible over channels such as existing analog phone lines or wireless networks where it may be difficult, if not impossible, to prioritize the data being transmitted. Hence, it becomes necessary to resort to error protection and error concealment techniques to mitigate the effects of channel errors.

A. Error Protection

In order to protect the compressed video data from channel errors, we use an error control code to correct and detect errors. This code introduces a controlled amount of redundancy that enables error detection and correction. The scheme comprises two Reed-Solomon (RS) coders with an interleaver in between. Most block codes (like RS codes) are random error correcting codes and are not capable of correcting burst errors. The interleaver is used to combat the effect of burst errors in the channel. It mixes up the symbols from several code words so that the symbols from any given code word are well separated during transmission. When the code words are reconstructed by the deinterleaver, error bursts introduced by the channel are broken up and spread across several

code words. The interleaver/deinterleaver pair thus creates an effectively random channel. We used a crossinterleaver in our design of the error control code. The architecture of the error control scheme is similar to that used in the audio CD [30].

The RS coders used are over the GF(64) (Galois Field). The compressed video data is input to the RS coder C_2 which maps k , m -bit ($m = 6$, in the case of GF(64)) input symbols (words): $\{m_0, m_1, \dots, m_{k-1}\}$ onto n_2 , m -bit words: $\{c_0, c_1, \dots, c_{n_2-1}\}$. The output of C_2 is input to the cross interleaver before being encoded by the C_1 RS coder. The cross interleaver introduces a delay of $(i-1)D$ words for the code word c_i (we set $D = 1$). C_1 encodes n_2 , m -bit words into n_1 output symbols: $\{d_0, d_1, \dots, d_{n_1-1}\}$ which are then transmitted over the channel. Fig. 6 shows a schematic diagram of the coder architecture. The decoder is a mirror image of the coder and comprises an RS decoder C_1 followed by a cross deinterleaver and a C_2 decoder. Fig. 7 shows the decoder architecture.

We used the above error-correction scheme for 2 bit rates of video traffic: 24 Kb/s and 48 Kb/s. In the former case, we chose $(k, n_2, n_1) = (27, 31, 34)$. Note that for this design choice the average delay introduced by the interleaver is less than 250 ms. In this case, we choose to have decoder C_1 correct all single-error patterns. When the C_1 decoder sees a higher-weight error pattern (double-error patterns and any pattern causing a decoder failure, i.e., a detectable error), the decoder outputs 31 erased symbols. The crossdeinterleaver spreads these erasures over 27 C_2 code words. The C_2 decoder can correct any combination of e errors and s erasures, where $2e + s \leq (n_2 - k) = 4$. However, we chose to decode

erasures only owing to the small probability of a C_1 decoder error (error which is not detectable). Whenever the number of erasures is greater than $(n_2 - k)$, it outputs k symbols with the erasures substituted by an arbitrary GF(64) word. This technique performs better in the decoding of the video data than the case when the C_2 decoder outputs k erased symbols.

In the second case we choose $(k, n_2, n_1) = (38, 43, 48)$. Once again, the average delay introduced by the crossinterleaver is less than 250 ms. The decoder is similar to the one described earlier except for the C_1 decoder. In this case, the C_1 decoder is chosen to correct all single and double-error patterns and detect all other higher weight error patterns which result in a decoder failure. The C_2 decoder is designed to correct s erasures, where $s \leq (n_2 - k)$.

The performance of this error-correcting technique can be further improved by a computationally inexpensive modification described below. We use *feedback* from the RS decoder C_1 to C_2 to improve the error-correcting capability of C_2 , which, in turn, improves the overall performance. More specifically, let us assume that the decoder C_2 corrects e_2 errors and detects t_2 errors where $e_2 + t_2 < d_2$ with d_2 being the minimum distance for the decoder C_2 . Let C_1 be capable of correcting s_2 erasures in any given block of n_2 symbols. Further assume that at time T , C_2 detects x errors in the input block of n_1 symbols (B , say) of n_1 m -bit symbols, where $x > e_2$, thereby leading to a decoding failure. Hence, C_2 outputs n_2 erased symbols as was mentioned earlier. We modify the decoder here by storing the n_1 input symbols to C_2 and the time instant at which the decoding failure occurred (namely, T) in a buffer M . The output of C_2 is used as input by the crossinterleaver which spreads out the erased symbols over the next n_1 blocks (the erased symbols from B appear as input to C_1 at times $T, T + D, T + 2D, \dots, T + (n_1 - 1)D$).

Consider the time instant $t = T + iD$, where $i = 0, 1, \dots, n_1 - 1$. If the number of erased symbols in the input block for decoder C_1 at time $t \leq s_2$ then we can correct all these erasures. One of these corrected erasures (y , say) is a consequence of the block erased by C_2 at time T . This corrected erasure y can be one of the following.

- 1) It is one of the symbols that were in error in the block B or
- 2) It was not one of the symbols that were in error.

It is easily determined which category y belongs to by comparing the content of the buffer M at the appropriate location (y' , say) with y . If $y = y'$, then condition 2) is true, in which case we continue without any further modification. If, however, $y \neq y'$, then condition 1) is true. This implies that we have, by correcting the erasure at decoder C_1 , successfully corrected one of the errors that occurred in the block B that was previously uncorrectable. If this is the j th such error that was corrected in the block B (by means of the erasure corrections at C_1 and the subsequent feedback to the buffer N), then we have reduced the number of errors in B to $x - j$. If $x - j \leq e_2$, then we are in a position to correct all the other errors in B . Upon correction of all the other errors we proceed to update the content of the crossinterleaver appropriately. This, in turn, reduces the number of erased symbols in the

input blocks to C_1 at subsequent times, thereby increasing the overall probability of error-correction. If, however, it is not possible to correct all the errors in B , we continue without any further modification in the decoding.

Hence, decoding with feedback, in the worst case situation, performs as well as a decoder without feedback. In general, the probability of an uncorrected channel error can reduce by up to a factor of two. Note that the overall codec delay may be increased due to the feedback.

B. Error Concealment

Although the above error protection scheme is effective in removing most of the errors introduced by the channel, it is not possible to remove all error patterns. Moreover, a better error protection scheme (that is, one with a lower probability of an uncorrectable error) can be designed only at the expense of an increased bit rate. Since the primary objective is compression, it is clearly not a pragmatic approach to search for such a protection scheme. Hence, it becomes necessary to resort to other techniques which will be able to detect these uncorrectable channel errors (which we will call *residual errors*) and prevent a serious degradation of picture quality. These *error concealment* techniques detect when a bit stream is in error, resynchronize, and use the correctly decoded data to minimize the effect of the erroneous data.

As was mentioned earlier, one of the most severe effects that can result from errors in compressed video data is that of the decoder losing synchronization with the encoder. To enable the decoder to regain synchronization, we introduce resynchronization (resynch) words into the bitstream. These resynch words are characterized by the fact that they are unique and no valid bit-sequence in the compressed video data can be the same as them. They are introduced at the end of each of the following for every detected moving object in a frame:

- 1) the shape data (the data that describes the shape of a moving object);
- 2) the motion vector data (the data describes the motion of the given object);
- 3) the DCT data (the data that encodes the differences, if any, between the motion compensated object and the actual object).

Once the decoder detects an error in the received bitstream, it tries to find the nearest resynch word and re-establishes synchronization with minimal loss of coded data. An error is detected at the decoder if:

- 1) an invalid codeword is found;
- 2) an invalid mode is found in the syntax;
- 3) the resynch word does not follow a decoded block of data;
- 4) a motion vector points outside the image;
- 5) a decoded DCT coefficient value lies outside permissible limits;
- 6) the shape information is found to be invalid.

Upon detection of an error and subsequent resynchronization, the decoder attempts to minimize the effect of the error on the picture quality. If an error is detected in the shape data,

then the moving object corresponding to this shape description is dropped and is made a part of the background. This exploits the temporal correlation present in the video sequence. If an error is detected in the motion vector data, then the mean motion vector for the object (which is also transmitted as part of the bitstream) is applied to the entire object. Here we have exploited the spatial correlation in a given frame. If the error detected is in the DCT data, then all the DCT coefficients for that object are set to zero and the decoder attempts to resynch at the earliest possible resynchronization point. This prevents serious picture quality degradation because the DCT coefficients are usually zero or very small in magnitude. These concealment strategies prove to be very effective in combating the residual errors and maintaining adequate picture quality. In summary, we find that in addition to providing good quality video at very low bit rates, the object-based video compression techniques also make graceful error concealment possible due to the object-oriented structuring of the coded bitstream and the associated syntax.

The MPEG-4 error robustness tests consisted of two kinds of tests: 1) resilience and 2) recovery. The resilience tests required the coded video bitstream to be subjected to three different kinds of error conditions: random errors, short burst errors, and a combination of both. The random errors are iid with the probability of a bit-error = $1.0e^{-3}$. The burst error conditions consist of three short bursts. The duration of each is uniformly distributed between 16 ms and 24 ms. It is further characterized by a probability of bit-error = 0.5. Moreover, any two bursts are a minimum of 2 s apart. It is also assumed that the first 1.5 s of the bitstream is transmitted without any channel error. The recovery tests consisted of subjecting the coded bitstream to a long burst of 2 s duration. These error conditions are used to typify wireless communication channels. The test data was divided into different classes based on the complexity of the test material. There are four video test sequences with low spatial detail and low motion (*Class A* sequences) and four video sequences of medium spatial detail and low motion or vice versa (*Class B* sequences). The *Class A* sequences are coded at 24 Kb/s including the overhead for protection. The *Class B* sequences are coded at 48 Kb/s. The duration of each sequence is 10 s. The proposals had to demonstrate the coder performance on all the sequences in a particular class. The evaluation stage consisted of blind subjective evaluation of all the competitive proposals in a *single stimulus* test [13].

Our proposal competed in both the resilience and recovery tests for *Class A* and *Class B* sequences. Our proposal was ranked as best in class in *Class B* sequences and second best in *Class A* sequences for both resilience and recovery tests. The error-protection scheme was able to correct most of the channel errors. The residual errors are detected by the concealment strategy and the loss of picture quality is minimized.

Fig. 8(a) shows a typical frame of one of the *Class A* test sequences called *Akiyo*. Fig. 8(b) shows the same frame upon decoding a corrupted bitstream. There are residual errors in this frame but the effect of these errors is *minimized* and localized to specific locations in the image. Fig. 9 shows the

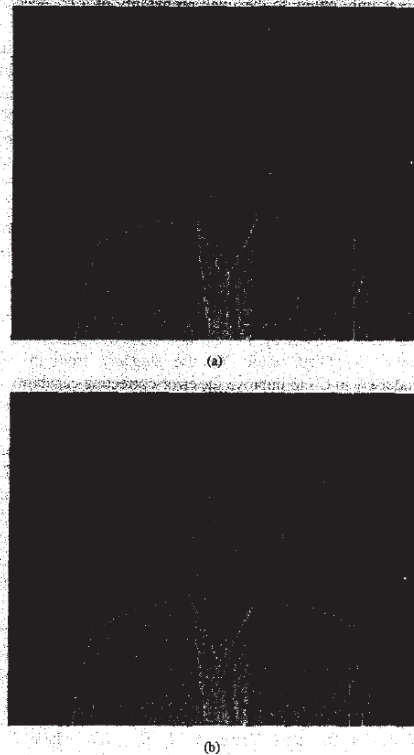


Fig. 8. The error protection and concealment techniques enable the object-based video coder to sustain the effects of channel errors (a) decoded without errors and (b) decoded with random bit errors and burst errors.

peak signal-to-noise ratio (PSNR) (Y) values of all the 100 coded frames (10 s of video coded at 10 frames/s) in this sequence with and without channel errors. Notice that the PSNR (Y) value drops in the middle of the sequence due to residual errors. However, the coding and concealment strategies enable the decoder to gracefully recover and maintain an acceptable quality, with a PSNR (Y) close to the original, during the rest of the sequences.

IV. CONTENT SCALABILITY

Content scalability refers to the property of a compressed video bitstream that can have a subset of the encoded information removed, for example all of the objects representing a particular person, and the remaining bitstream will still decode correctly. In the context of MPEG-4, *object scalability* refers to the ability of a coder to selectively encode/decode specific objects within a video sequence and achieve a lower bit rate when a lesser number of objects are encoded [6]. *Quality scalability* refers to the ability of the coder to selectively

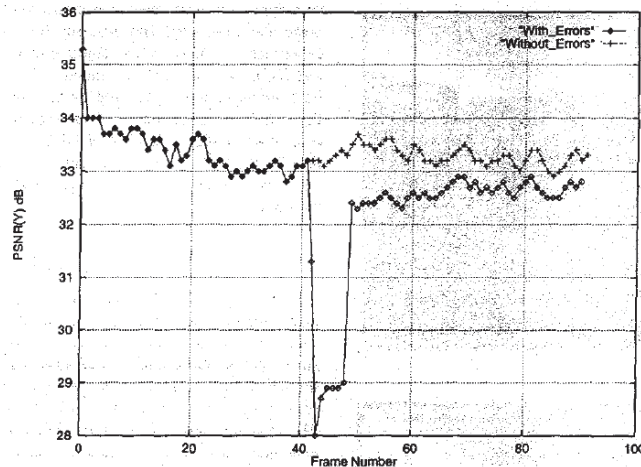


Fig. 9. PSNR (Y) values for the frames in the Akiyo sequence with and without errors.

enhance the quality of a particular object in the video sequence by spending more bits on that particular object. The quality of the object can be enhanced either by improving its spatial or its temporal resolution.

In order to construct a content scalable bit stream in the object-based video compression technique, the following steps are performed. The first frame of the sequence is coded as an intraframe, a frame composed solely of intra objects. Other intraframes may occur at regular or irregular intervals thereafter. Intra objects are objects coded only with reference to themselves in the current frame; they are not limited to intra frames. Each frame of the sequence is composed of background and foreground objects. The shape of the background object (the frame with all the foreground objects removed) is first coded using the chosen shape coding technique (block-based or spline-based). The background texture information is then encoded as DCT-encoded intra blocks. Next, for each foreground object in the frame, the shape and the texture are also coded similarly. As each object is put into the bitstream, it is preceded by a header of fixed length wherein the object number, object type (such as I object), and object length (in bits) are recorded. After all of the objects in the first frame have been coded, a reconstructed frame is made, combining decoded images of the background and each object into one frame. An average pixel value is calculated for each of the three color channels in the reconstructed frame and the three averages are written to the bitstream.

Following the intraframe, each subsequent frame of video is encoded as an interframe, until the next, if any, intraframe. As with intraframes, the shape information for each object is encoded first. Next, the block motion vectors are calculated for the object with respect to its reconstructed version in the previous frame. In order to do this, the reconstructed object

is put into an empty frame and its nonmask area filled with the average pixel values calculated in the last intraframe. This isolates the motion compensation to a per object basis. The average pixels smooth out the object/nonobject boundary so that the block matching done during motion estimation will perform better on the border blocks than if the object was surrounded with zero-valued (or similarly contrasting) pixels. The estimated motion vectors are then applied to create a motion-compensated object. The motion vectors are efficiently encoded and put into the bitstream. Next, the differences between the motion-compensated object and the current object are DCT encoded as INTER and INTRA blocks as usual and are put into the bitstream. This step of ensuring that the motion vectors of the currently encoded object only point to its previous instantiation is a key step in achieving a scalable bitstream.

Finally, uncovered background is handled. The uncovered background created by the motion of an object is calculated and coded as a separate object in the bitstream. To calculate the uncovered background, the object's original (not the reconstructed) shape mask is differenced with its previous shape mask to isolate the pixels that are in the current shape and not in the previous shape. This uncovered background shape is then coded coarsely using a block-based representation. The texture values in the uncovered background image are DCT encoded as INTRA blocks (making the uncovered background objects intra objects). This separate treatment of the uncovered background (along with the per object motion compensation) is what makes the bitstream scalable for video objects. The bitstream can be played as created; an object and its uncovered background can be removed to excise the object from the playback; or an object can be extracted to play on its own or be added to a different bitstream. On the decoding side, the

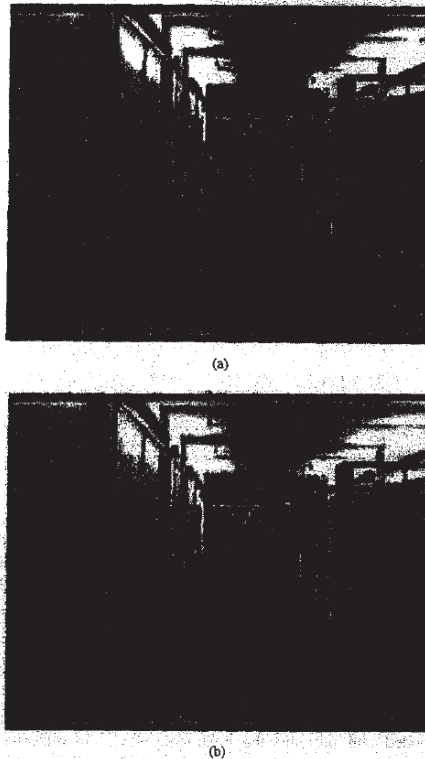


Fig. 10. The content scalability functionality of object-based video compression supports selective encoding and decoding of objects: (a) frame of the "Hallway" MPEG video test data and (b) the same frame with the image of the person on the left removed.

scalable OBVC decoder works the same as the non-scalable version except that it decodes an object at a time instead of a frame at a time. When dropping objects, the decoder merely reads the object header to find out how many bits long it is, reads that many bits, and throws them away.

In order to achieve quality scalability, we define *enhancement objects* for those objects whose quality we need to improve. These objects, whose loss would not remove the object from the scene, just lower the quality of its visual appearance or omit audio or other data linked to the object.

In our coding scheme, the accuracy of content scalability depends on how accurately the moving object detection step segments objects from the background. Precise segmentation is important for content scalability applications such as multimedia video editing. Automatic moving object detection methods are unlikely to produce accurate segmentations, so interactive segmentation is needed to satisfy this requirement for content scalability. The 1995 MPEG-4 tests [6] provided

an accurate set of segmentation masks to idealize the moving object detection step. The bitstream for each object in these experiments comprised the object contour, motion vectors, and DCT data, as well as contours and DCT data for the background regions uncovered by the objects as they move. Fig. 10 shows an example of an experiment in which this content scalability capability of OBVC was applied to send data for the background and a moving object to the decoder, while data for another moving object was excluded from the bit stream.

V. CONCLUSIONS

In this paper we described a hybrid video coding approach that combines the advantages of both the traditional block-based coders and object-based coders. By using an efficient shape representation that is well matched to the application, the scheme achieves improved coding efficiency and provides compressed video with reduced artifacts at very low bit rates. The modular nature of the codec architecture enabled us to experiment with both DCT and wavelet coding techniques in the residual coding stage. The object-based nature of the coding technique and the bit stream architecture enable the coder to support the content scalability by providing methods to selectively encode, decode, and manipulate individual objects in the compressed bit stream. In addition, the error protection and concealment strategies developed enable robust transmission of compressed video with acceptable video quality in the face of severe channel degradations. Graceful error concealment is made possible due to the object-oriented structuring of the coded bitstream syntax. This technique was proposed by Texas Instruments as a candidate for the evolving ISO MPEG-4 standard. The proposal performed impressively in the first round of MPEG-4 tests. When complete, the ISO MPEG-4 standard will provide robust video coding algorithms that will enable a host of future video communications applications.

REFERENCES

- [1] F. Pereira, "MPEG-4: A new challenge for the representation of audio-visual information," in *Proc. Picture Coding Symp., PCS '96*, Melbourne, Australia, Mar. 1996, pp. 7-16.
- [2] R. Talluri, "MPEG-4 status and directions," in *Proc. SPIE Critical Reviews of Standards and Common Interfaces for Video Information Systems*, Philadelphia, PA, Oct. 1995, pp. 252-262.
- [3] "MPEG-4 Proposal Package Description (ppd)—revision 3," ISO/IEC/JTC1/SC29/WG11, Doc. no. N0998, July 1995.
- [4] "MSDL ad hoc group, requirements for the MPEG-4 syntactic description language," ISO/IEC/JTC1/SC29/WG11, Doc. no. N1022, July 1995.
- [5] "MSDL ad hoc group, MSDL specification version 1.0 testing logistics," ISO/IEC/JTC1/SC29/WG11, Doc. no. N1164, Jan. 1996.
- [6] "MPEG-4 test and evaluation procedures document," ISO/IEC/JTC1/SC29/WG11, Doc. no. N0999, July 1995.
- [7] H. G. Musmann, "A layered coding scheme for very low bit rate video coding," *Signal Processing: Image Commun.*, vol. 7, pp. 267-278, Nov. 1995.
- [8] J. Ostermann, "Object-oriented analysis-synthesis coding based on the source model of moving, flexible 3D objects," *IEEE Trans. Image Processing*, vol. 3, Sept. 1994.
- [9] T. Ebrahimi, E. Rousens, and W. Li, "New trends in very low bitrate video coding," *Proc. IEEE*, vol. 83, pp. 877-891, June 1995.
- [10] P. Gerken, "Object-based analysis-synthesis coding of image sequences at very low bit rates," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 4, pp. 228-235, June 1994.

- [11] "Video coding for low bitrate communication, draft recommendation H.263," ITU-T, Study Group 15, Oct. 1995.
- [12] A. Das and R. Talluri, "Error protection and concealment techniques for object-based video compression schemes," in *Proc. Picture Coding Symp., PCS '96*, Melbourne, Australia, Mar. 1996, pp. 579-584.
- [13] "Report of the ad hoc group on MPEG-4 video testing logistics," ISO/IEC/JTC1/SC29/WG11, Doc. no. N0578, Nov. 1995.
- [14] R. Talluri, "A hybrid object-based video compression technique," in *Proc. Int. Conf. Image Processing, ICIP '96*, Lausanne, Switzerland, Sept. 1996, pp. 387-390.
- [15] R. Talluri and B. Flinchbaugh, "Video coding below twenty kilo-bits per second," presented at *Proc. European Conf. Multimedia, Applications, Services and Techniques, ECMAST '96*, Louvain-la-Neuve, Belgium, May 1996.
- [16] D. H. Ballard and C. M. Brown, *Computer Vision*. Englewood Cliffs, NJ: Prentice-Hall, 1982.
- [17] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*. Reading, MA: Addison-Wesley, 1993.
- [18] V. Markandey, S. Wang, M. Burton, and A. Reid, "Image analysis for moving target indication applications," *Texas Instruments Tech. J.*, vol. 8, Sept. 1991.
- [19] M. Irani, B. Ruso, and S. Peleg, "Computing occluding and transparent motions," *Int. J. Computer Vision*, vol. 12, pp. 5-16, Feb. 1994.
- [20] G. Medioni and Y. Yasumoto, "Corner detection and curve representation using cubic b-splines," *Computer Vision, Graphics and Image Processing*, vol. 59, pp. 267-278, Nov. 1987.
- [21] K. Rijkse, "ITU standardization of very low bitrate video coding algorithms," *Signal Processing: Image Commun.*, vol. 7, pp. 553-565, Nov. 1995.
- [22] K. Oehler, "Region-based wavelet compression for very low bitrate video coding," in *Proc. Int. Conf. Image Processing, ICIP '96*, Lausanne, Switzerland, Sept. 1996, pp. 573-576.
- [23] J. Villaseor, B. Belzer, and J. Liao, "Wavelet filter evaluation for image compression," *IEEE Trans. Image Processing*, vol. 4, pp. 1053-1060, Aug. 1995.
- [24] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 31, pp. 614-627, Dec. 1989.
- [25] Y. Q. Zhang, "Variable bit-rate video in the broadband ISDN environment," *Proc. IEEE*, vol. 79, pp. 214-222, Feb. 1991.
- [26] M. Ghanbari, "Two-layer coding of video signals for VBR networks," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 771-781, June 1989.
- [27] M. Nomura, T. Fujii, and N. Ohta, "Layered packet-loss protection for variable rate video coding using DCT," presented at *Proc. Int. Workshop on Packet Video*, Sept. 1988.
- [28] P. Haskell and D. Messerschmitt, "Resynchronization of motion compensated video affected by ATM cell loss," in *Proc. 1992 Int. Conf. Acoustics, Speech, Signal Processing*, Mar. 1992, pp. 545-548.
- [29] A. R. Reisman, "DCT based embedded coding for video," *Signal Processing: Image Commun.*, vol. 3, pp. 333-343, Sept. 1991.
- [30] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Englewood Cliffs, NJ: Prentice-Hall, 1994.



Raj Talluri (S'89-M'91) received the Bachelor of Engineering degree in electronics and communications engineering from Andhra University, India, in 1984, the Master of Engineering degree in electrical engineering from Anna University, India, in 1986, and the Ph.D. degree from the University of Texas at Austin, in 1993.

He worked as a Scientist at the Indian Space Research Organization's Satellite Center, Bangalore, from 1986 to 1987 and as a Graduate Research Assistant at the Computer and Vision Research Center, University of Texas at Austin from 1988 to 1992. He has been with Texas Instruments Corporate Research since 1993, currently as the Manager of the Video Technology Branch, DSP R&D Center, Dallas, TX. He has been actively representing Texas Instruments at the ISO MPEG Standards body. He is also a participant in the ITU Low Bitrate Coding activities. He holds two patents and has authored a number of research articles in the areas of computer vision, robotics, and video compression. His current research interests include video coding, error resilience, and computer vision.

Dr. Talluri is the Co-Program Chair of the Third IEEE Workshop on Applications of Computer Vision, December 1996.



Karen Oehler (S'85-M'90) received the B.S. degree *summa cum laude* in electrical engineering from Rice University, Houston, TX, in 1987. As a Winston Churchill Scholar, she received a Certificate of Postgraduate Study in engineering from Cambridge University, England, in 1988. She received the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1989 and 1993, respectively.

Since 1993, she has been a Member of Technical Staff at the DSP R&D Center of Texas Instruments, Dallas, TX. She has also participated in the ITU H.263 and ISO MPEG-4 video compression standardization efforts. Her research interests include digital video compression and processing.

Dr. Oehler is a member of Eta Kappa Nu, Tau Beta Pi, and Sigma Xi.



Thomas Bannon received the B.S. degree in electrical engineering from Texas A&M University, College Station, TX, in 1984, and the M.S. degree in computer science from The University of Texas at Dallas, Richardson, in 1991.

After pre-graduate work at The University of Texas at Austin in radio-television-film and computer science, he joined Texas Instruments, Plano, in 1986. He is currently a Member of Technical Staff in the Vision Research branch in the Personal Systems Laboratory of TI Corporate R&D in Dallas.

He has published a number of papers and holds two patents in the area of object-oriented databases.



Jonathan D. Courtney (S'92-M'93) received the B.S. degree with honors in computer engineering and computer science and the M.S. degree in computer science both from Michigan State University, East Lansing, in 1991 and 1993, respectively.

He subsequently joined Texas Instruments, Dallas, where he is a member of the Multimedia Systems Branch of TI Corporate R&D. His current research interests include multimedia information systems and virtual environments for cooperative work.



Arnab Das received the B. Tech. degree in electronics and communications engineering in 1992 from the Indian Institute of Technology, Madras. He received the M.S. degree in electrical engineering in 1996 from the University of Maryland, College Park, where he is currently engaged in a Ph.D. program in electrical engineering.

He worked as a Summer Intern at the Texas Instruments Corporate Research Labs in Summer 1994 and Summer 1995. His research interests include image and video compression, high-speed networks, and queuing theory.



Judy Liao received the B.S. degree in electrical engineering and computer science from the University of California at Berkeley in 1991, the M.S. degree from the University of California at Los Angeles (UCLA) in 1993, and the Ph.D. degree from UCLA in 1996.

She is currently with Rockwell Semiconductor Systems in Newport Beach, CA. Her research interests include image/video compression and wireless video communication systems.

Dr. Liao is a member of Tau Beta Pi.

Wavelet-Based Lossless Compression of Coronary Angiographic Images

Adrian Munteanu,* Jan Cornelis, *Member, IEEE*, and Paul Cristea, *Senior Member, IEEE*

Abstract—The final diagnosis in coronary angiography has to be performed on a large set of original images. Therefore, lossless compression schemes play a key role in medical database management and teleradiology applications. This paper proposes a wavelet-based compression scheme that is able to operate in the lossless mode. The quantization module implements a new way of coding of the wavelet coefficients that is more effective than the classical zerotree coding. The experimental results obtained on a set of 20 angiograms show that the algorithm outperforms the embedded zerotree coder, combined with the integer wavelet transform, by 0.38 bpp, the set partitioning coder by 0.21 bpp, and the lossless JPEG coder by 0.71 bpp. The scheme is a good candidate for radiological applications such as teleradiology and picture archiving and communications systems (PACS's).

Index Terms—Coronary angiography, lifting scheme, lossless compression, medical image coding.

I. INTRODUCTION

TECHNOLOGICAL improvements and significant cost reductions of digital angiographic units have triggered an increasing demand for acquiring and handling the coronary angiograms in a digital form. Moreover, various types of X-ray digitizers are currently being used to replace data storage on 35-mm cine films with digital storage of coronary angiograms. The main advantages of digital data representation consist of the simple archiving and manipulation of the images and the fast image transfer over large networks. Inevitably, the compression of digital coronary angiograms is of paramount importance in picture archiving and communications systems (PACS's) and teleradiology applications, due to the huge amount of patient data (e.g., in a typical sequence of 2500 images per patient, each image of 512×512 pixels, represents 655 MB [1]). Such applications require both efficient storage of images and low-cost data transmission through digital communication lines. The compression ratios required for transmitting 512×512 pixels coronary angiographic images at 5, 10, and 15 frames per second, using ISDN with channel capacities of $p \times 64$ kbit/s, are illustrated in Fig. 1.

Note that if one considers that for this class of images, the standard lossless JPEG provides an average compression ratio

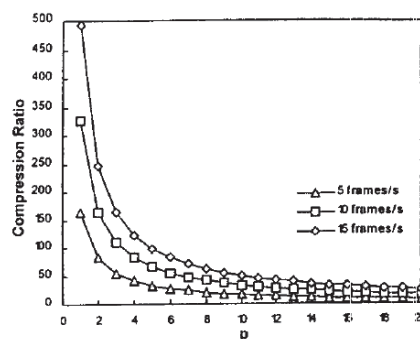


Fig. 1. Compression ratio required for transmission of 512×512 pixels coronary angiograms using ISDN with channel capacities of $p \times 64$ kbit/s.

of three (see Table I), more than 54 channels are required to achieve a throughput of five frames/s. We conclude that applying compression results in improved efficiency, since it reduces both the storage costs and the image transfer time.

Various methods, both for lossy (irreversible) and lossless (reversible) image compression are proposed in the literature. The recent advances in the lossy compression techniques include different methods, such as vector quantization [2], wavelet coding [3], [4], and some less widespread approaches like neural networks [5] and fractal coding [6]. Although these methods can achieve high compression ratios (of the order 50:1, or more), they do not allow one to reconstruct exactly the original version of the input data. Lossless compression techniques, however, permit the perfect reconstruction of the original image, but the achievable compression ratios are only of the order 2:1 up to 4:1. Currently, the lossy coding methods are not accepted for coronary angiogram compression [7]. In recent years, several methods have been proposed for lossless (medical) image compression, such as context-based adaptive lossless image coding (CALIC) [8], hierarchical interpolation (HINT) [9], two-dimensional (2-D) multiplicative autoregressive model-based coding [10], and segmentation-based image coding [11]. Although these coders are among the most performant ones as far as the compression ratio is concerned, the lack of a progressive transmission capability makes them unsuitable for radiological applications expanded over large networks. The concept of progressive data transmission is of particular importance in an electronic

Manuscript received October 28, 1997; revised January 31, 1999. This work was supported in part by BARCO N.V. and IWT under the SAMSET Project ITA/950 204. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was C. Roux. Asterisk indicates corresponding author.

*A. Munteanu and J. Cornelis are with the Electronics and Information Processing Department, Vrije Universiteit Brussel, Brussels B-1050, Belgium. P. Cristea is with the Digital Signal Processing Laboratory, Politehnica University of Bucharest, Bucharest 77206, Romania.
Publisher Item Identifier S 0278-0062(99)04100-2.

0278-0062/99\$10.00 © 1999 IEEE

TABLE I
AVERAGE COMPRESSION RATE, AVERAGE BIT RATE (BPP), AND DIFFERENCE IN
AVERAGE BIT RATE (BPP), WITH RESPECT TO ALGORITHM A, OBTAINED IN THE
LOSSLESS COMPRESSION OF THE TEST SET OF CORONARY ANGIOGRAMS

Algorithm	Compression Ratio	Bit Rate (bpp)	Difference (bpp)
LJPG	3.0608	2.6137	+0.7134
BTPC	3.6024	2.2207	+0.3204
EZW & IW	3.4933	2.2901	+0.3898
SPHT	3.7880	2.1119	+0.2116
Algorithm A	4.1891	1.9003	0.0000
Algorithm B	4.2026	1.9036	+0.0033
CALIC	4.2072	1.8820	-0.0183

radiology environment, including applications such as fast telebrowsing through large data sets or telediagnosis [12]. In ΔS and teleradiology are becoming available over large networks and, as distance interactivity grows, the available network bandwidth has a tendency to collapse. For economical data transmission over large networks, it is therefore desirable to send progressively the compressed data and to refine the images gradually at the destination site, retaining the option of perfect reconstruction.

Additionally, intelligent image transfer scheduling and extra functionalities of the employed compression technique can significantly reduce the access time to image data for the end users working in a network environment. In this context, the JPEG 2000 committee defines a set of profiles for still image compression covering typical applications, including medical ones (e.g., remote sensing, medical, digital photography, etc.). Some of the most prominent requirements [13] from the medical profile are enumerated in the following.

- 1) *Lossy-to-Lossless Compression*: Lossy-to-lossless compression based on a unique compressed bit stream, as well as a much better rate-distortion performance at low bit rates than the current standard JPEG [14]. These two requirements are essential for medical-image handling scenarios based on progressive transmission.
- 2) *Definition of Regions-of-Interest (ROI's) at the Sender Site*: Usually, medical images are taken under the assumption of a certain pathology. Therefore, based on some contextual knowledge, one can *a priori* define several ROI's which can be refined at the receiver site much faster than their surroundings, resulting in a faster data interaction, less network congestion, and better image database access.
- 3) *Definition of ROI's by the Receiver*: In many cases, at an early stage of transmission the end user can indicate the ROI's within an image or the particular 2-D slices within a 3-D data set on which he would like to carry out a detailed analysis. We have recently reported [15] on a case study of tumor delineation carried out on a set of ROI's while the transmission of the 3-D MRI data set is going on over a 64 kbit/s ISDN line: while for the uncompressed data, the full transmission would have taken around 1.5 h, progressive transmission and ROI support allowed to delineate and render the tumor in 3-D in less than 2 min [15].

- 4) *Error Resilience*: In some cases, the errors that might occur due to transmission can be reduced by coding the most important information (e.g., the coarse version of the image) with a very robust error correction, while weaker error correction methods can be applied in the least significant parts of the transmitted bit stream. Possible applications are the transmission of medical images from a portable medical imaging device in an ambulance or from distant areas to a nearby hospital.

These requirements are not supported by the current JPEG standard [14]. Among the possible alternatives for a lossless compression scheme with progressive transmission capabilities, the transform-based techniques are the most attractive. The wavelet transform can be implemented in compression schemes featuring progressive transmission in which one transmits the coarser version of the image first, followed by successive transmissions of the refinement details. The paper presents a wavelet-based compression method that is able to operate in the lossless mode. There is a significant difference between this method and the standard wavelet compression techniques [3], [4], which cannot reconstruct the lossless version of the original image because of the rounding to integers of the wavelet coefficients. Our alternative is the use of the lifting scheme [16] to compute an integer wavelet representation. Using lifting, it is particularly easy to build nonlinear wavelet transforms that map integers to integers and which permit a lossless representation of the image pixels [17], [18]. Moreover, the factorization of the wavelet transform into lifting steps reduces the computational complexity by a factor of two, as compared with the standard algorithm implementing the fast wavelet transform [19].

The paper is organized as follows. Section II briefly describes the integer wavelet transform. The quantization module described in section III implements a new modality for the coding of the wavelet coefficients that is more effective than the classical zerotree coding [3]. Section IV reports on comparative experimental results obtained on 20 coronary angiograms with the proposed algorithms and other lossless compression schemes. An example is given to illustrate the progressive transmission capabilities of the proposed methods and a visual comparison to a JPEG block transform scheme shows the superior performance of our algorithms.

II. INTEGER WAVELET TRANSFORM

In most of the cases, the filters used in wavelet transforms have floating point coefficients. Since the input images are matrices of integer values, the filtered output no longer consists of integers and losses will result from rounding. For lossless coding it is necessary to make an invertible mapping from an integer image input to an integer wavelet representation. The solution is given in [18] where it is shown that the integer version of every wavelet transform employing finite filters can be built with a finite number of lifting steps.

The lifting scheme is used in [16] to construct symmetric biorthogonal wavelet transforms starting from interpolating Deslauriers-Dubuc scaling functions [20]. A family of (N, \tilde{N}) symmetric biorthogonal wavelets is derived, where N is the

number of vanishing moments of the analysis high-pass filter and \tilde{N} is the number of vanishing moments of the synthesis high-pass filter. An instance of this family of transforms is the (4,2) interpolating transform. The integer version of it, given in [18], is implemented in the first stage of our coding algorithm. In this case, the integer wavelet representation of a one dimensional signal $A^0(n)$ having N nonzero samples is given by

$$\begin{aligned} \forall n: D^{i+1}(n) &= A^i(2n+1) - \left[\sum_k p_k A^i(2(n-k)) + 1/2 \right] \\ &\quad 0 \leq i < J, \quad 0 \leq n < 2^{-(i+1)}N, \\ &\quad -2 \leq k \leq 1 \\ \forall n: A^{i+1}(n) &= A^i(2n) + \left[\sum_k u_k D^{i+1}(n-k) + 1/2 \right] \\ &\quad 0 \leq i < J, \quad 0 \leq n < 2^{-(i+1)}N, \\ &\quad 0 \leq k \leq 1 \end{aligned} \quad (1)$$

where $[x]$ represents the integer part of x , J is the number of scales, $A^{i+1}(n)$ and $D^{i+1}(n)$ denote, respectively, the approximation and the detail of the original signal calculated at the scales $(i+1)$, $0 \leq i < J$.

The coefficients of the filters in (1) are given in [18]

$$\begin{aligned} p_{-2} &= p_1 = -1/16, & p_{-1} &= p_0 = 9/16 \\ u_0 &= u_1 = 1/4. \end{aligned} \quad (2)$$

The extension of (1) to the 2-D case consists of the successive application of the one-dimensional (1-D) wavelet transform, as described in [21]. This results in the representation of any arbitrary input image f via the set

$$(A_J, \tilde{W}X_k, WY_k, WXY_k)_{1 \leq k \leq J} \quad (3)$$

where A_J is the approximation of f at the coarser resolution 2^{-J} , and $W\tilde{X}_k$, WY_k , and WXY_k , are the details of the image in horizontal, vertical, and diagonal orientation, respectively, at the resolutions 2^{-k} , $1 \leq k \leq J$.

III. WAVELET COEFFICIENTS CODING

The coefficient's significance maps, introduced by Shapiro in [3], are binary maps indicating the positions of the significant wavelet coefficients. It is proven in [3] that the zerotree representation can efficiently encode the significance maps by predicting the absence of significant information across scales. The image model used in the zerotree coding (EZW) is extremely general and is based on the zerotree hypothesis [3]. This hypothesis assumes that if a wavelet coefficient w at a certain scale is insignificant with respect to a given threshold T , i.e., $|w| < T$, then all the coefficients of the same orientation in the same spatial location at finer scales are also insignificant with respect to the threshold T . As shown in [3], the zerotree hypothesis is satisfied with a high probability. As a consequence, the interband redundancies are efficiently exploited by grouping the insignificant coefficients in trees growing exponentially across the scales, and by coding them with zerotree symbols. In contrast to EZW [3] and set

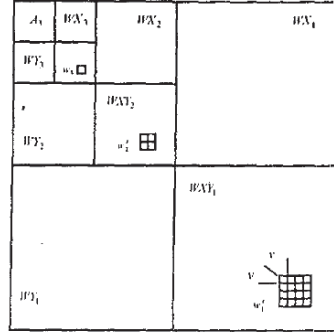


Fig. 2. Three-level wavelet representation ($J = 3$): the squares $w_1^i, 1 \leq i \leq 16$ and $w_j^j, 1 \leq j \leq 4$ delimit all the descendants of the root coefficients in w_3 .

partitioning into hierarchical trees (SPIHT) [22], which are zerotree based coders, the new coding method presented in the following exploits only intraband dependencies.

Define the tree associated to an arbitrary root coefficient $w(i, j)$ from a band $W_k \in \{W\tilde{X}_k, WY_k, WXY_k, 1 \leq k \leq J\}$, as the set:

$$D(w) = \{w(2^p i + m, 2^p j + n), 0 \leq p < k, 0 \leq m, n < 2^p\}.$$

The wavelet image is partitioned into a lattice of squares of width v , each square delineating a set of wavelet coefficients. Consider a parent square w_J located in the subband W_J and select the v^2 trees corresponding to all the root coefficients w from w_J . It can be proven that the coefficients from $\bigcup_{w \in w_J} D(w)$ located in an arbitrary subband $W_k \in \{W\tilde{X}_k, WY_k, WXY_k, 1 \leq k \leq J\}$ are delimited by the squares w_k^i with $1 \leq i \leq 4^{J-k}$. It follows that the total number of squares delimiting the selected v^2 trees is

$$N_{\text{squares}} = \sum_{k=1}^J 4^{J-k}.$$

For reasons of simplicity, we depict in Fig. 2 the particular case of $J = 3$ where $w_1^i, 1 \leq i \leq 16$ and $w_j^j, 1 \leq j \leq 4$ delimit all the descendants of the root coefficients in w_3 .

Choose an arbitrary square w_k^i located in a subband $W_k \in \{W\tilde{X}_k, WY_k, WXY_k, 1 \leq k \leq J\}$ and associate with it the logarithmic threshold T_k^i defined as

$$T_k^i = \lfloor \log_2 w_{\max} \rfloor \quad (4)$$

where w_{\max} is the maximum absolute value of the wavelet coefficients delimited by w_k^i .

The value T_k^i indicates that all the coefficients in w_k^i are not significant with respect to the whole set of thresholds: $2^{T_k^i}$, for any T satisfying $T_k^i < T$

$\forall w$ inside w_k^i

$$|w| \leq w_{\max} = 2^{\log_2 w_{\max}} < 2^{\lfloor \log_2 w_{\max} \rfloor + 1} = 2^{T_k^i + 1} \leq 2^T. \quad (5)$$

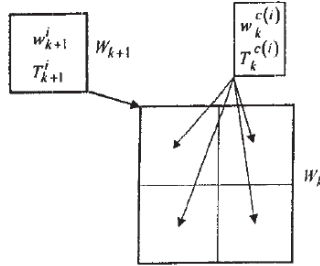


Fig. 3. All the descendants from $w_{k+1}^{(i)} \in W_{k+1}$ in the finer subband W_k are located in the squares $w_k^{(i)}$; the associated logarithmic thresholds are $T_k^{(i)}$. For the set of coefficients included in $w_{k+1}^{(i)}$, respectively, $T_k^{(i)}$ for the corresponding descendants located in $w_k^{(i)}$.

Also, the definition (4) of $T_k^{(i)}$ indicates that at least one coefficient from $w_k^{(i)}$ is significant with respect to $2^{T_k^{(i)}}$

$$w_{\max} = 2^{\log_2 w_{\max}} \geq 2^{\lceil \log_2 w_{\max} \rceil} = 2^{T_k^{(i)}}. \quad (6)$$

We deduce from (5) that the $T_k^{(i)}$ representation can be used to code the significance of all the coefficients delimited by $w_k^{(i)}$ with respect to the set of thresholds 2^T , ($T_k^{(i)} < T$). For a typical nine-bits-per-pixel wavelet image (including one bit per pixel for the sign), the range of the $T_k^{(i)}$ values is $[0, 7]$, therefore one can use three bits to represent them. Hence, the total number of bits used to code the significance of the coefficients from the considered v^2 trees is

$$N = 3 \cdot N_{\text{squares}} = 3 \cdot \sum_{k=1}^J 4^{(J-k)} = 4^J - 1. \quad (7)$$

We conclude from (6) that significant coefficients are expected in the square $w_k^{(i)}$ for thresholds lower than $2^{T_k^{(i)}}$. The significance of the coefficients from $w_k^{(i)}$ with respect to the thresholds 2^T , $0 \leq T \leq T_k^{(i)}$ is determined by applying successive approximation quantization (SAQ).

Assuming, as in the case of the zerotree-like coders [3], [22], that the wavelet coefficients satisfy the zerotree hypothesis, it is shown hereafter that our proposed coding method outperforms the classical zerotree coding, when used in the lossless mode.

To make a comparison with the zerotree coding scheme, we have to determine the number of zerotree symbols used to code the significance of the coefficients from the same set of trees.

Select an arbitrary square $w_{k+1}^{(i)}$ from the subband W_{k+1} and denote by $w_k^{(i)}$ the squares delimiting the corresponding descendants in the finer subband W_k (see Fig. 3). If $T_{k+1}^{(i)}$ respectively $T_k^{(i)}$ are the associated logarithmic thresholds, from their definitions we write

$$T_{k+1}^{(i)} = \lceil \log_2 (w_{k+1}^{(i)})_{\max} \rceil, \quad T_k^{(i)} = \lceil \log_2 (w_k^{(i)})_{\max} \rceil$$

where $(w_{k+1}^{(i)})_{\max} = \max_{w \in w_{k+1}^{(i)}} |w|$ and $(w_k^{(i)})_{\max} = \max_{w \in w_k^{(i)}} |w|$ are the maximum absolute values of the wavelet coefficients in the squares $w_{k+1}^{(i)}$, respectively $w_k^{(i)}$.

From (5) and (6) we write

$$2^{T_{k+1}^{(i)}} \leq (w_{k+1}^{(i)})_{\max} < 2^{T_{k+1}^{(i)}+1}, \quad (8)$$

$$2^{T_k^{(i)}} \leq (w_k^{(i)})_{\max} < 2^{T_k^{(i)}+1}. \quad (9)$$

The zerotree hypothesis, which we assume to be valid, states that if the parent coefficients in $w_{k+1}^{(i)}$ are not significant with respect to $2^{T_{k+1}^{(i)}+1}$, then all the corresponding descendants in the finer subbands are also not significant. Therefore

$$\forall w \text{ inside } w_k^{(i)}, \quad |w| \leq (w_k^{(i)})_{\max} < 2^{T_k^{(i)}+1}. \quad (10)$$

From (9) and (10) we deduce $T_k^{(i)} < T_{k+1}^{(i)} + 1$ which is equivalent to $T_k^{(i)} \leq T_{k+1}^{(i)}$. This shows that the $T_k^{(i)}$ values are increasing with k

$$T_k^{(i)} \leq T_{k+1}^{(i)}, \quad \forall k, 1 \leq k < J, \quad \forall i, 1 \leq i \leq 4^{J-k-1}. \quad (11)$$

If we apply the thresholds 2^T with $T_k^{(i)} < T \leq T_{k+1}^{(i)}$, the coefficients delimited by $w_{k+1}^{(i)}$ and all the descendants from the finer subbands are not significant. This can be shown by choosing an arbitrary descendant coefficient with value w belonging to $W_p \cap D(w_{k+1}^{(i)})$, $p < k$ and having the associated logarithmic threshold T_p . From (5) and (11) we can write $|w| < 2^{T_p+1} \leq 2^{T_k^{(i)}+1} \leq 2^T$ which proves the nonsignificance of w with respect to 2^T ($T_k^{(i)} < T \leq T_{k+1}^{(i)}$).

The result is that the significance of the coefficients in $w_k^{(i)}$ and the significance of their descendants in W_p , $p < k$ is coded with zerotree symbols for all the thresholds 2^T with $T_k^{(i)} < T \leq T_{k+1}^{(i)}$. Hence, for all the squares $w_k^{(i)}$, with $1 \leq i \leq 4^{J-k}$, the number of zerotree symbols is (see Fig. 3)

$$v^2 \cdot \sum_{i=1}^{4^{(J-k)}} (T_{k+1}^{(i)} - T_k^{(i)}).$$

The number of zerotree symbols used to code the significance with respect to the thresholds 2^T , $T_J < T \leq 7$ of the roots from W_J and of all the corresponding descendants is $v^2 \cdot (7 - T_J)$.

The total number of zerotree symbols needed to code the significance of the selected v^2 trees amounts to

$$N_{\text{ZTR}} = v^2 \cdot \left[(7 - T_J) + \sum_{k=1}^{J-1} \sum_{i=1}^{4^{J-k}} (T_{k+1}^{(i)} - T_k^{(i)}) \right] \stackrel{\text{notation}}{=} v^2 \cdot A. \quad (12)$$

The proposed coding scheme is more efficient than the zerotree coding if N [given by (7)] is smaller than the number of bits needed to code the zerotree symbols (two bits for each symbol)

$$2 \cdot N_{\text{ZTR}} > N \Leftrightarrow 2 \cdot v^2 \cdot A > 4^J - 1. \quad (13)$$

This inequality is satisfied for

$$v \geq \frac{2^J}{\sqrt{2 \cdot A}} \quad (14)$$

Relation (14) specifies a theoretical bound for which the coding of the significance map through the T_k^i representation has the same efficiency as the zerotree coding.

Assume a practical case in which the width of the squares is $v = 4$ and consider a wavelet decomposition with $J = 6$ of a 512×512 pixels input image. If equality is considered in (14) we get $A = 128$.

The number of wavelet coefficients contained in one tree having the root in an arbitrary square $w_6 \in W_6$ is

$$\sum_{k=1}^6 4^{(J-k)} = 1365.$$

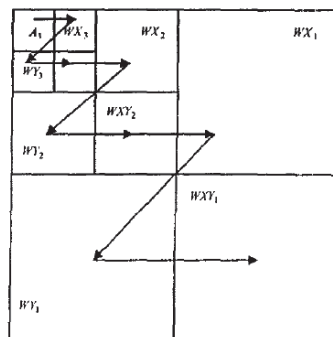
From the above example we conclude that our method and EZW have equal efficiency if the significance of these 1365 coefficients is coded with only 128 zerotree symbols (equivalent to 0.19 bits per coefficient). In practical situations the value of A is much higher, which explains the gain attained with this method, even for small values of v (typically $v = 2$ or $v = 4$).

We could incorrectly conclude from (13) that the efficiency of our method increases when a large value of v is chosen. However, there is another (intuitive) constraint that imposes small values of v , which is illustrated in the following example.

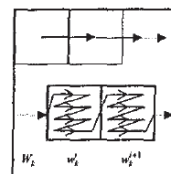
Suppose we have an extreme case where only one coefficient from the square w_k^i is significant with respect to 2^{T_1} (and not significant with respect to 2^{T_1+1}) and all the other $v^2 - 1$ coefficients are significant with respect to 2^{T_2} (and not significant with respect to 2^{T_2+1}), with $T_1 > T_2$. The significance of the $v^2 - 1$ coefficients with respect to the thresholds 2^T , $T_2 < T \leq T_1$ has to be coded with $(v^2 - 1) \cdot (T_1 - T_2)$ bits. Let us now split the square w_k^i into four squares of half width w_k^j , $1 \leq j \leq 4$ and associate to each of them the logarithmic thresholds T_k^j . A number of $3 \cdot (v/2)^2$ coefficients delimited by three squares w_k^j have the T_k^j value equal to T_2 , while the other $(v/2)^2$ coefficients have the associated logarithmic threshold equal to T_1 . It follows that only for the last ones we do need to code the significance with respect to the thresholds 2^T , $T_2 < T \leq T_1$. Halving the width of w_k^i brings a gain of $(3v^2/4) \cdot (T_1 - T_2) - 9$ bits. The last term comes from the nine additional bits used to code the three additional T_k^j values.

We conclude from the above example that for high values of v the influence in a certain square w_k^i of the isolated coefficients with high values increases the overhead needed for the significance coding of the coefficients with small values. It is the case of the w_k^i regions located around edges characterized by wavelet coefficients of high values. Increasing the width v causes more overhead information to be used for the significance coding of the coefficients with small values, corresponding to the smooth areas in the image.

Taking into account the constraints regarding the value of the square's width, we choose $v = 4$ for the experiments reported in Section IV.



(a)



(b)

Fig. 4. (a) Three-level wavelet representation and the scanning order of the subbands symbolized by arrows. Note that the scanning starts with the lowest frequency band A_3 and ends with the highest frequency band WXY_1 . (b) In any subband W_k the corresponding squares w_k^j are scanned in raster order. For any scanned w_k^j the corresponding coefficients are also scanned in raster order.

A. Coding Scheme A

A coding scheme that implements the previously described significance representation is explained in the following. The wavelet image is partitioned into a lattice of squares of width $v = 4$ and for each square, the associated logarithmic thresholds are determined according to (4). The maximum T_{\max} of the determined T_k^i values is recorded as side information to inform the decoder about the starting threshold $2^{T_{\max}}$.

The wavelet image is scanned, starting with the lowest frequency subband A_J and ending with the highest frequency subband WXY_1 [see Fig. 4 (a)]. In each subband W_k the corresponding set of squares w_k^j are scanned in raster order. For any scanned square w_k^j , the corresponding coefficients are scanned also in raster order, as shown in Fig. 4(b).

During encoding SAQ is applied. This consists of the use of a set of thresholds 2^T which are sequentially applied, starting with $T = T_{\max}$ and ending with $T = 0$. No coding is performed for all the wavelet coefficients w belonging to w_k^j if $T > T_k^j$. Conversely, for each $T \leq T_k^j$, the significance with respect to 2^T of all the coefficients w belonging to w_k^j is recorded into a significance list, list1.

If a certain coefficient is declared as significant with respect to 2^T , we record its sign into a second list (sign list), list2.

The quantization module is followed by arithmetic data compression. Each threshold T_k^i is recorded as a matrix

element having a position within the matrix corresponding to the spatial location of the associated square w_k^i in the wavelet image. The matrix is scanned in raster order and the resulting stream of symbols is entropy coded with a high-order arithmetic coder.

The list of significance (list1) is also entropy coded by using the arithmetic coder described in [23] and a classical adaptive zero-order model [23]. There is no need to code the sign list (list2), since there is almost no gain in the compression ratio.

We observe that the most important coefficients (having a higher T_k^i) are coded before the least important ones (having a small T_k^i). Therefore, the T_k^i representation allows an ordering of the coefficients, starting from the ones of high values and ending with the ones of low values. This enables the coder to transmit progressively the data, which is gradually refined at the decoder site, until perfect reconstruction is achieved.

It is also interesting to observe the spatial extension of the v -based lattice partitioning of the wavelet image. In the particular case of Fig. 2, we notice that the selected v^2 trees corresponding to all the root coefficients w from w_3 are related to a square region of $8v \times 8v$ pixels located in the spatial domain. These trees contain all the wavelet coefficients located in the subbands $WXY_k, 1 \leq k \leq 3$ that correspond to this region. More generally, due to the dyadic nature of the wavelet transform, for a wavelet decomposition with J levels the v^2 trees corresponding to all the root coefficients w from a square $w_J \in WXY_J$ are related to a square region of $2^J v \times 2^J v$ pixels located in the spatial domain. Similarly, one can identify the squares selecting the wavelet coefficients corresponding to this spatial domain region and located in the subbands $WX_k, WY_k, 1 \leq k \leq J$ and A_J .

The correspondence between a certain spatial region and a set of squares delimiting wavelet coefficients in the wavelet domain can be used if one is interested in coding several ROI's in the image. Suppose that we require that a certain rectangular ROI is coded before coding the background information. The squares from the lattice that correspond to the ROI are determined and marked, and the coefficients belonging to these squares are coded first. Similarly, the coding process continues with the coding of all the other ROI's, and ends with the coding of the remaining wavelet coefficients (included in the squares which are not marked and which actually correspond to the background information). This coding process is very useful when several predefined diagnosis-based ROI's are delineated, coded, and transmitted before sending the background information, resulting in the advantages mentioned in Section I.

The algorithm described above is identified as Algorithm A in the experiments reported in Section IV.

B. Coding Scheme B

Another scheme implementing the previously described approach for the coefficient's significance coding can be derived if we impose that the square's width v is one. The algorithm described in the following is related to Algorithm A, and it gives an interesting insight concerning the achievable compression results for a minimal value of v . Although in this case we cannot claim that the proposed algorithm is more

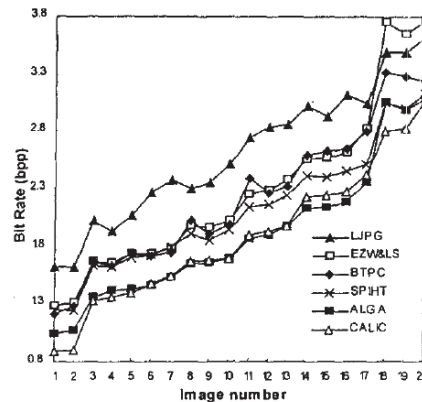


Fig. 5. Bit rates achieved in the lossless compression of 20 coronary angiograms with various algorithms. Note that Algorithm A and CALIC provide the best performances for the whole set of images.



Fig. 6. Coronary angiogram, 512×512 pixels, 8 bpp: the area of stenosis is indicated by the white rectangle.

efficient than the zerotree coding [(14) might not be satisfied], good performances are expected in lossless coding due to the nature of the wavelet transform.

If $v = 1$, the logarithmic thresholds T_k^i are coding the significance of each wavelet coefficient apart. In this case, T_k^i indicates that the corresponding coefficient is not significant with respect to $2^T, T_k^i < T \leq 7$ and significant with respect to $2^{T_k^i}$ if it has a nonzero value. It results that the most significant $(8 - T_k^i)$ bits (excluding the sign bit) from the binary representation of any wavelet coefficient are represented with three bits (since $\forall T_k^i, 0 \leq T_k^i \leq 7$), therefore, a gain of $(5 - T_k^i)$ bits is feasible.

Note that the values 0, 1 and -1 are identified by the same value $T_k^i = 0$, thus, a distinction between them has to be made using separate information.

An intuitive explanation of the efficiency of such significance representation is given by the observation that the

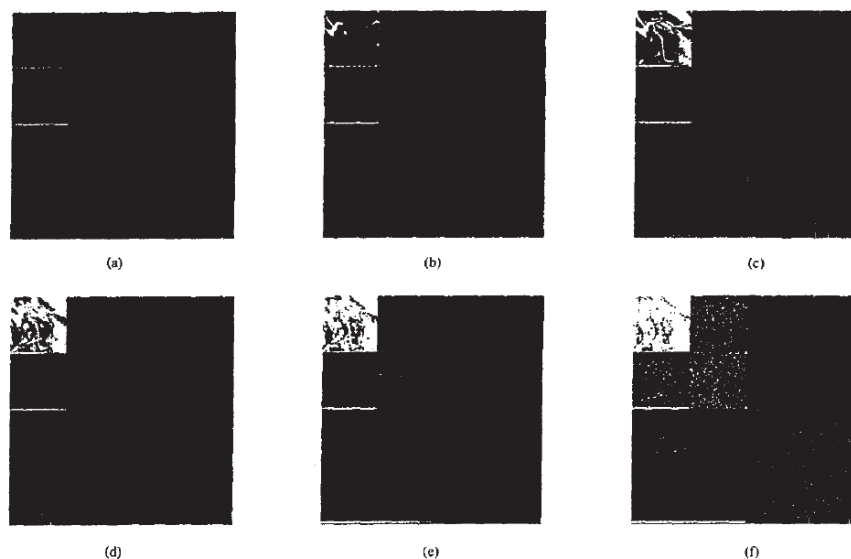


Fig. 7. Illustration of different significance maps obtained by the thresholding with (a) 64, (b) 32, (c) 16, (d) 8, (e) 4, and (f) 2 of the images obtained by the two-level wavelet decomposition of the angiogram in Fig. 6. The black and white pixels correspond to the nonsignificant and significant coefficients, respectively.

majority of the wavelet coefficients have small values. We deduce that a gain is feasible if a high number of wavelet coefficients satisfy $T_k^i \leq 5$.

Denote by p_T the probability of an arbitrary coefficient w to be significant with respect to 2^T and nonsignificant with respect to 2^{T+1} where $1 \leq T \leq 7$

$$p_T = p(2^T \leq |w| < 2^{T+1}). \quad (15)$$

Also, denote by p_0 the probability of an arbitrary coefficient w to satisfy $|w| < 2$ (i.e., $|w| \in \{0, 1\}$).

If the total number of the wavelet coefficients is N , then the number of the coefficients that satisfy $2^T \leq |w| < 2^{T+1}$, ($1 \leq T \leq 7$) is $N_T = p_T \cdot N$ and the number of 0, ± 1 values is $N_0 = p_0 \cdot N$.

As was previously explained, the gain obtained for each coefficient with an associated logarithmic threshold T is $(5 - T)$ bits, therefore the total gain is

$$G = \sum_{T=0}^7 [(5 - T) \cdot p_T \cdot N].$$

However, separate information makes the distinction between the values 0 and ± 1 , for which we need $p_0 \cdot N$ bits. It follows that the gain achieved by using this representation of the coefficient's significance is

$$G = \sum_{T=0}^7 [(5 - T) \cdot p_T \cdot N] - p_0 \cdot N. \quad (16)$$

The gain has a positive value if

$$4 \cdot p_0 + 4 \cdot p_1 + 3 \cdot p_2 + 2 \cdot p_3 + p_4 > p_6 + 2 \cdot p_7. \quad (17)$$

Although the argumentation is empirical, in general

$$p(2^T \leq |w| < 2^{T+1}) \geq p(2^{T+1} \leq |w| < 2^{T+2}) \quad (18)$$

which means that in a wavelet image the values $|w|$ have a higher probability of occurrence than the higher values $2 \cdot |w|$.

From (15) and (18) we write $p_T \geq p_{T+1}$, which implies that (17) is satisfied.

The coding scheme implementing this method is broadly the same as the Algorithm A. The difference consists in the side information used to make the distinction between the values 0, and ± 1 . We identify this method as Algorithm B in the section reporting the experimental results.

IV. EXPERIMENTAL RESULTS

We evaluate the lossless compression performances obtained with the proposed algorithms on a set of 20 digitized coronary angiograms and we compare them with the results obtained with other lossless compression techniques such as CALIC [8], binary tree predictive coding (BTPC) [24], lossless JPEG (LJPG) [14], [25], the EZW coder [3] combined with the integer wavelet transform (EZW & IWT), and the set partitioning into hierarchical trees coder (SPIHT) [22].

It is important to mention that in contrast to our proposed techniques operating in the wavelet domain, both BTPC and CALIC are predictive coders operating in the spatial domain. BTPC uses a noncausal shape-adaptive predictor to decompose

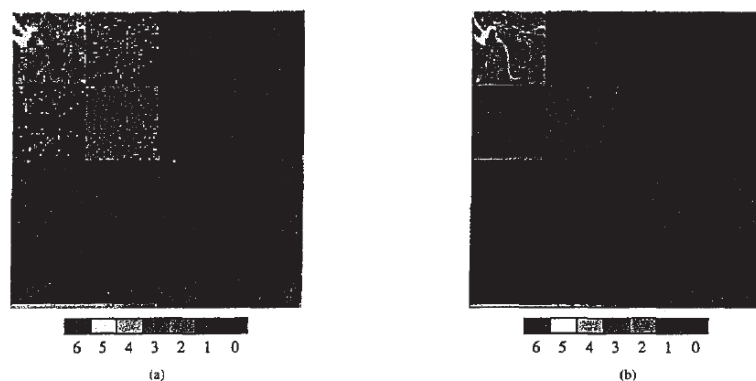


Fig. 8. Logarithmic thresholds images corresponding to the angiogram in Fig. 6 for (a) Algorithm A ($r = 4$), and (b) Algorithm B ($r = 1$). The colors displayed in the colorbars correspond to the seven possible values of the logarithmic thresholds.

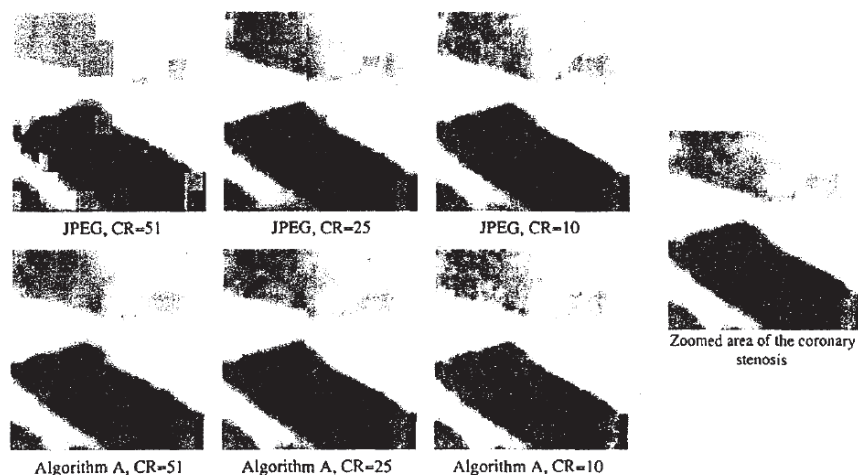


Fig. 9. The image of Fig. 6 is compressed at 51:1, 25:1 and 10:1 with JPEG, and Algorithm A. The zoomed area of the coronary stenosis indicated by the white rectangle in Fig. 6 is refined progressively up to the lossless version depicted on the right.

an image into a binary tree of prediction errors and zero blocks. BTPC is designed to perform both lossless and lossy still-image compression and is very effective for compression of photographs, graphics, and compound images containing both text and continuous-tone images. The best lossless compression performances reported in the literature for continuous-tone images are mostly obtained with CALIC. This technique is a context-based, adaptive, predictive coder designed for the lossless/nearly lossless compression of continuous-tone images. CALIC uses a large number of modeling contexts to condition a nonlinear predictor and makes it adaptive to varying source statistics. Efficient techniques for forming and

quantizing modeling contexts assure the low time and space complexities of CALIC.

In the chart in Fig. 5 we plot the bit rates achieved with the tested algorithms applied on the individual images. Note that the curves corresponding to Algorithm A and CALIC are below the curves corresponding to the other algorithms for all the images from the test set.

Note also from Table 1 that, on average, Algorithm A outperforms with 0.21 bpp the SPIHT coder, with 0.32 bpp the BTPC coder, with 0.38 bpp the EZW & IWT coder, and with 0.71 bpp the LJPEG. Algorithm B has the same performances as Algorithm A (the difference between them is only 0.003 bpp).

Also, the difference between the average bit rate achieved with Algorithm A and CALIC is only 0.018 bpp. Additionally, it is important to remark that although CALIC provides a slightly better compression performance for the tested set of images, it cannot be used for progressive image transmission.

Notice that the results reported in Table 1 are obtained for two values of v : the value $v = 4$ is the size of the square lattice used in Algorithm A, while the extreme case $v = 1$ is used in Algorithm B. The average lossless-coding results do not critically depend on the value of v . We have experimentally observed, however, that for this class of images and for values of v higher or equal to 16, the efficiency of Algorithm A decreases. The explanation is offered in Section III, where we show that for high values of v the coding of the coefficients with high values increases the overhead needed for the significance coding of the coefficients with small values.

To give an intuitive interpretation of how the significance information is conveyed and to show how SAQ operates, we illustrate in Fig. 7 several significance maps corresponding to the image shown in Fig. 6. For a wavelet decomposition of this image on two levels, we find that the highest threshold applied by SAQ is $2^{T_{\max}} = 64$. The significance map corresponding to this threshold is shown in Fig. 7(a). The white and the black pixels indicate the positions of the significant and nonsignificant coefficients respectively. As expected, the number of significant coefficients increases as the value of the applied threshold is lower, as we notice in Fig. 7(b)–(f), where the significance maps corresponding to the thresholds 32, 16, 8, 4, and 2, respectively, are shown. As explained in Section III A, at each applied threshold 2^T the significance of the coefficients within a given square w_k^T is encoded if T is lower or equal to the corresponding logarithmic threshold T_k^T . The matrix of logarithmic thresholds is depicted in Fig. 8(a). The seven colors used to draw this image correspond to the seven possible values of the applied thresholds 2^T , $0 \leq T \leq 6$ and are shown in the color bar in Fig. 8(a). The significance information for the coefficients delimited by the squares with the highest associated logarithmic threshold ($T_{\max} = 6$) is coded first; these squares are indicated with a dark-brown color in Fig. 8(a). For the lower threshold of value 32, the significance of the wavelet coefficients within the squares with an associated logarithmic threshold of five is coded; these squares are yellow-brown colored in Fig. 8(a). Similarly, the coding process continues for the all lower applied thresholds until the significance information is fully encoded, i.e., up to the lossless stage. For Algorithm B the coding process is similar, but in this case the matrix of logarithmic thresholds [shown in Fig. 8(b)] operates at pixel level ($v = 1$).

We note that the use of the logarithmic thresholds allows the encoder to order the significant coefficients in their relative order of importance and to encode the coefficients with higher values first. If at a certain point in the coding process the target bit rate is met, then the coding is simply stopped and the image is decoded. This property was used to produce, with Algorithm A, the images shown in Fig. 9.

In this figure, the progressive refinement of a zoomed area from a coronary stenosis is shown comparatively, both for Algorithm A and for the standard JPEG [14]. Although the

progressive transmission scheme implemented in Algorithm A is not optimal in the rate distortion sense (our purpose is to demonstrate the progressive transmission capabilities), we note that our technique preserves a better visual quality at high compression ratios, while typical blocking artefacts are present in JPEG compression.

The CPU times spent on a 143-MHz ULTRA-SPARC computer to encode/decode a 512×512 8 bpp coronary angiogram with Algorithm A are approximately 0.8 and 0.85 s, respectively. For Algorithm B these times increase up to approximately 1.2 s for encoding and 1.3 s for decoding of the same image. Since the programs implementing these algorithms are not optimized for speed, these figures are just an indication of the achievable speed on usual computer architecture.

V. CONCLUSIONS

Radiological applications such as PACS and teleradiology are expanding over large networks, requiring an efficient lossless compression scheme with progressive transmission capabilities.

Considering the reported results (lossless compression ratio's equal to the state of the art CALIC technique) and the low complexity of the algorithms, we conclude that the proposed compression schemes are good candidates for radiological applications for archiving, for economical image transfer over limited bandwidth channels, and for fast interaction with coarser versions of the image data. It is the progressive transmission capability of the algorithms that makes them attractive for compression schemes in which a subset or all the received images have to be gradually refined up to the lossless version of the input data.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

REFERENCES

- [1] J. Cornelis, P. Suetens, and I. Lemahieu, "Kijken in de mens: Medische beelden en beeldvorming," (Deel 1/3), *Het Ingenieursblad*, KIVI, no. 3, pp. 70–83, 1998.
- [2] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.
- [3] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3452, Dec. 1993.
- [4] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 205–220, Feb. 1992.
- [5] R. D. Dony and S. Haykin, "Neural networks approaches to image compression," *Proc. IEEE*, vol. 83, pp. 288–303, Feb. 1995.
- [6] Y. Fisher, Ed., *Fractal Compression: Theory and Application to Digital Images*. New York: Springer-Verlag, 1994.
- [7] R. P. Lewis, "President's page: Developing standards for the digital age: The Dicom project," *J. Amer. Coll. Radiol.*, vol. 28, no. 6, pp. 1631–1632, 1995.
- [8] X. Wu, "Lossless compression of continuous-tone images via context selection, quantization, and modeling," *IEEE Trans. Image Processing*, vol. 6, pp. 656–664, May 1997.
- [9] P. Roos, M. A. Vergever, M. C. A. van Dijke, and J. H. Peters, "Reversible intraframe compression of medical images," *IEEE Trans. Med. Imag.*, vol. 7, pp. 328–336, Dec. 1988.

- [10] M. Das and S. Burgert, "Lossless compression of medical images using two-dimensional multiplicative autoregressive models," *IEEE Trans. Med. Imag.*, vol. 12, pp. 721-726, Dec. 1993.
- [11] L. Shen and R. M. Rangayyan, "A segmentation-based lossless image coding method for high-resolution medical image compression," *IEEE Trans. Med. Imag.*, vol. 16, pp. 301-307, June 1997.
- [12] S. E. Elnaahs, K. Tzou, J. R. Cox, R. L. Hill, and R. G. Jost, "Progressive coding and transmission of digital diagnostic pictures," *IEEE Trans. Med. Imag.*, vol. MI-5, pp. 73-83, June 1986.
- [13] T. Hamid, DICOM Requirements for JPEG 2000, ISO/IEC JTC1/SC 29/WG1, N944, 1998.
- [14] G. K. Wallace, "The JPEG still picture compression standard," *Comm. ACM*, vol. 34, no. 4, pp. 30-44, 1991.
- [15] P. Schelkens, A. Munteanu, A. Salomie, R. Deklerck, V. Enescu, Y. Christophe, and J. Cornelis, "Validation of the usefulness of the wavelet based compression algorithms for progressive transmission of large medical images," ISO/IEC JTC1/SC29/WG1 N884, 1998.
- [16] W. Sweldens, "The lifting scheme: A custom design construction of biorthogonal wavelets," *J. Appl. Comput. Harmonic Analysis*, vol. 3, no. 2, pp. 186-200, 1996.
- [17] S. Dewitte and J. Cornelis, "Lossless integer wavelet transform," *IEEE Signal Processing Lett.*, vol. 4, pp. 158-160, June 1997.
- [18] A. R. Calderbank, I. Daubechies, W. Sweldens, and B. Yeo, "Wavelet transforms that map integers to integers," Dept. Math., Princeton Univ., Tech. Rep., 1996.
- [19] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," Bell Labs-Lucent Technol., Murray Hill, NJ, Tech. Rep., 1996.
- [20] G. Deslauriers and S. Dubuc, "Interpolation dyadique," in *Fractals, Dimensions non Entières et Applications*. Paris, France: Masson, 1987, pp. 44-55.
- [21] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, pp. 674-693, July 1989.
- [22] A. Said and W. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst.*, vol. 6, pp. 243-250, Mar. 1996.
- [23] A. Moffat, R. M. Neal, and I. H. Witten, "Arithmetic coding revisited," in *Proc. IEEE Data Compression Conf.*, Los Alamitos, CA, 1995, pp. 202-211.
- [24] J. A. Robinson, "Efficient general-purpose image compression with binary tree predictive coding," *IEEE Trans. Image Processing*, vol. 6, pp. 601-608, Apr. 1997.
- [25] H. Kongji and B. Smith, "Lossless JPEG codec. Version 1.0," June 1994. Available <http://www.cs.cornell.edu/Info/Projects/multimedia/Projects/LJPG.html>.

Signal Analysis and Compression Performance Evaluation of Pathological Microscopic Images

Akira Okumura,* Junji Suzuki, Member, IEEE, Isao Furukawa, Sadayasu Ono, Member, IEEE, and Tsukasa Ashihara

Abstract—Digitizing high-quality microscopic images and developing input/output technology for displaying those results is critical to telepathology in which pathological microscopic images are transferred to remote locations where they are diagnosed by specialists. This paper will discuss the results achieved by directly digitizing (nonfilm process) pathological microscopic images at a $2k \times 2k$ resolution, and then using a super-high-definition imaging system to analyze their signals and evaluate compression performance. We will start off by digitizing samples that a pathologist will actually use in making a diagnosis, and then analyze their color distribution and spatial frequencies characteristics by comparing them to general images. This will make it apparent that such pathological images characteristically contain high spatial frequency in their chrominance components. We will also discuss the evaluation results of color differences for $L^*a^*b^*$ space and compression ratios achieved when using JPEG to encode pathological images. We will also present a subjective evaluation of the influence subsampling of chrominance components has on image quality.

Index Terms—JPEG compression, pathological microscopic images, Super-high-definition (SHD) images, telepathology.

1. INTRODUCTION

Super-high-definition (SHD) images [1] have a spatial resolution of 2000×2000 pixels or more and a temporal resolution of 60 frames/s or more (progressive scanning). They are particularly suited to fields where high image quality is in demand, such as medical imaging and printing. Amongst the various applications for SHD images, their use as medical images has been attracting attention in recent years due to the major societal impact of medical images in the areas of diagnosis, observation, and the education of doctors [2]. Medical information consists of charts containing text, one-dimensional data including waveforms, audio, and two-dimensional (2-D) and three-dimensional (3-D) images. Research currently underway is digitizing all such medical information in order to unify management of patient information, speed-up information searches, prevent loss or deterioration, and improve diagnostic performance [3]. The ultimate goal of the digitization of medical information is to create a

filmless/paperless system, making it possible to quickly obtain the necessary patient information, including images, anytime and anywhere.

There is presently a wide range of medical images used for making diagnoses. Simple chest X rays (monochrome) and pathological microscopic images (color) require the highest levels of quality among all others. Simple chest X rays must precisely reproduce images of the lungs and different regions of the body consisting of diverse structures such as organs or bones, while pathological microscopic images must faithfully reproduce at various magnifications images including minute structures and subtle changes in cellular texture. Those are the respective reasons why the said images must be of high quality. The transferring of medical information including X rays and pathological microscopic images over communication lines in order to make diagnosis are called teleradiology and telepathology, respectively. Various systems for both are under investigation [2], [4]. In addition to high-speed digital systems and wideband communications systems, it is imperative that we develop input/output systems that can faithfully reproduce image information which demands high fidelity. This will make it possible to advance the medical system and make it more efficient through the digitization of all medical information.

In this paper we will discuss the assembly of a system that can faithfully input (digitize) and display pathological microscopic images, and then present the resulting signal analysis and compression performance of the image thereby obtained. The estimated resolution required for the general display of pathological microscopic images is between approximately 4–6 million pixels [5]. At present, the SHD image system we mentioned above is capable of displaying such a high pixel resolution. Compression must be applied to efficiently transfer and store SHD images. A single SHD image is 100 Mb. When these are used to create moving images at 60 frames/s, we get a data stream amounting to 6 Gb/s. A number of compression standards have been developed, including JPEG and MPEG1 and 2. Tests have been underway for some time on the application of these and other compression methods to medical images [6]–[9]. However, little attention has been paid to their application to pathological microscopic images.

In Section II, we will discuss the assembly of an image input/output system that centers on a charge-coupled device (CCD) camera for digitizing pathological microscopic images. In preparation of an investigation into the compression of pathological microscopic images, Section III will present the analysis results of color distribution and spatial frequency

Manuscript received July 29, 1996; revised June 29, 1997. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was H. K. Huang. Asterisk indicates corresponding author.

*A. Okumura is with NTT Optical Network Systems Laboratories, 1-1 Hikari-no-oka, Yokosuka-shi, Kanagawa 239, Japan (e-mail: nura@esa.onlab.ntt.co.jp).

J. Suzuki, I. Furukawa, and S. Ono are with NTT Optical Network Systems Laboratories, Yokosuka-shi, Kanagawa 239, Japan.

T. Ashihara is with the Kyoto Prefectural University of Medicine, Kawaramachi-Hirokoji, Kamikyo-ku, Kyoto 602, Japan.

Publisher Item Identifier S 0278-0062/97/0707580-0.

0278-0062/97\$10.00 © 1997 IEEE

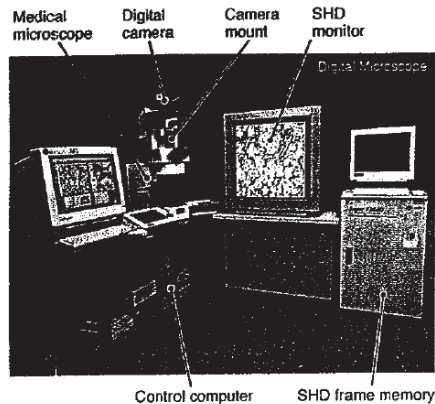


Fig. 1. Image digitizing system from pathological microscope.

characteristics to elucidate the differences between general images (including portraits and landscapes) and pathological microscopic images. In Section IV, we will present the JPEG compression results on pathological microscopic images as well as the results of our evaluation of color differences that arise due to subsampling of chrominance components. Finally, in Section V, we will present our subjective evaluation of the effect of subsampling of chrominance components on pathological microscopic images.

II. IMAGE INPUT/OUTPUT

A. System Configuration

Fig. 1 shows the super-high-definition input/output system for pathological microscopic images. This system is composed of an input system centering on a microscope for medical use and CCD area sensor-equipped digital camera and an output system centering on SHD image frame memory and a monitor [10]. Table I shows the principle data for the pathological image input/output system. A minutely adjustable camera mount unit has been placed between the microscope and CCD digital camera so that the focus of the microscope image will fit on the CCD surface. This means that a pathologist can digitally capture the sample he is observing under the microscope simply by using the computer's mouse. The CCD digital camera can input resolutions of 2048×2048 and pixel depths of 14 b/pixel. In order to display the image on the SHD monitor, gamma correction is performed on the 14-b/pixel data, converting it to 8 b/pixel. The time required to transfer data on this system is about 30 s when capturing pathological microscopic images to the system control computer and about 60 s when loading data from the system control computer to the SHD frame memory.

B. Images Used for Evaluations

The samples we used to create images for making evaluations were preparations of pathological samples created

TABLE I
PRINCIPAL DATA FOR THE PATHOLOGICAL IMAGE INPUT/OUTPUT SYSTEM

Medical microscope	A X80 (made by Olympus Optical Corporation Ltd., Tokyo Japan), 10 \times ocular lens and 2 \times , 4 \times , 10 \times , 20 \times , 40 \times , and 100 \times objective lens
Digital camera	Leaf Digital Camera (made by Scitex Corporation Ltd., Herzliya B. Israel), Cooled CCD, 2,048 \times 2,048 pixel resolution, RGB sequential input, 14-bit color, Capturing speed: About 30 seconds/image
Camera mount	Focus adjustment ring, continuously adjustable within ± 1 mm (made by NTT, Tokyo Japan)
Control computer	Macintosh 8100/80 (made by Apple Computer Inc., California U.S.A.), Loading speed: About 60 seconds/image
SHD frame memory	Super FM-III (made by NTT, Tokyo Japan), 2,048 \times 2,048 pixels, 8-bit RGB color (16 million colors), 60 frames/sec progressive scanning, total of 256 frames
SHD monitor	DDM-2802C2 (made by SONY Corporation, Tokyo Japan), 28-inch CRT, 2048 \times 2048 pixel resolution, Maximum brightness: More than 80 cd/m ²

TABLE II
SUPER-HIGH-DEFINITION IMAGE SAMPLES

Pathological microscopic images		General images	
image	contents	image	contents
M1	kidney cancer (magnified by 100)	G1	still life (fruits and vegetables)
M2	kidney cancer (magnified by 40)	G2	portrait with color chart
M3	uterus cancer (magnified by 400)	G3	portrait
M4	uterus cancer (magnified by 200)	G4	landscape (cafeteria)
M5	bronchial pneumonia (magnified by 100)	G5	still life (fruits basket)
M6	colonic cancer (magnified by 100)	G6	still life (wine bottle and tableware)
M7	enlargement of prostate (magnified by 100)	G7	still life (bicycle with color chart)

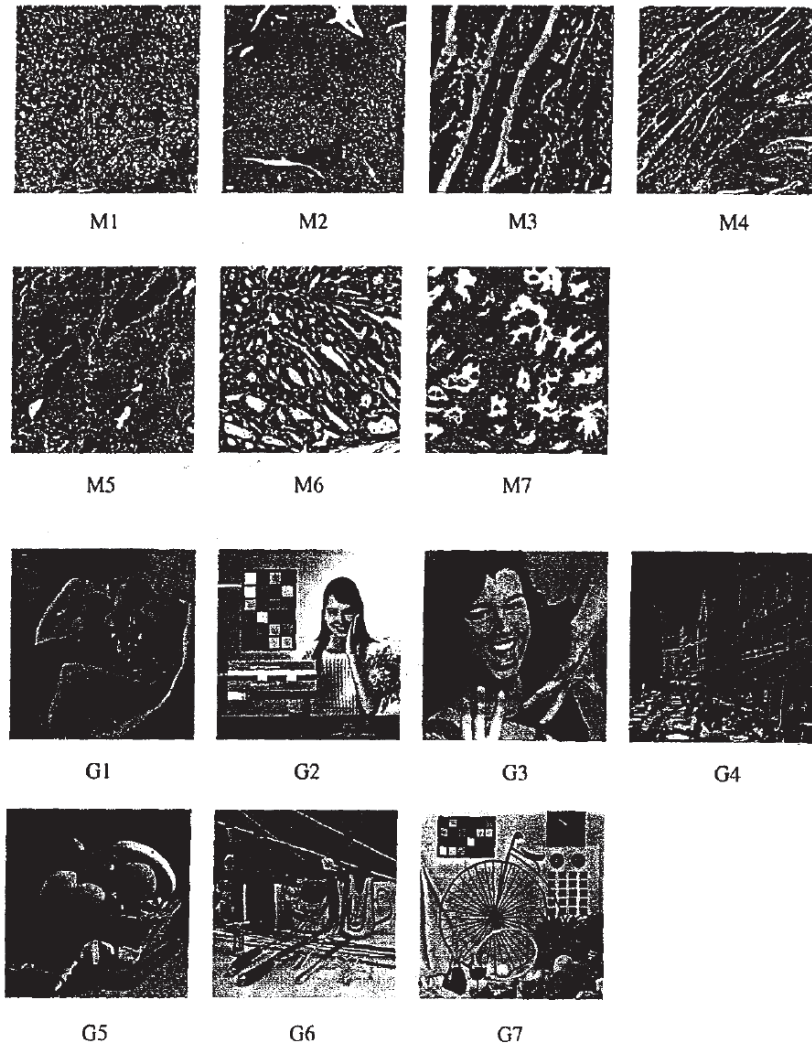
by the Department of Pathology at the Kyoto Prefectural University of Medicine, Kyoto, Japan. Fig. 2(a) shows seven pathological microscopic images (images M1–M7) digitized by the input system discussed above. Table II shows the content and magnification of each sample, most of which contain typical cancer cells. Actual diagnosis of pathological samples using a microscope are conducted by first using a low level of magnification to locate the area that needs to be observed, and then raising the magnification until the desired image is obtained. Consequently, as shown in Table II, a change from low magnification to higher magnifications (40, 100, 200, and 400 power) is required when inputting pathological images.

In order to analyze signals and evaluate compression on pathological microscopic images in the following tests, we also used seven general images (images G1–G7) including portraits and landscapes as a comparison. Fig. 2(b) shows these seven general images. The content of these general images is also shown in Table II. These general images are digitized from film-based photographs, and used for image-quality tests of color printers, CRT monitors, image-coding algorithms, and several kinds of image processing operations. Images G1 and G2 are super-high-definition test images created at NTT Laboratory, [11]. Images G3–G7 are high-definition standard color image data (SCID) images used in printing [12]. All images used in our tests have 2048×2048 pixels and 24 b/pixel (8 b for each RGB color).

III. PROPERTIES OF PATHOLOGICAL MICROSCOPIC IMAGES

A. Color Distribution

Normal pathological samples are almost both colorless and transparent when they are first obtained from diseased tissue.



(b)

Fig. 2. Pathological microscopic image samples. All images have 2048×2048 pixel resolution. (a) Pathological microscopic images (M1–M7). (b) General images (G1–G7).

If left as is, the sample cannot be diagnosed since its cell structure cannot be seen. This means that prepared samples must be created by using special dyes to produce color. Pathological examinations are, therefore, performed by using dye to colorize the structure of the cells to be observed. The appropriate dyes have been added to colorize the images

shown in Fig. 2(a), thereby making visible cancer cells and other items important for a diagnosis. Fig. 3 shows the color distribution of pathological microscopic image M2 and general image G4 on Commission Internationale de l'Eclairage (CIE)- xy chromaticity diagrams. In the diagrams, the z -axis shows a histogram of pixels having each chromaticity value on the

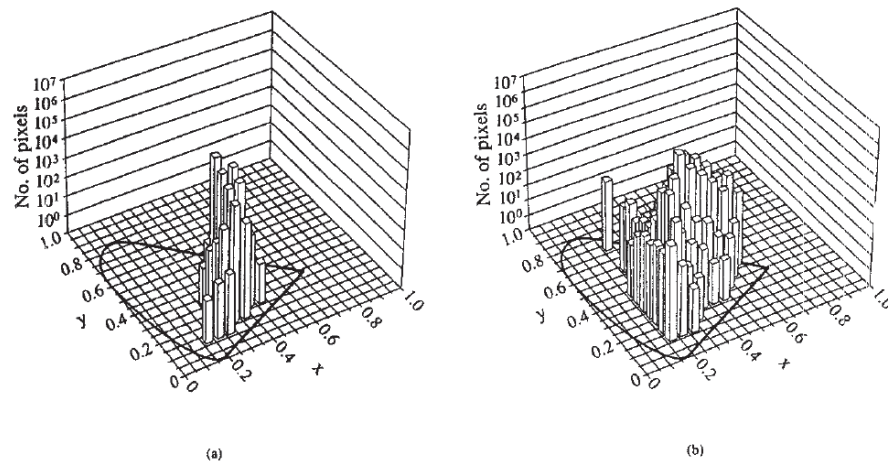


Fig. 3. Color distribution on CIE- xy chromaticity coordinate. The region within solid line on xy -plane corresponds to the area visible to human eyes. (a) Pathological microscopic image M2. (b) General image G4.

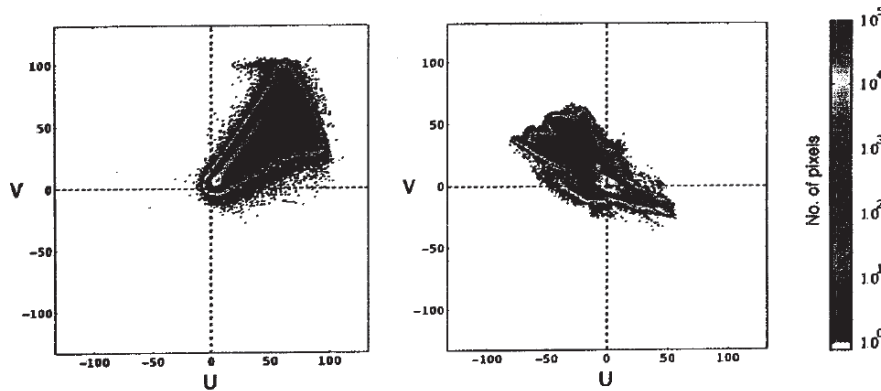


Fig. 4. Distribution of pixel values in UV -plane. Number of pixels is distinguishable with color variations. (a) Pathological microscopic image M2. (b) General image G4.

CIE- xy coordinates [14]. These diagrams make it apparent that pathological microscopic images are more concentrated in specific color components than are general images. The region where this color distribution is concentrated on the CIE- xy chromaticity diagram is dependent on the colorizing properties of the dye used on the pathological sample.

Fig. 4 shows the distribution of pixels in images M2 and G4 represented on the UV -plane after the original images

consisting of RGB signals were converted to a luminance component (Y) and two chrominance components (U , V). The same figure uses color variations to represent a histogram of pixels that have UV component values. It is evident from this figure that the U and V chrominance components in the pathological microscopic image have a normal correlation, are concentrated linearly, and are distributed in a wide area around the concentrated portion. In contrast, the chrominance

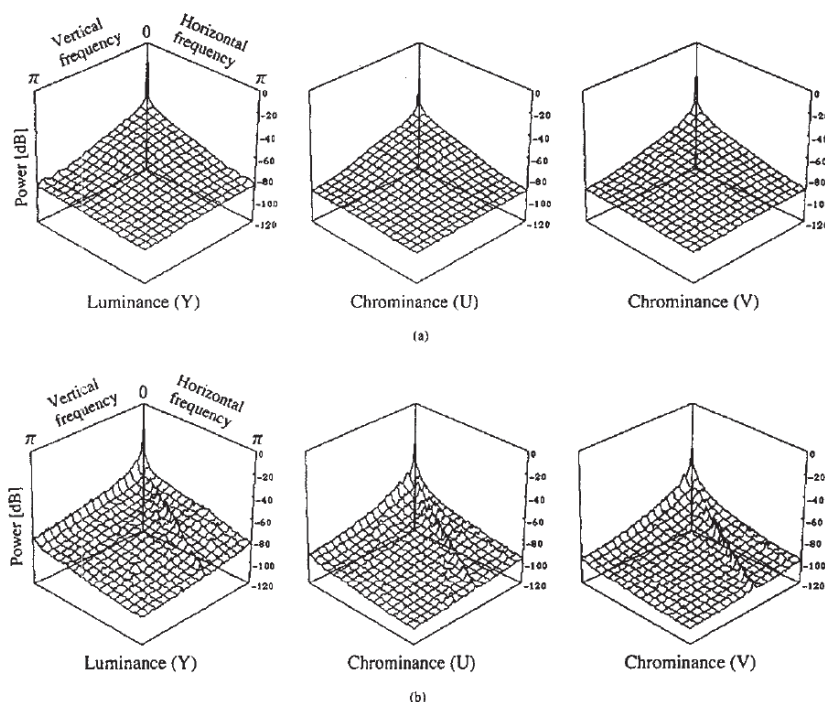


Fig. 5. Spatial spectrum of luminance and chrominance components converted from original RGB color signals. (a) Pathological microscopic image M2. (b) General image G4.

Components of the general image are distributed in an area around the origin, and as a whole, the UV chrominance components have an inverse correlation. Note that although the color distributions shown in Figs. 3 and 4 only show the results for images M2 and G4, other pathological microscopic images and general images show similar properties.

B. Spatial Frequency

We use image-compression techniques that take advantage of the redundancy in image signals and the characteristics of human vision to efficiently transmit and store image data. In order to obtain a high compression performance without lowering subjective quality when using conventional irreversible image compression, such as the discrete cosine transform (DCT)-based method of JPEG (still-image coding standard) and MPEG 1 and 2 (moving-image coding standard), we use subsampling of the chrominance components and quantization of DCT coefficients that takes into account the characteristics of human vision. The subsampling of chrominance components in such compression method takes advantage of the fact that in the signals of general images, the spatial frequency of chrominance components decay more rapidly in the high-

frequency than does the luminance component, as well as the fact that the effects of distortion in chrominance components is rather low for human vision.

We will now elucidate the basic properties and spatial-frequency properties of chrominance components to test methods of compressing pathological microscopic images. Fig. 5 shows the results obtained from the spatial frequency of each component after converting the signals from the original RGB color images (pathological microscopic image M2 and general image G4) to a luminance component (Y) and two chrominance components (U , V). The spatial frequencies in Fig. 5 are obtained by averaging the spatial frequencies of partial images with 256×256 pixel resolution extracted from the original image by overlapping 128×128 pixels from adjacent partial regions. The overlap of the partial images reduces the differences in spatial weight that would, otherwise, be created by the Hamming windowing process used to compute the spatial frequency. We choose to use G4 as our general image because amongst all our other general images it contained the highest spatial-frequency components. Fig. 5 shows that compared to general images, the spatial-frequency band of chrominance components in pathological

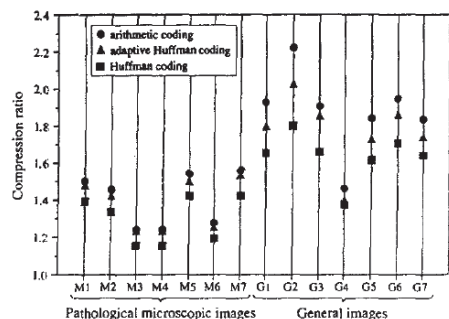


Fig. 6. Compression ratios of JPEG reversible coding. Entropy coding employs arithmetic, adaptive Huffman, and Huffman coding using a fixed table.

microscopic images is remarkably wide. Note that like the color distribution characteristic discussed earlier, this spatial-frequency characteristic will tend to be nearly the same as the results shown in Fig. 5 for other pathological microscopic images and general images. This means that the compression of pathological microscopic images can result in a great deal of color distortion if the chrominance-component subsampling is employed. In Section IV-C, we will discuss our quantitative evaluation of color distortion that occurs during chrominance-component subsampling when the JPEG algorithm is used for irreversible compression.

IV. COMPRESSION AND COLOR DISTORTION

A. Reversible Compression

Here, we discuss the characteristics of reversible compression when applied to pathological microscopic images. The coding algorithms employ the JPEG spatial method [13]. This method is based on DPCM coding, and it employs a predictor in which it predicts the average values of the two pixels above and to the left of encoded pixels. Note that entropy coding employs both Huffman coding and arithmetic coding. In addition, Huffman coding uses both the recommended Huffman tables specified in the JPEG Annex K "Examples and guidelines" [13] and adaptive Huffman tables best suited to each image.

Fig. 6 shows the compression ratios achieved when using JPEG reversible compression on pathological microscopic images and general images. Our definition of compression ratio here is the size of the original image data divided by the size of the compressed data. The highest compression ratio was achieved by arithmetic coding followed by adaptive Huffman coding, and Huffman coding using a fixed table. Unfortunately, even when the best performing method, arithmetic coding, is used, the resulting compression ratio for pathological microscopic images is on the level of a meager 1.2–1.6. This is due to the fact that, as shown in Section III-B, pathological microscopic images contain high

TABLE III
SUBSAMPLING FORMAT FOR THE CHROMINANCE COMPONENTS. THE VALUE L DENOTES IMAGE SIZE

4:4:4	4:2:2	4:1:1
$y(n, m)$	$y(n, m)$	$y(n, m)$
$u(n, m)$	$u(2n', m)$	$u(2n', 2m')$
$v(n, m)$	$v(2n', m)$	$v(2n', 2m')$

n, m, n' and m' are integers defined by:
 $0 \leq n, m < L$
 $0 \leq n', m' < L/2$

spatial-frequency components, which means that compared to general images, they have a high entropy in their differential signals when reversible coding is employed.

B. Irreversible Compression

Irreversible compression algorithms employ the JPEG DCT-based method [13]. Entropy coding only employs the Huffman coding with the previously mentioned Huffman tables recommended in JPEG Annex K. The subsampling methods of the two chrominance components are in the 4:4:4, 4:2:2, and 4:1:1 format shown in Table III. Fig. 7(a) and (b) shows the relationship between compression ratios and signal-to-noise ratio (SNR) of luminance when irreversible compression is applied to pathological microscopic images and general images, where the SNR of luminance is defined by (1)

$$\text{SNR} = 10 \cdot \log_{10} \left\{ \frac{(2^N - 1)^2}{\sum_m \sum_n [y'(n, m) - y(n, m)]^2} \right\} \quad (1)$$

In (1), $y(n, m)$ denotes the luminance of the original image and $y'(n, m)$ denotes the luminance of compressed images; (n, m) is the spatial coordinate, and N is the wordlength of luminance ($N = 8$). The subsampling format for the chrominance components is 4:1:1 in which the pixel values of the two chrominance components are subsampled by one-half each, in both horizontal and vertical directions. This means that the degradation in images reproduced after irreversible compression includes both quantization noise and loss that comes with chrominance-component subsampling. Note that in order to eliminate any effects on the SNR evaluation for distortion due to the transform between RGB and YUV by finite word-length arithmetic processing, the RGB/YUV converted image is used as a reference signal for the SNR calculation. According to Fig. 7, obtaining for pathological microscopic images an SNR of the same luminance component as general images requires that the compression ratio be set relatively lower.

The compression ratio in irreversible compression depends on the format the chrominance components are subsampled. Fig. 8 shows the improvement in the compression ratio for the 4:2:2 and 4:1:1 formats against the compression ratio for the 4:4:4 format, in which chrominance-component subsampling does not take place. The 4:2:2 format mentioned here subsamples by one-half only the pixels of the two chrominance components in the horizontal direction. Improvements in compression ratio is defined by the following (2), where the

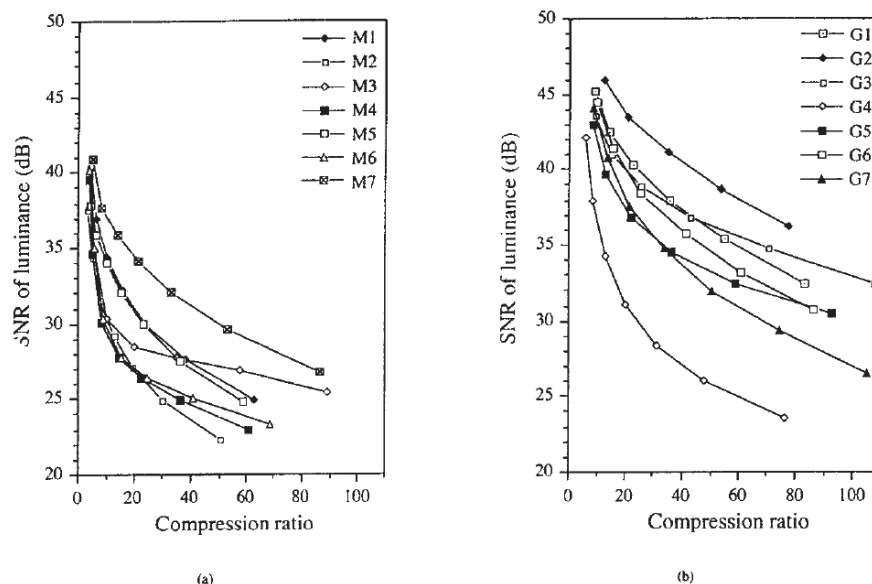


Fig. 7. Compression ratios and SNR of irreversible coding (JPEG DCT-based method) using the 4:1:1 chrominance format. (a) Pathological microscopic images. (b) General images.

compression ratio for the 4:4:4, 4:2:2, and 4:1:1 formats are γ_{444} , γ_{422} , and γ_{411} , respectively

$$(\text{Compression ratio improvement}) = \frac{[\gamma_{422} \text{ (or } \gamma_{411}) - \gamma_{444}]}{\gamma_{444}} \quad (2)$$

Compression ratios generally improve as the subsampling rate of chrominance components increases. If we were to hypothetically assume that the luminance component and chrominance component were compressed at the same level of efficiency, the compression ratio improvement for (2) with the 4:2:2 format would be 0.5, and that with the 4:1:1 format would be 1.0. In reality, the compression-ratio improvement for general images is about 0.15 with the 4:2:2 format and 0.3 with the 4:1:1 format, as can be seen in Fig. 8. The compression-ratio improvement for pathological microscopic images is about 0.3 with the 4:2:2 format and about 0.6 with the 4:1:1 format, or about double that for general images. This can be explained by the fact that compared with general images, pathological microscopic images contain high spatial frequencies in their chrominance components. However, when the subsampling rate for chrominance components increases, the distortion in the chrominance components also increases. This means that the SNR evaluation of the luminance component is insufficient for evaluations that take into account

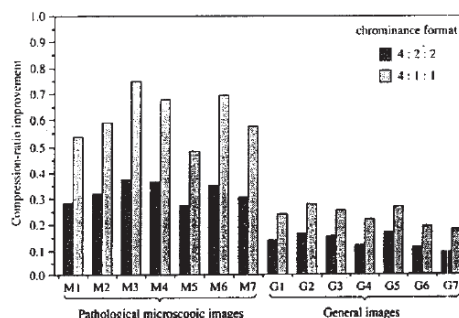


Fig. 8. Improvement of compression ratio against the 4:4:4 chrominance format where the luminance-component SNR is around 40 dB. The improvement of zero means the compressed data size is the same as that using the 4:4:4 format, while the improvement of one means the compressed data size is just one-half of that using the 4:4:4 format.

distortion in the color signals and, therefore, we must use the color difference discussed in the Section IV-C.

C. Color Difference Evaluation

Quantitative evaluations of color differences transform the RGB signals of the original and decoded images into $L^*a^*b^*$

uniform color space and then uses the difference between, or in other words the distance, as measure of evaluation on the color space. To evaluate on $L^*a^*b^*$ space, the color difference of RGB signals after irreversible coding and decoding, first it is transformed from RGB space to XYZ space and the result is then transformed to $L^*a^*b^*$ space according to (3) [14], [15]

$$\left. \begin{aligned} L^* &= 116f\left(\frac{Y}{Y_n}\right) - 16 \\ a^* &= 500\left\{f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right\} \\ b^* &= 200\left\{f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right\} \end{aligned} \right\} \quad (3)$$

where X_n , Y_n , and Z_n can be represented as shown in (4) in case of the CIE's standard light source C which approximates average daylight and white for the National Telecommunication System Committee (NTSC) system [16]

$$\left. \begin{aligned} X_n &= 100 \times \frac{0.3101}{0.3163} = 98.04 \\ Y_n &= 100 \\ Z_n &= 100 \times \frac{0.3736}{0.3163} = 118.12 \end{aligned} \right\} \quad (4)$$

In addition, the function $f(x)$ in (3) is defined by the following equation:

$$f(x) = \begin{cases} x^{1/3}, & \text{if } x > 0.008856 \\ 7.787x + \frac{16}{116}, & \text{if } x \leq 0.008856. \end{cases} \quad (5)$$

The color difference that is ultimately applied in (5) is equivalent to the average value of the geometric distance of the original and decoded images on $L^*a^*b^*$ uniform color space. Note that L_1^* , a_1^* , and b_1^* are the $L^*a^*b^*$ values of the original image, and L_2^* , a_2^* and b_2^* are the $L^*a^*b^*$ values of the decoded image.

$$\Delta C = \sqrt{(L_1^* - L_2^*)^2 + (a_1^* - a_2^*)^2 + (b_1^* - b_2^*)^2}. \quad (6)$$

Fig. 9 shows for pathological microscopic images M2 and M5 and general images G3 and G4, the relationship between JPEG compression ratios and the color difference of decoded images as calculated by (5). This figure also simultaneously shows the compression ratios attained by reversible compression that employs arithmetic coding for its entropy coding. Chrominance-component subsampling during irreversible compression uses the 4:4:4, 4:2:2, and 4:1:1 chrominance formats. In addition, the dotted line links the compression ratio in a straight line, where the luminance component will be 40 dB for each chrominance format. The luminance component in the region to the left of this dotted line will be 40 dB or higher. The unique feature of pathological microscopic images is that compared to general images, they experience a large distortion in their chrominance components due to subsampling. This is a simple conclusion derived from the spectrum evaluation results for the chrominance components discussed in Section III-B. Consequently, even if you set compression ratios low for pathological microscopic images to decrease quantization distortion in the case that subsampling of chrominance components is employed, the color difference

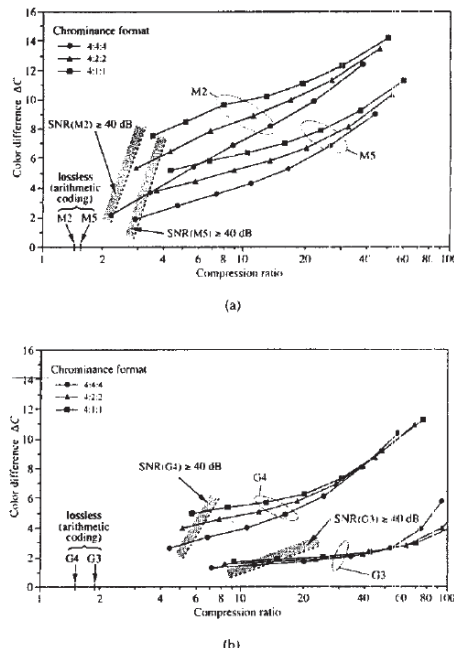


Fig. 9. Compression ratio and color difference using 4:1:1, 4:2:2, and 4:4:4 chrominance formats. Compression ratios attained by reversible compression with arithmetic coding are also included. (a) Pathological microscopic images M2 and M5. (b) General images G3 and G4.

value does not reduce due to distortion in the chrominance components. On the other hand, it is clear that for general images, subsampling of chrominance components has little effect on color difference. Rather, quantization distortion plays a much larger role in color signal distortion.

This means that for the irreversible compression of pathological microscopic images, even if a high-luminance component SNR value of say 40 dB or higher was achieved, there will still be cases where the color difference value will be high. For example, keeping the color difference value to 2.0 or less for pathological microscopic images M2, requires that the compression ratio be lowered to around two even if we use the 4:4:4 format in which chrominance component subsampling does not take place. This approaches the compression ratio that can be achieved by using reversible compression that combines DPCM and arithmetic coding. On the other hand, with an irreversible compression ratio of between ten and 20 for general image G3, it is possible to fulfill this condition regardless of the use of chrominance-component subsampling, and the difference between compression ratios of irreversible coding and reversible coding is considerable. This points to the major difference between pathological microscopic images and general images.

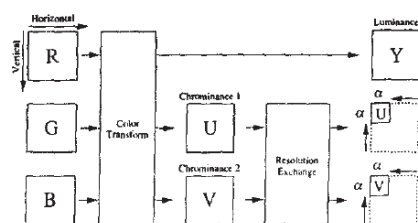


Fig. 10. Chrominance-component subsampling. The chrominance subsampling is implemented by resolution exchange by α ($0 < \alpha < 1$) times the number of pixels of the UV components in both the horizontal and vertical directions.

V. SUBJECTIVE EVALUATION OF CHROMINANCE SUBSAMPLING

Section IV-C discussed the compression of pathological microscopic images and evaluation of the resulting color distortion. This color distortion includes both distortion due to chrominance subsampling and quantization during encoding. This quantization distortion during encoding depends on the coding algorithm, making it difficult to make generalizations about it. We will, therefore, only discuss the results of subjective evaluations of the distortion that comes about due to chrominance subsampling.

Fig. 10 shows the chrominance subsampling algorithm we are discussing here. The chrominance subsampling is implemented by resolution exchange by α ($0 < \alpha < 1$) the number of pixels of the UV components in both the horizontal and vertical directions after transforming the original RGB image signal to luminance (Y) and chrominance (UV) components. The UV component is broken down into 16×16 pixel regions, and then for each region, 2-D DCT with a base length of 6×6 is executed. The higher-frequency component of the transform coefficient is eliminated, and resolution reduction processing is conducting on the remaining transform coefficient using 2-D inverse discrete cosine transform (IDCT) with a base length less than 16×16 . For example, the 4:1:1 format corresponds to the case of $\alpha = 0.5$.

Fig. 11 shows the color difference value ΔC on $L^*a^*b^*$ color space for subsampling rate α when chrominance subsampling is applied to pathological microscopic images M2 and M5 and general images G3 and G4, all of which were evaluated for color difference values in Fig. 9 based on the same method. It is apparent from the results from Fig. 11 and Section IV-C that the color difference value ΔC for pathological microscopic images is much higher than for general images, even when the quantization distortion is not taken into account for the evaluation of color difference.

Next, we will discuss for all of the pathological microscopic images and general images found in Table II, the results of our subjective evaluation of degradation in picture quality that occurs when the chrominance subsampling rate is changed. Table IV shows the test conditions for this evaluation. We used the SHD monitor shown in Table I to display the images. We

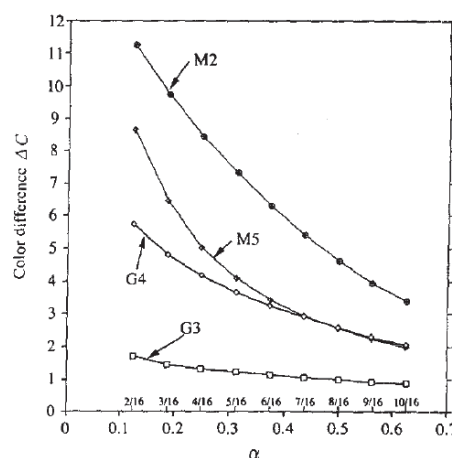


Fig. 11. Color difference value ΔC against chrominance-component subsampling rate α varied from 2/16–10/16.

TABLE IV
SUBJECTIVE EVALUATION CONDITIONS

Lighting conditions	No illumination in a darkroom (0.3 lx or lower with display turned off)
Viewing range	No limit (0 to 1.5 H)
Presentation method	Alternately display the compressed image and original image (controllable by the evaluator)
Observation time	Up to the evaluator (generally 20 sec/image)
Evaluation method	Try to find whether difference can be detected between the two images

made these subjective evaluations by measuring the detectable limit of image degradation due to chrominance subsampling. The results are shown in Fig. 12. This figure shows that removing the general image G7, which contains the highest spatial frequency component, the detectable limits of chrominance subsampling rate α for the general images are lower than that for the pathological microscopic images. Degradation in the general images could be detected mainly as blurring and color shifts occurring around their contours. On the other hand, deterioration in the pathological microscopic images could be detected mainly as changes in the color and texture of specific parts of cells. In addition, general images G2 and G7 have a color chart with sharp edges which makes it easy to detect image quality degradation. When α is a relatively high value, image-quality degradation can only be detected in the color charts. Consequently, the detectable limit value α for G2 and G7 will go even lower if we evaluate the images without their color charts.

Our evaluations found that removing the general images G2 and G7 that include color charts resulted in detectable image degradation at a chrominance subsampling rate of $\alpha = 3/16$ – $5/16$ for general images. This contrasted with our findings for pathological microscopic images in which the detectable image degradation was at a chrominance subsam-

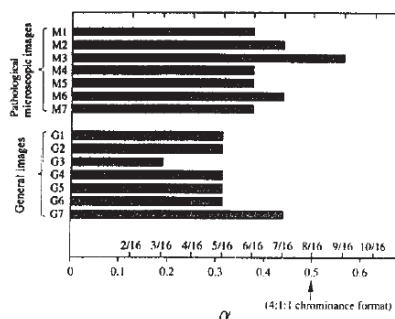


Fig. 12. Chrominance-component subsampling rate where distortion can be detected (the filled-in areas show where distortion can be detected). $\alpha = 0.5$ corresponds to 4:1:1 chrominance format.

pling rate of $\alpha = 6/16$ – $9/16$. The subjective evaluation tests made it clear that the image quality degradation resulting from chrominance subsampling is easier to detect in pathological microscopic images than it is in general images. Furthermore, the results of subjective evaluations of images M2, M5, G3, and G4 agree with the relationships from the evaluation results of color difference values shown in Fig. 11, and this indicates that evaluations using color difference are appropriate for evaluating degradation resulting from chrominance subsampling.

VI. CONCLUSION

We used our super-high-definition input/output system with a spatial resolution of 2048×2048 pixels to create digital images from pathological preparation samples that pathologists captured through medical microscopes. We then evaluated the compression performance and analyzed the signals of these pathological microscopic images by comparing them with general images. We found that pathological microscopic images have much wider chrominance-component spectrums (spatial frequency) than do general images. This indicates that when compressing pathological microscopic images, a large color difference is produced as a result of subsampling chrominance components. Such results have also been confirmed by subjective tests. Since the chrominance-component spectrums for general images have relatively narrow bandwidths, the conventional chrominance component subsampling formats 4:2:2 and 4:1:1 have been applied to raise the compression ratios. For pathological microscopic images, however, the compression ratio of irreversible coding has been forced

to approach the compression ratio of reversible coding, in order to suppress image degradation below detectable levels. Furthermore, the compression ratio of reversible coding for pathological microscopic images is lower than that for general images and, therefore, when comprehensively studied from the perspective of the cost of coding, decoding, storing, and transferring pathological microscopic images, it cannot be said that the current level of compression for these images is sufficient when using current JPEG algorithms.

We plan to test color transformation methods to raise compression ratios for pathological microscopic images, and also plan to test the effects of compression and chrominance-component subsampling on diagnostic accuracy.

REFERENCES

- [1] S. Ono, N. Ohta, and T. Aoyama, "All-digital super high definition images," *Signal Processing, Image Communication*, vol. 4, nos. 4, 5, pp. 429–444, Aug. 1992.
- [2] S. Akselsen, A. K. Eidsvik, and T. Folkow, "Telemedicine and ISDN," *IEEE Communicat. Mag.*, vol. 31, no. 1, pp. 46–51, Jan. 1993.
- [3] S. Ono, J. Suzuki, and N. Ohta, "Super high definition image coding for medical applications," presented at *26th Int. Conf. DSP*, Cyprus, June 1995.
- [4] J. Suzuki, "Application of super high definition images to teleradiology and telepathology," in *Proc. Digital Compression Technologies and Systems for Video Communications, SPIE*, Oct. 1996, vol. 2952, pp. 482–503.
- [5] A. Kayukawa, "Medical application of high-vision," *J. ITE Japan*, vol. 48, no. 3, pp. 251–253, Mar. 1994.
- [6] R. B. Arps and T. K. Truong, "Comparison of international standards for lossless still image compression," *Proc. IEEE*, vol. 82, no. 6, pp. 889–899, June 1994.
- [7] G. R. Kuduvali and R. M. Rangayyan, "Performance analysis of reversible image compression techniques for high-resolution digital teleradiology," *IEEE Trans. Med. Imag.*, vol. 11, no. 3, pp. 430–445, Sept. 1992.
- [8] P. C. Cosman, R. M. Gray, and R. A. Olshen, "Evaluating quality of compressed medical images: SNR, subjective rating, and diagnostic accuracy," *Proc. IEEE*, vol. 82, no. 6, pp. 919–932, June 1994.
- [9] K. Chen and T. V. Ramabadran, "Near-lossless compression of medical images through entropy-coded DPCM," *IEEE Trans. Med. Imag.*, vol. 13, no. 3, pp. 538–548, Sept. 1994.
- [10] J. Suzuki, I. Furukawa, A. Okumura, and T. Ashihara, "Performance analysis of compression techniques for pathological microscopic images," in *Proc. The European Symp. Advanced Network and Services*, Mar. 1995, vol. 2451, pp. 68–79.
- [11] I. Furukawa, K. Kashiwabuchi, and S. Ono, "Super high definition still/motion images digitizing system and standard test images," in *Proc. SPIE's Visual Communication and Image Processing '92*, Nov. 1992, vol. 1818, pp. 1545–1553.
- [12] M. Nakajima, "Very high quality digital test images for printing field," in *Proc. PCS'94*, Sept. 1994, vol. 5–09, pp. 138–141.
- [13] (JPEG) ISO CD 10918, "Digital compression and coding of continuous-tone still images," ISO/IEC JTC1/SC2/WG10, Jan. 1991.
- [14] G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulas*, 2nd ed. New York: Wiley, 1982.
- [15] R. W. G. Hunt, *Measuring Color*, 2nd ed. Chichester, U.K.: Ellis Horwood, 1991.
- [16] A. N. Netravali and B. G. Haskell, *Digital Pictures*. New York: Plenum, 1988.

An Active Scene Analysis-Based Approach for Pseudoconstant Bit-Rate Video Coding

Kuo-Chin Fan, Member, IEEE, and Kou-Sou Kan

Abstract—Digital video transmission across the network is now widely used in the Internet community. Visual information transmission will no doubt play an important role in a future national information infrastructure (NII) [1]. Very often, however, the channel bandwidth is fixed. Therefore, video coding based on an appropriate bit-rate control is a key to the success of accommodating visual communication in NII. In this paper, an extremely efficient rate-control algorithm for pseudoconstant bit-rate video coder is proposed. The proposed technique, adaptive scene analysis, intelligently collects required information to obtain estimated bit counts from the previous, present, and future frames. This collected intelligent information is then employed to govern the resource allocation so that better objective quality can be achieved under the bit-rate restriction. Combined with a series of dynamic rate-control schemes, the proposed algorithm demonstrates impressive quality transition between successive frames. In addition, the bit-rate irregularities caused by dramatic scene transition can be smoothed out in a very short period. A comparison with the MPEG-2 Test Model 5 is also presented in this paper.

Index Terms—Active sequence analysis, motion estimation, quantizer scaling factor, rate buffer control.

I. INTRODUCTION

THE RAPID development in very-large-scale-integration design, digital signal processing, and video compression in recent years stimulates the progress of visual communication. In the future national information infrastructure (NII) environment, video communication will play a major role in information transaction. NII applications include distribution of entertainment programming, educational information, and visualized information. Nevertheless, without appropriate video coding, delivery of massive video data will create serious problems in NII environment such as time delay, network congestion, video jitter, video damage, etc.

Motion-compensation-based hybrid discrete cosine transform (DCT)/interframe video compression algorithms have been proven to be successful in removing spatial and temporal redundancy of digitized video signals. The scheme has been widely adopted in the international standards, such as H.261 [2], [3], Moving Pictures Experts Group (MPEG) [4], [5],

and North America full-digital high-definition television [6]. However, the hybrid transform/interframe coding schemes will generate variable bit-rates (VBR's) intrinsically, where the number of the transmitted bits may change from time to time. In general, VBR is designed to accommodate constant quality. Moreover, VBR is easier to implement than the constant bit-rate (CBR) scheme. However, the practical transmission media usually exhibits fixed channel capacity. A large buffer is required to smooth out the variation in compressed data. The buffer delay varies according to the buffer size; larger buffer size will cause longer frame delay. However, smaller buffer size may result in unpleasant video quality during scene transition. The encoder output rate should be carefully regulated to adapt to existing channel limitation and produce pleasant visual quality. Therefore, an adaptive rate-regulation algorithm is essential for realistic visual communication applications. A successful rate-control algorithm should provide efficient channel resource utilization and maintain consistent video quality.

Rate control for the hybrid DCT/interframe video coding has been actively studied recently, where typical research topics are associated with the quantizer scaling factor adjustment. The adjustment is made according to the short-term buffer occupation and variation tendency state, local scene activity, and the establishment of the statistical rate model from test images [7]–[13]. In most of the algorithms, the coding parameters and quantizer scaling factors are determined based on information derived from coded data. The proposed rate-control strategy, adaptive scene analysis (ASA), is aimed at maintaining a smooth buffer variation and smaller buffer size. A large buffer can always result in a smooth bit stream. However, larger buffer size usually causes longer delay.

The rest of this paper is organized as follows. In Section II, we briefly review the coding scheme of the MPEG standard. A series of bit-rate control strategies are described in Section III. Details of the proposed ASA strategy and simulation results are given in Section IV. Last, concluding remarks are given in Section V.

II. OVERVIEW OF THE MPEG CODING

A. Primary MPEG Coding Strategy

In MPEG, each frame is coded in I-picture, B-picture, or P-picture, as sketched in Figs. 1–3. I-picture indicates intraframe coding. P-picture is forward-predictive coded using information from the previous P- (or I-) picture, and B-

Manuscript received June 10, 1996; revised September 10, 1997.

K.-C. Fan is with the Institute of Computer Science and Information Engineering, National Central University, Chung-Li, Taiwan 320, R.O.C.

K.-S. Kan is with the Institute of Computer Science and Information Engineering, National Central University, Chung-Li, Taiwan 320, R.O.C. He is also with ChungHwa Telecom Laboratories, Chung-Li, Taiwan 320, R.O.C. (e-mail: kcfan@ncu.ee.ncu.edu.tw).

Publisher Item Identifier S 1051-8215(98)02642-1.

1051-8215/98\$10.00 © 1998 IEEE

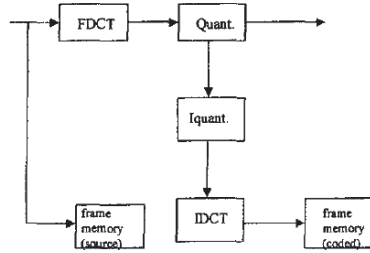


Fig. 1. Intra-coded picture type.

picture is bidirectional-predictive coded using information from the neighboring previous and future P- (or I-) pictures. To accommodate random access, a video sequence is partitioned into self-contained groups of pictures (GOP's) with N frames each. Usually, N is between 12 and 15. There are $(M-1)$ B-pictures between two consecutive P-pictures or between a P-picture and its preceding I-picture. Usually, the encoding sequence of the frames is different from the display order as indicated in Fig. 4.

The I-picture removes the spatial redundancy within the present frame. This coding type is independent of other frames. Hence, I-picture is used to encode the frame at the beginning of a GOP, and this picture coding type can also provide the functionality of periodical synchronization within a sequence to reduce the damages from the transmission errors. P-picture can be coded more efficiently by removing temporal domain redundancy from the previous I- or P-picture. On the other hand, B-picture achieves the highest degree of coding efficiency by using the neighboring previous and future pictures as reference for the motion estimation and motion compensation. Consequently, B-picture results in higher complexity and more delay than P-picture. Prediction is called forward if the reference frame is a previous frame and called backward if it is a future frame. B-picture is not referred to by any other prediction. Therefore, coding errors will be confined to B-frame and will not propagate. The detailed algorithm can be found in [4], [5], and [15].

B. MPEG-2 Test Model 5 (TM5) Rate-Control Mechanism [5]

The rate control of MPEG-2 TM5 regulates bit-rate from the coded information of previous pictures. Basically, three coding parameters are derived and used to adjust bit-rate.

1) *Average Macroblock Quantization Step Assignment and Readjustment*: The bit-counts assignment and adjustment is performed according to the global complexity, which is calculated based on GOP and is defined as follows:

$$X_I = S_I Q_I \quad (1)$$

$$X_P = S_P Q_P \quad (2)$$

$$X_B = S_B Q_B \quad (3)$$

where S_I , S_P , and S_B are the bit counts created by encoding the I-, P-, and B-pictures, respectively. Q_I , Q_P , and Q_B

are the global quantizer step size for I-, P-, and B-pictures, respectively. The subsequent target bit-counts assignment is defined based on GOP

$$T_I = \max \left\{ \frac{R}{1 + \frac{N_P X_P}{K_P X_I} + \frac{N_B X_B}{K_B X_I}}, \frac{\text{bit_rate}}{8 \times \text{picture_rate}} \right\} \quad (4)$$

$$T_P = \max \left\{ \frac{R}{N_P + \frac{N_B K_P X_B}{K_P X_P}}, \frac{\text{bit_rate}}{8 \times \text{picture_rate}} \right\} \quad (5)$$

$$T_B = \max \left\{ \frac{R}{N_B + \frac{N_P K_B X_P}{K_P X_B}}, \frac{\text{bit_rate}}{8 \times \text{picture_rate}} \right\} \quad (6)$$

where K_P and K_B are empirical constants, R is the number of available bits in the present group of pictures before subsequent coding, and N_P , N_B are the number of pictures remaining in the present group of pictures.

The number of R , N_P , and N_B will be reduced after each coding, and the number will be reset for a new group of pictures.

2) *Buffer Compensation Factor*: The buffer fullness status is predicted as

$$d_{I,j} = d_{I,0} + B_{j-1} - \frac{T_I \times (j-1)}{MB_{\text{number}}} \quad (7)$$

$$d_{P,j} = d_{P,0} + B_{j-1} - \frac{T_P \times (j-1)}{MB_{\text{number}}} \quad (8)$$

$$d_{B,j} = d_{B,0} + B_{j-1} - \frac{T_B \times (j-1)}{MB_{\text{number}}} \quad (9)$$

where $d_{I,0}$, $d_{P,0}$, and $d_{B,0}$ are the initial buffer contents for I, P, and B pictures, individually, B_{j-1} is the number of bit counts created before macroblock j , and MB_{number} is the total number of macroblocks within a picture. The quantizer step size of macroblock j is adjusted according to

$$Q_j = \frac{31 \times d_j}{2 \times \text{bit_rate/picture_rate}} \quad (10)$$

3) *Macroblock Quantizer Step Size Fine Adjustment*: The local spatial activity of macroblock j is first computed as

$$\text{act}(j) = 1 + \min_{\text{block}} [\text{variance}(j)] \quad (11)$$

Then normalized local activity of macroblock j is defined as

$$\text{Normalized_act}(j) = \frac{2 \times \text{act}(j) + \text{average_act}}{\text{act}(j) + 2 \times \text{average_act}} \quad (12)$$

Last, the final quantizer step size of macroblock j is calculated as

$$m_{\text{quant}_j} = Q_j \times \text{Normalized_act}(j) \quad (13)$$

where the average_act indicates the average spatial activity of the previous coded picture. Detailed MPEG-2 TM5 rate-control strategy can be found in [5].

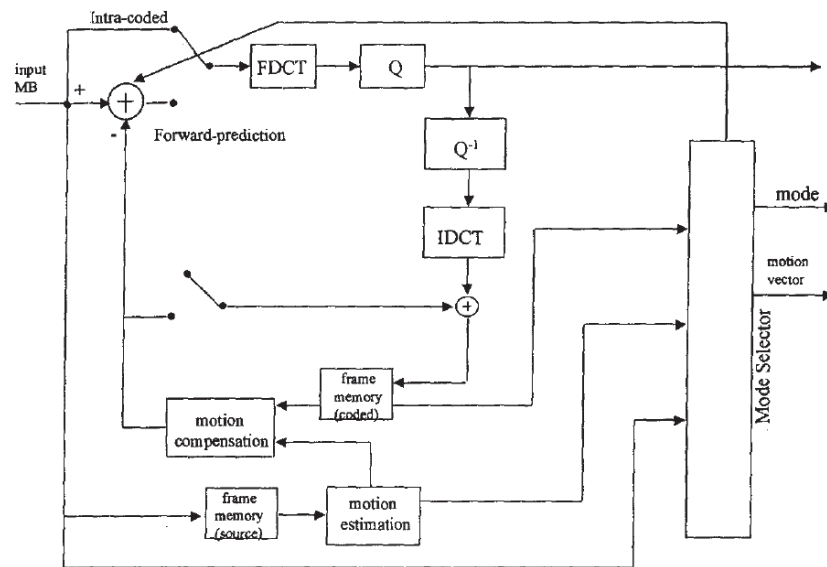


Fig. 2 Functional block diagram for P-picture coding type.

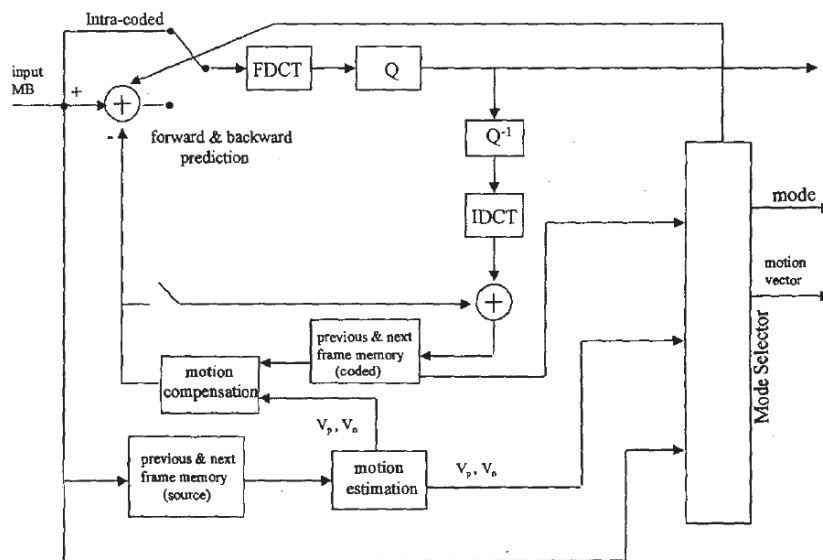


Fig. 3. Functional block diagram for B-picture coding type.

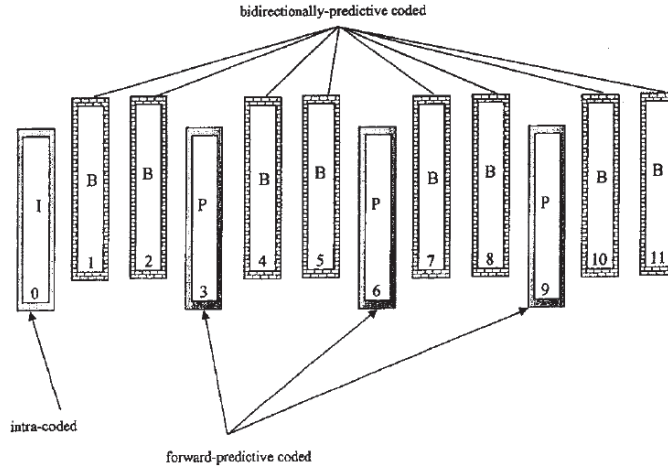


Fig. 4. MPEG group coding structure in display order.

III. ACTIVE SEQUENCE ANALYSIS STRATEGY

As mentioned earlier, the compressed bit stream created by the MPEG-2 TM5 rate-control algorithm is entirely based upon the coding history. Generally, a video sequence with complicated variation is difficult to code based on limited historical knowledge. The proposed technique, ASA, collects the required coding information to estimate bit counts from the preceding, present, and future frames. The proposed technique extends the knowledge collection basis to improve the efficiency of rate control.

The proposed algorithm includes bit-rate prediction of variable length coding (VLC) and active forward-looking and backward-looking quantizer.

A. Bit-Rate Prediction Model of VLC Coding

The video buffer occupation needs to be controlled via macroblock quantizer adjustment so that the desired bit-rate can be achieved. Consequently, characterizing the bit counts of different picture coding types is a key issue in rate control. Bit-rate prediction of VLC coding proceeds after the transformed prediction error has been generated. Two VLC simulators are proposed: one to estimate the rate for DCT coefficients and one to estimate the rate for motion vectors.

Let $slice_{acp}$ denote the average activity within slice. The average activity of slice m is defined as follows:

$$slice_{acp}(m) = c_0 \times \left\{ \sum_j AP(j) \right\} / mb_width$$

$$AP(j) = \left\{ \sum_k block_sum_j(k) \right\} / 4$$

$$block_sum_j(k) = \left(\left\{ \sum_i abs[block_k(i)] \right\} / 63 \right),$$

$$j = [m \times mb_height, (m+1) \times mb_height - 1],$$

$$k = [0, 3],$$

$$i = [1, 63] \quad (14)$$

where $block_k$ denotes the transformed coefficient of the 8×8 DCT block of macroblock j , $AP(j)$ denotes the average activity of macroblock j , c_0 is an empirical constant, mb_width denotes the macroblock number within each slice, and mb_height denotes the total number of slices in the input picture.

The first VLC simulator is aimed at imitating the functionality of the VLC encoding scheme by using the VLC tables defined in the MPEG standard [4], which is used to encode the transformed coefficients. Each 8×8 transformed block data is first normalized by slice activity $slice_{acp}$ as defined in (14). Both luminance and chrominance components are considered for bit-counts estimation. The purpose of normalization is to reduce the potential activity irregularity possessed among the slice. As defined in the MPEG standard VLC table [4], each VLC code word is referred to a certain combination of zero run and level magnitude. The bit count for each normalized and quantized DCT block can then be obtained by using the MPEG VLC table. Therefore, we can create a code word length reference table to represent the particular zero run and level combination. For example, the corresponding code word is "0000 0001 0010 0" for the zero run and level combination (4,3) according to MPEG VLC table [4]. The bit count of this code word is 13. The last bit in the preceding code word is sign bit. As described previously, no actual VLC coding action is performed except for a simple table lookup to obtain

the word length. Therefore, the extra computational effort can be ignored.

The second VLC simulator is used to estimate the bit count of the motion vectors. The motion vectors are extracted after the motion estimation. As recommended by the MPEG standard, a first-order differential pulse code modulation is further employed to produce the differential vectors. The proposed VLC simulator for motion vectors estimates the bit count of these differential vectors by using the MPEG VLC vector codeword table.

When there exists high temporal variation such as multiple irregular moving objects or scene change among the consecutive frames, the prediction errors between the successive frames may increase significantly. The actual bit counts corresponding to the transformed prediction errors and the motion vectors may be substantially higher. Consequently, we developed the following scheme to improve the accuracy for dynamic scenes.

B. Active Forward- and Backward-Looking Macroblock Quantizer Assignment

Let I_d , P_d , and B_d denote the desired bit counts for I, P, and B pictures, respectively, which are defined as

$$\begin{aligned} I_d &= G_I * R_d / I_N \\ P_d &= G_P * R_d / P_N \\ B_d &= G_B * R_d / B_N \end{aligned} \quad (15)$$

where R_d indicates the desired channel rate, G_I , G_P , and G_B are the predefined bit-count ratio for I-, P-, and B-pictures in each second, and I_N , P_N , and B_N denote the number of pictures for each picture coding type per second.

Let r_I , r_P , and r_B denote the deviation between the desired bit counts I_d , P_d , and B_d and the estimated bit counts I_P , P_P , and B_P for I-, P-, and B-pictures, respectively

$$\begin{aligned} r_I &= (I_P - I_d) / I_d \\ r_P &= (P_P - P_d) / P_d \\ r_B &= (B_P - B_d) / B_d. \end{aligned} \quad (16)$$

Estimated bit counts I_P , P_P , and B_P can be obtained from the two VLC simulators, which are defined in Section III-A.

As indicated in (16), the coding of I-picture is independent of any other picture coding type. Therefore, r_I value can be used to represent the spatial variation of the scene. Basically, the picture contents are similar within a short time period. Therefore, r_I can be further considered as an indication of the spatial domain variation of the group of pictures. A large positive r_I value usually indicates high spatial activities, and the small positive or negative r_I value indicates typical spatial variation within the scene.

Since I-picture is coded without using any information from its neighboring pictures, the scene transition cannot be identified from I-picture. Motion estimation and compensation has to be exploited to remove the temporal redundancy of P- and B-pictures. Therefore, both r_P and r_B are used to derive the scene transition information. Usually large r_P and r_B indicate violent scene changes. Since this proposed scheme

utilizes temporal transition information for rate control, it is superior to traditional approach, which uses historical statistics based on a single frame.

Here, the macroblock quantizer assignment is defined before the actual coding operation is initiated. For I-, P-, and B-pictures, several parameters obtained *a priori* are considered in macroblock quantizer assignment. These parameters include r_I , r_P , r_B , $slice_{acp}$, AP , and buffer fullness status.

Let f_0 , f_P , and D_P denote the frame number of the first I-picture, the frame number of P-picture, and the distance between neighboring P-pictures in the related group, respectively. For input picture f , the quantizer step $Q(j)$ of macroblock j within slice m is defined as follows:

$$\begin{aligned} &\text{for I-picture} \\ &\quad Q(j) = act_index(j) \times r_I(f) + fine_adjust(f) \\ &\text{for P-picture} \\ &\quad Q(j) = act_index(j) \times [r_I(f_0) + (r_{PP} + r_{PL})/2] \\ &\quad \quad + fine_adjust(f) \\ &\quad \quad r_{PP} = r_P(f_P) \\ &\quad \quad r_{PL} = r_P(f_P - D_P) \\ &\text{for B-picture} \\ &\quad Q(j) = act_index(j) \times [r_I(f_0) + (r_{PP} + r_{PL})/2] \\ &\quad \quad + r_B(f)] + fine_adjust(f) \end{aligned} \quad (17)$$

where

$$\begin{aligned} act_index(j) &= [AP(j) + slice_{acp}(m)]/2 \\ R_c(f-1) &= codedbits(f-1) + I_d \times N_{Ir}(f-1) \\ &\quad + P_d \times N_{Pr}(f-1) + B_d \times N_{Br}(f-1) \\ &\text{if } \{[R_c(f-1)/R_d] > 1\} \\ &\quad fine_adjust(f) = 2 \times \lfloor R_c(f-1)/R_d \rfloor \\ &\text{else} \\ &\quad fine_adjust(f) = (-1) \times 2 \times \lfloor R_d/R_c(f-1) \rfloor. \end{aligned} \quad (18)$$

For P-picture, r_{PP} and r_{PL} denote the estimated bit-rate deviation of the present and previous P-pictures, respectively. Here, R_c denotes the estimated bit-rate after previous picture coding, $codebits$ represents the bits generated after previous picture coding, and N_{Ir} , N_{Pr} , and N_{Br} denote the numbers of unprocessed pictures of I-, P-, and B-pictures after previous picture coding, respectively.

According to the experimental data, the actual bit count of coded picture can be properly estimated by the proposed strategy with a simple scaling factor. The contents of I-, P- and B-picture are considered as uniform variation if the predicted bit-count deviations r_I , r_P , and r_B are smaller than one, two, and three, respectively. The minimum values of r_I , r_P , and r_B are empirically determined to be one.

For I-pictures, the factors that affect the macroblock quantizer step $Q(j)$ consist of slice average activity index $slice_{acp}$, local activity index $AP(j)$, estimated bit-rate variation r_I , and $fine_adjust$. If the picture exhibits high spatial activity, a large value of r_I and activity index act_index will result in a coarse quantizer step $Q(j)$ to reduce the bit-rate. Otherwise, finer quantization is employed to retain the picture quality.

For P-picture, the determination of macroblock quantizer requires the estimated information from previous I- and P-pictures and present P-picture. The predicted variation of previous I- and P-pictures can contribute to present P-picture by means of r_I and r_{PL} . If there exists high spatial activity scene content and scene transition, the values of r_I , r_{PL} , and activity index act_index will exhibit large values. Therefore, a coarse quantizer will be assigned to reduce the rate. Otherwise, quantization will be compensated by distinct weighting factors of r_I , r_{PL} , and activity index act_index . The integrated resulting quantizer $Q(j)$ is designed to preserve both the picture fidelity and bit-rate.

As to the B-picture, the determination of quantizer consists of the estimated information of its previous and future I- and P-pictures and present B-picture. If there exists violent scene variation and scene transition in any neighboring I- and P-pictures, and present B-picture, the values of r_I , r_{PL} , r_{PP} , and r_B will exhibit large values. Then, coarse quantization will be used to reduce the rate. The rate variation, which is the ratio between the estimated bit-rate R_C and target bit-rate R_d , is further considered in the quantization. If the estimated bit-rate R_C is higher than the target bit-rate R_d , then the quantizer step is slightly increased, as shown in (18). Otherwise, the quantizer step is slightly decreased.

A further evaluation strategy, which is implemented based on the buffer variation within each slice, is proposed to fine adjust $Q(j)$. For input macroblock j of slice m , the final macroblock quantizer $Mquant(j)$ is defined as

$$\begin{aligned}
 & \text{if}(\text{slice_bit_counts}(m-1) > \text{slice_data_rate}) \\
 & \quad Mquant(j) \\
 & \quad = Q(j) + (\text{buf_content}_{\text{slice}}(m-1)/\text{slice_data_rate}) \\
 & \quad \text{else} \\
 & \quad Mquant(j) \\
 & \quad = Q(j) + (-1) \\
 & \quad \quad \times (\text{buf_content}_{\text{slice}}(m-1)/\text{slice_data_rate}) \\
 & \quad \text{for I - picture} \\
 & \quad \quad \text{slice_data_rate} = I_d/\text{mb_height} \\
 & \quad \text{for P - picture} \\
 & \quad \quad \text{slice_data_rate} = P_d/\text{mb_height} \\
 & \quad \text{for B - picture} \\
 & \quad \quad \text{slice_data_rate} = B_d/\text{mb_height}
 \end{aligned} \quad (19)$$

where $\text{slice_bit_count}_{\text{previous}}$ denotes the generated bit counts of the previous slice, $\text{buf_content}_{\text{slice}}$ denotes the buffer fullness status after the previous slice has been coded and counted in bit, and mb_height denotes the total slice number within a picture. The purposes of slice-based buffer content evaluation and related quantizer fine adjustment defined previously are: 1) to achieve desired and stable rate within each picture; 2) to avoid violent scene variation; and 3) to achieve consistent picture quality.

The purpose of the constraints defined in this section is not only to maintain a steady buffer state but also to maintain a smooth quality transition. Basically, the quantization of I-, P-, and B-picture is based on estimated information as defined

TABLE I
PARAMETERS USED IN TM5 RATE CONTROL
AND ACTIVE SCENE ANALYSIS ALGORITHM

Parameters used in TM5 ratecontrol	
1. $K_p=1.0$, $K_B=1.4$;	
2. $\text{bit_rate}=1.3$ Mbps	
3. MPEG (13,3) coding structure for 1 st group	
(15,3) coding structure for the rest of the sequence:	
4. initial bit allocation for I=162500 bits	
	P=58500 bits
	B=25350 bits
5. $\text{picture_rate}=30$	
Parameters used in proposed Active Scene Analysis Strategy	
1. $c_0=2.5$ for I-picture,	
= 3 for P and B-picture	
2. $G_I=0.25$, $G_P=0.36$, $G_B=0.39$;	
3. buffer size = 200 ms	
4. target bit rate = 1.2 Mbps and 1.3 Mbps	
5. motion estimation: full search with variable search window size	
6. MPEG (13,3) coding structure for 1 st group	
(15,3) coding structure for the rest of the sequence	

in (14)–(19). Since the quantizer scaling factors of the target frame are globally adjusted, the quality will be homogeneous within the scene.

IV. SIMULATION RESULTS

In the simulation, the video sequence is digitized according to the CCIR 601 format [14] with a color sampling ratio of 4:2:2. Each frame is low-pass filtered down to 352 (H) by 240 (V) [15]. The color sampling ratio is converted to 4:2:0. The luminance and chrominance samples are digitized into 8 bits/sample. Therefore, the precision of each pixel is 16 bits. The content of the video sequence for simulation is a sequence from a movie. The picture rate is simulated at 30 frames/s, and the target video bit-rate is chosen to be 1.20 Mbps including all necessary header bits in the MPEG syntax. The default quantization matrices of the MPEG standard are applied to both the luminance and chrominance components [15].

The full-search algorithm from public-domain software [16] is used to compute the motion vectors in the proposed strategy. The buffer allocated for I-, P-, and B-pictures is defined as a certain ratio of the given channel rate. Table I is a summary of all background parameters assigned in the simulation.

To collect essential parameters for compression, the global encoding procedure has been reorganized and the data flow has also been rescheduled to suit our special purpose. This new arrangement facilitates the acquisition of the necessary intelligence proposed in this paper for appropriate decision making. The operations of the incoming video frames are described as follows.

- 1) The first group of pictures adopt MPEG (13,3) coding structure. Coding structure (15,3) is chosen for the rest of the video sequence.
- 2) According to picture coding type, perform the full search to extract motion vectors.

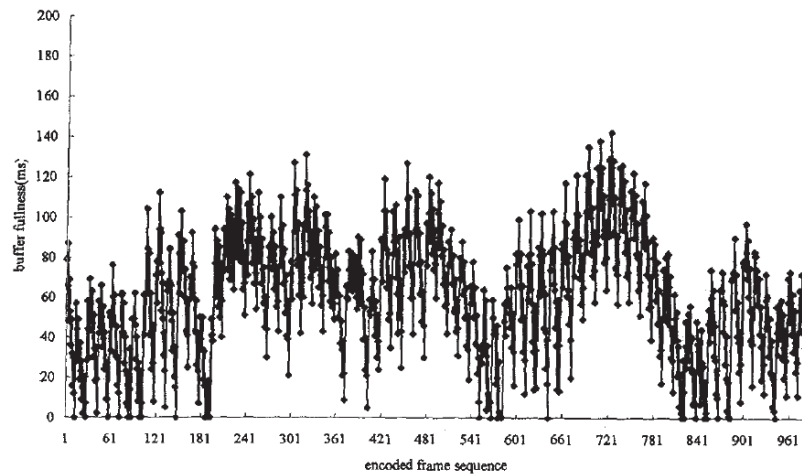


Fig. 5. Buffer fullness status of ASA at 1.2 Mbps with buffer size 200 ms.

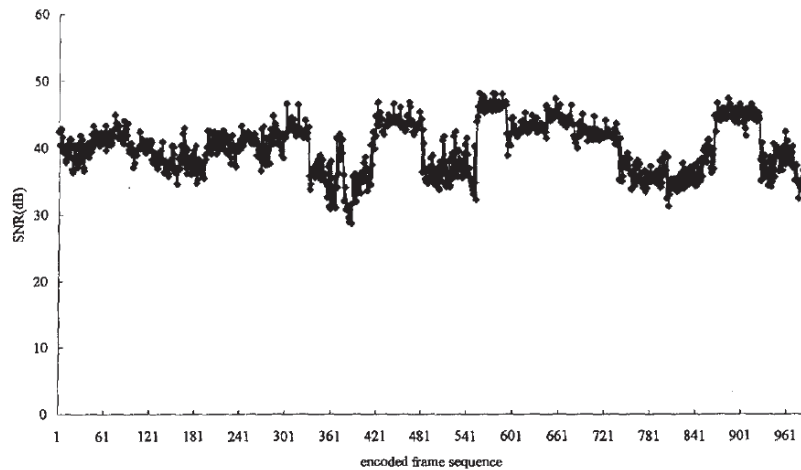


Fig. 6. SNR of ASA at 1.2 Mbps with buffer size 200 ms.

- 3) A macroblock coding type decision-making mechanism, which is used in the MPEG simulation model standard [15], is used to determine the coding type.
- 4) Two-dimensional DCT is applied to block data according to the macroblock coding type.
- 5) The macroblock and slice activity information is computed according to (14).
- 6) Estimate bit counts of each picture coding type from VLC simulators and calculate bit-count deviations based on (15) and (16).
- 7) Macroblock quantizer step of entire picture is determined according to (17) and (18) and is further adjusted from (19) after each VLC coding step.

The initial r_{PL} for the forward-predictive coded frame is set to one. The buffer fullness variation profiles, signal-to-noise ratio (SNR), and frame bit counts are shown in Figs. 5–7. It should be noted that the horizontal x axis in all figures represents the encoded frame sequence instead of the nature scene order. The buffer fullness is defined as

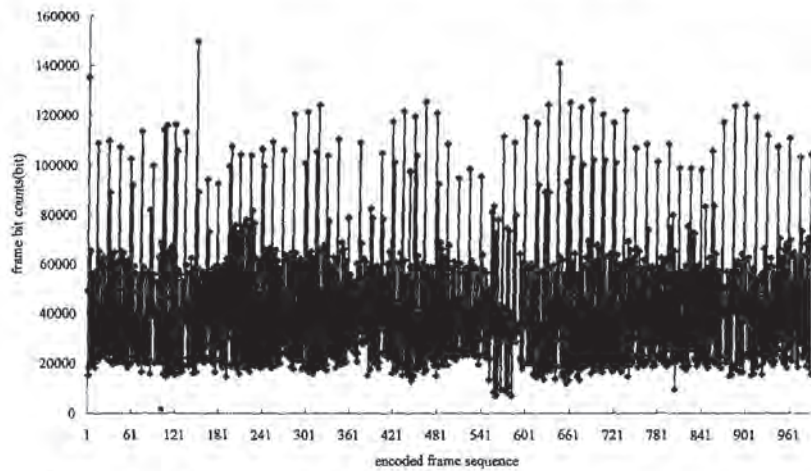


Fig. 7. Frame bit counts of ASA at 1.2 Mbps with buffer size 200 ms.

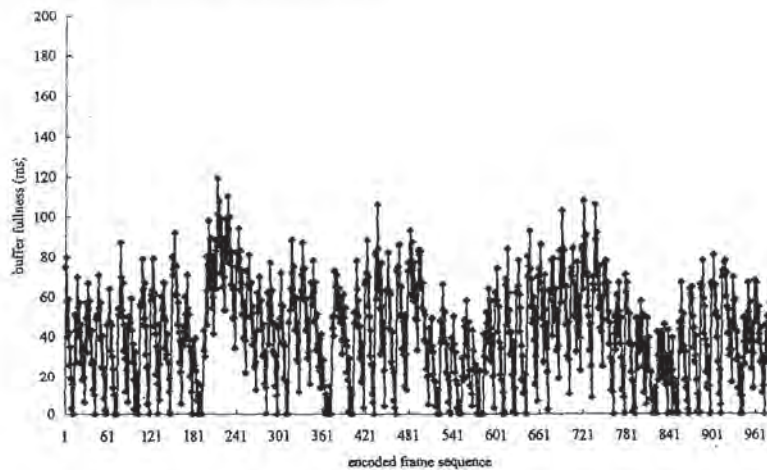


Fig. 8. Buffer fullness of ASA at 1.3 Mbps with buffer size 200 ms.

follows:

$$\begin{aligned}
 \text{buffer}(f) &= \text{buffer}(f-1) - \text{data_rate}_{\text{frame}} \\
 &\quad + \text{coded_rate}(f) \text{data_rate}_{\text{frame}} \\
 &= R_d / \text{frame_rate}.
 \end{aligned} \tag{20}$$

In the test video sequence, there exist several scenes. The resulting buffer status in the proposed algorithm is well confined within a desired level. The reconstructed pictures are uniformly varied under violent scene transition. Besides, the target bit-rate is under the desired rate. Note that the buffer overshoot condition is avoided by the proposed algorithm. As to the buffer undershot, the frame bit counts of each picture coding type as sketched in Fig. 7 in the proposed algorithm are

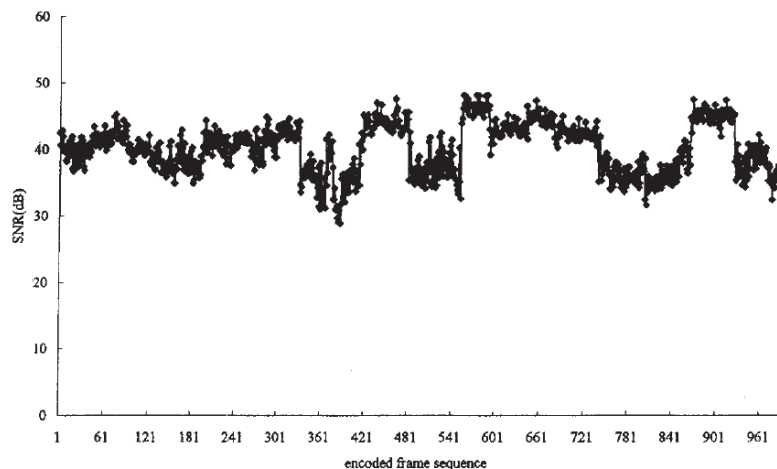


Fig. 9. SNR of ASA at 1.3 Mbps with buffer size 200 ms.

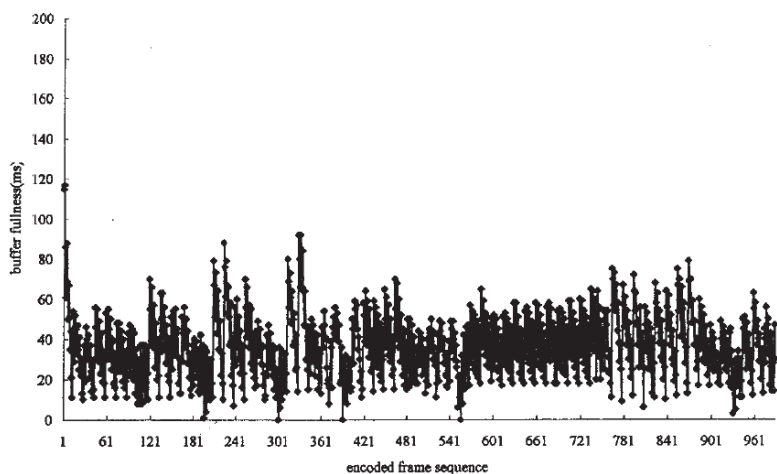


Fig. 10. Buffer fullness of TM5 rate control at 1.3 Mbps.

very close to the assigned number. Therefore, a small amount of stuffing bits can be used to maintain the buffer content.

MPEG-2 TM5, which is an excellent rate-control algorithm in the public domain, is chosen to compare with the proposed algorithm. The target bit-rate used here is 1.3 Mbps. The coding results of the proposed technique, ASA, are outlined in Figs. 8 and 9. The coding results of TM5 are outlined in Figs. 10 and 11. The necessary parameters used in the proposed technique, ASA, and TM5 are listed in Table 1. As

sketched in Fig. 12, the buffer requirement of TM5 is slightly smaller than the proposed technique. However, the resulting bit-rate of TM5 is higher than 1.3 Mbps and is unstable. The resulting bit-rate is slightly lower than 1.3 Mbps for the proposed algorithm and is confined within small range. Note that under the frame-by-frame-based SNR comparison, both algorithms exhibit minor differences. A further comparison for a 30-frame period containing violent scene transitions with buffer fullness and SNR is illustrated in Figs. 13 and 14. These

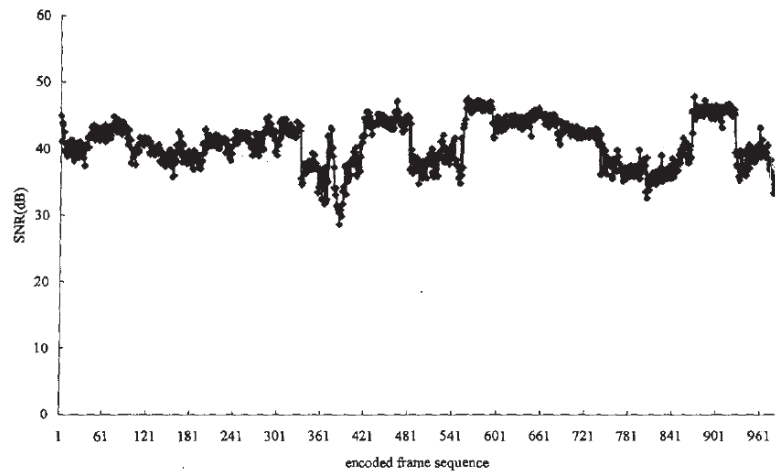


Fig. 11. SNR of TM5 rate control at 1.3 Mbps.

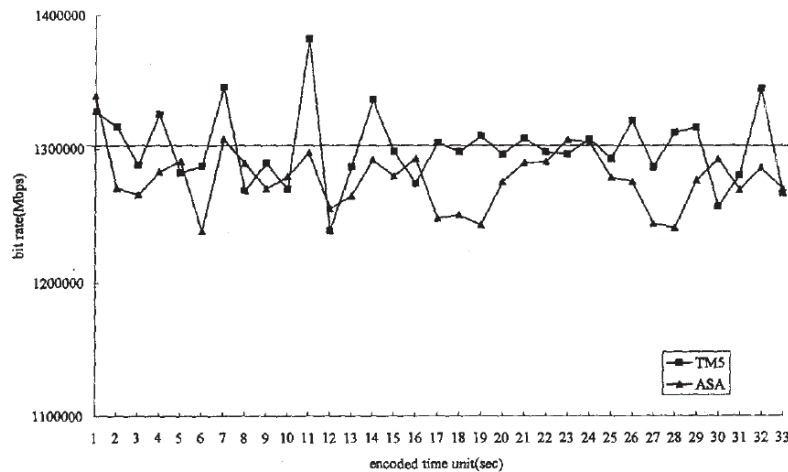


Fig. 12. Resulting bit-rate of TM5 and ASA at 1.3 Mbps.

frames are selected from 360 to 390 in the display order. As illustrated in Figs. 13 and 14, the buffer variation of both algorithms exhibits similar behavior. Moreover, the difference in SNR is minor between the two algorithms.

V. CONCLUSIONS

In this paper, we propose a unique rate-control scheme for DCT/interframe-based video coding by exploiting scene activity adaptively. This strategy is MPEG-standard compliant

since only different operational schedules and minor computational efforts are needed without violating the MPEG syntax. The simulation results indicate that the proposed scheme results in a more stable output rate and a more consistent perceptual quality. In conventional rate control strategies, the rate adjustment is determined solely by the historical record. It usually results in more fluctuations in decoded picture quality. Compared with the MPEG TM5 algorithm, the proposed strategy shows the equivalent or even better quality with

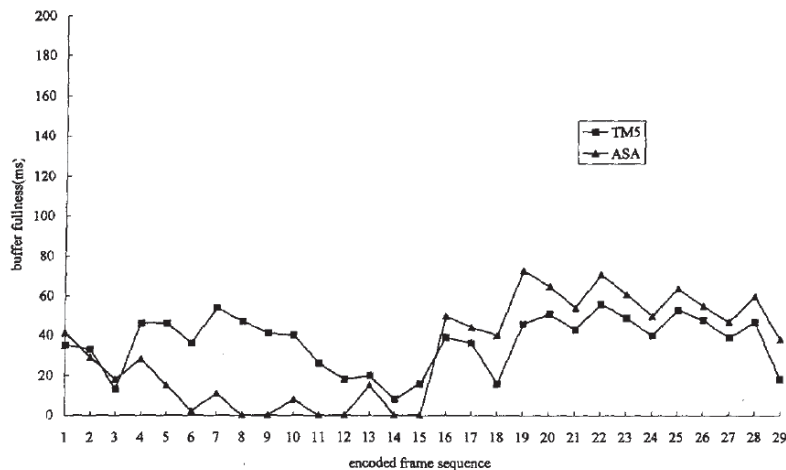


Fig. 13. Buffer fullness of TM5 and ASA at frame 360-390.

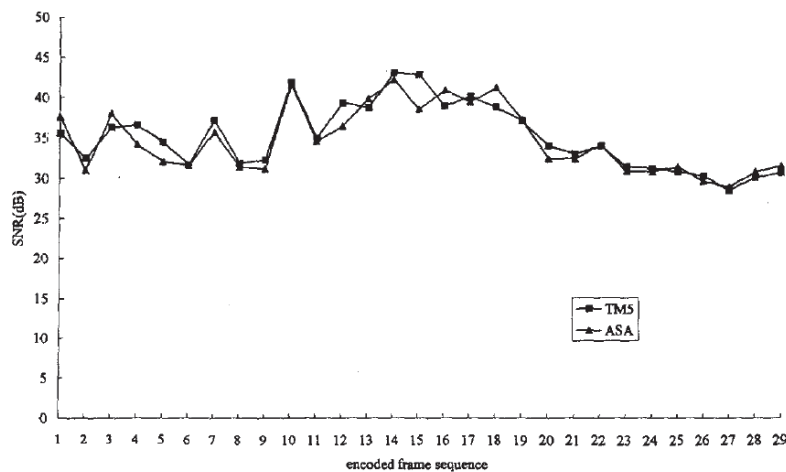


Fig. 14. SNR of TM5 and ASA at frame 360-390.

less resulting bit-rate. The proposed scheme also results in well-controlled buffer fullness and alleviates overflow and underflow problems. Furthermore, the annoying picture quality degradation during scene transition is greatly reduced by the scene analysis.

REFERENCES

- [1] NII Workshop Program Committee, "Summary report on the workshop on advanced digital video in the national information infrastructure," Washington, DC, May 10-11, 1994.
- [2] CCITT SG XV, "Draft revision of recommendation H.261," Doc. 572, Mar. 1990.
- [3] —, "Description of reference model 8 (RM8)," Document 525, June 1989.
- [4] ISO/MPEG, "Committee draft for MPEG video coding standard," ISO/IBC 11172-2, Nov. 1993.
- [5] ISO/MPEG II, "Test model 5," Doc. AVC-491, Apr. 1993.
- [6] Advisory Committee on Advanced Television Services Experts Groups, "Grand alliance HDTV system specification," version 2.0, Dec. 7, 1994.
- [7] M. R. Pickering and J. F. Arnold, "A perceptually efficient VBR rate control algorithm," *IEEE Trans. Image Processing*, vol. 3, pp. 527-532, Sept. 1994.

- [8] A. Puri and R. Aravid, "Motion-compensated video coding with adaptive perceptual quantization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 1, pp. 351-361, Dec. 1991.
- [9] N. B. Nilb, "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Trans. Commun.*, vol. COM-33, pp. 551-557, June 1985.
- [10] J. H. Choi and D. C. Park, "A stable feedback control of buffer state using the controlled Lagrange multiplier method," *IEEE Trans. Image Processing*, vol. 3, pp. 546-558, Sept. 1994.
- [11] C. T. Chen and A. Wang, "A self-governing rate buffer control strategy for pseudocostant bit-rate video coding," *IEEE Trans. Image Processing*, vol. 2, pp. 50-59, Jan. 1993.
- [12] J. Zdepski, D. Raychaudhuri, and K. Joseph, "Statistically based buffer control policies for constant rate transmission of compressed digital video," *IEEE Trans. Commun.*, vol. 39, pp. 947-957, June 1991.
- [13] A. Wang, C. T. Chen, D. J. L. Gall, F. C. Jeng, and K. M. Uz, "MCPIC: A video coding algorithm for transmission and storage applications," *IEEE Commun. Mag.*, pp. 24-32, Nov. 1990.
- [14] CCIR Recommendation 601, "Encoding parameters of digital television for studios," in *CCIR recommendation and reports*, vol. XI, International Telecommunications Union, Geneva, Switzerland, 1982.
- [15] ISO/MPEG, "MPEG video simulation model three (SM3)," Draft 1, MPEG90.
- [16] MPEG Software Simulation Group, "MPEG-2 video encoder," version 1.0, May 1994.

Kuo-Chin Fan (S'88-M'88) was born in Hsinchu, Taiwan, R.O.C., on June 21, 1959. He received the B.S. degree in electrical engineering from National Tsing-Hua University, Taiwan, in 1981 and the M.S. and Ph.D. degrees in electrical engineering from the University of Florida, Gainesville, in 1985 and 1989, respectively.

In June 1983, he was with the Electronic Research and Services Organization, Taiwan, as a Computer Engineer. From 1984 to 1989, he was a Research Assistant with the Center for Information Research at the University of Florida. In 1989, he joined the Institute of Computer Science and Information Engineering at National Central University, Taiwan, where he became a Professor in 1994. Currently, he is the Chairman of the Computer Science and Information Engineering Department. His current research interests include image analysis, optical character recognition, and machine intelligence.

Dr. Fan is a member of the International Society for Optical Engineering.

Kou-Sou Kan was born in Taipei, Taiwan, R.O.C., on May 17, 1963. He received the B.S. and M.S. degrees in atmospheric physics from National Central University, Taiwan, in 1985 and 1987, respectively. He is working toward the Ph.D. degree at the Institute of Computer Science and Information Engineering, National Central University, Taiwan.

In October 1987, he joined Telecommunications Laboratories, Chung-Hwa Telecom Co., Ltd., in Chung-Li, Taiwan, as an Assistant Researcher. There, he was engaged in digital image compression, algorithm, and architecture. From 1995 to 1996, he was an Exchange Visitor with Bell Communication Research (Bellcore), Red Bank, NJ, where he was engaged in the establishment of the testbed of Digital Storage Media Command and Control, real-time MPEG video delivery over local-area networks, MPEG bit-rate control, and MPEG-2 transcoding. In July 1997, he became a Researcher with Telecommunication Laboratories. His research interests include digital image/video compression, real-time digital video transmission over network, video on demand (VoD), and MPEG-2 video transcoding. His current work is to establish the xDSL-based "Interactive MultiMedia Trial Project" to provide near VoD, Internet access, and live TV broadcasting for the traditional local-loop customer of ChungHwa Telecom Co., Ltd., at Taipei.

Image Segmentation and Contour Detection Using Fractal Coding

Takashi Ida, Member, IEEE, and Yoko Sambonsugi

Abstract—Fractal coding was applied to image segmentation and contour detection. The encoding method was the same as in conventional fractal coding, and the compressed code, which we call fractal code, was used for image segmentation and contour detection instead of image reconstruction. An image can be segmented by calculating the basin of attraction on a mapping that is a set of local maps from the domain block to the range block. The local maps are parameterized using the fractal code, and contours of the objects in the image are detected by the inverse mapping from the range block to the domain block. Some objects in the test image Lena were segmented, and the contours were detected well. The proposed methods are expected to enable compressed codes to be used directly for image processing.

Index Terms—Chaos, fractals, image coding, image edge analysis, image segmentation.

I. INTRODUCTION

DUE to the increased use of digital image processing, image compression has become an important technology for the transmittal and storage of images. For lossy compression of gray-scale images, the discrete cosine transform is used widely, and it has been adopted in some international standards for image compression. There are many coding methods that have been researched actively, including vector quantization, subband coding, and wavelet transformation.

Fractal image coding is a novel method that was originally proposed by Barnsley [1] as a compression method for binary images, and applied to gray-scale images by Barnsley [2], [3] and Jacquin [4], [5]. Fractal coding has received considerable attention since it was proposed [3], [4] because of its use of the self-similarity in images, an innovation that had not been used previously in the field of image compression. Many improvements in compression efficiency and decoded image quality have been reported [6]–[8].

This paper presents new functions of fractal coding which go beyond just compression. If images have been compressed by fractal coding, segmented images and contour-detected images can be produced directly from the compressed code without decoding the images [9]. The encoding algorithm is the same as in the conventional method, but the compressed code (which we call fractal code) is used not only for image reconstruction, but also for segmentation and contour detection. Since the image representations referred to above are

used in many image processing systems, the new functions are expected to expand the applications of fractal coding.

Fractal coding is summarized in Section II. In Section III, dynamical systems in which parameters are fixed by self-similarity information in the fractal code are defined, and image representation methods using limit sets of the systems are proposed. The experimental results are presented in Section IV, and finally, the conclusions are outlined in Section V.

II. FRACTAL IMAGE CODING

In fractal image coding, a gray-scale image which supports $A \subset R^2$ is divided into domain blocks D_i ($i = 1, \dots, I$, where I is the number of domain blocks) as shown in Fig. 1, and range blocks R_i are found for every D_i . Replacement of the intensity value in D_i with the transformed value from R_i is called fractal transformation [3]. The fractal transform operator is defined as

$$(F\mu)(x, y) = v_i(\mu(f_i(x, y))) \quad \text{when } (x, y) \in D_i. \quad (1)$$

Here, $\mu(x, y)$ is the intensity at $(x, y) \in A$, and v_i , which is an affine transformation of intensity z , is defined as

$$v_i(z) = p_i z + q_i. \quad (2)$$

The affine map f_i from D_i to R_i is defined as

$$\begin{aligned} f_i(x, y) = & (r_i(x - x_i) \cos \theta_i - r_i(y - y_i) \sin \theta_i + s_i + x_i \\ & r_i(x - x_i) \sin \theta_i + r_i(y - y_i) \cos \theta_i + t_i + y_i) \end{aligned} \quad (3)$$

where x_i and y_i are the coordinate values of the center of D_i . The following simplified notation is used to represent (3):

$$f_i \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r_i \cos \theta_i & -r_i \sin \theta_i \\ r_i \sin \theta_i & r_i \cos \theta_i \end{pmatrix} \begin{pmatrix} x - x_i \\ y - y_i \end{pmatrix} + \begin{pmatrix} s_i + x_i \\ t_i + y_i \end{pmatrix}. \quad (4)$$

The affine map f_i performs rotation by θ_i , centered on (x_i, y_i) , enlargement by r_i , and translation by (s_i, t_i) .

Fig. 2 shows the process described in terms of the right-hand side of (1). $(x, y) \in D_i$ is mapped to $f_i(x, y) \in R_i$ by (4), and $\mu(f_i(x, y))$, which is the intensity value at $f_i(x, y)$, is transformed by (2). In addition, the intensity value at (x, y) is replaced by $v_i(\mu(f_i(x, y)))$. The transformation parameters $r_i, \theta_i, s_i, t_i, p_i$, and q_i are compressed codes (fractal codes).

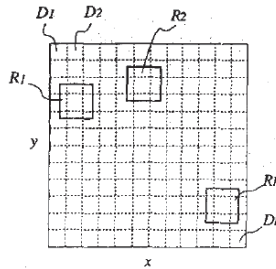
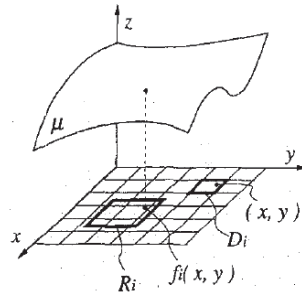
When each R_i is larger than D_i and each v_i is contractive

$$r_i > 1 \quad (5)$$

Manuscript received February 6, 1996; revised April 28, 1997 and April 15, 1998. This paper was recommended by Associate Editor L. A. Ferrari.

The authors are with the Laboratory VI, Communication and Information Systems Research Laboratories, R&D Center, Toshiba Corporation, 1, Komukai-Toshiba-cho, Saiwai-ku, Kawasaki 210, Japan.

Publisher Item Identifier S 1051-8215(98)07873-2.

Fig. 1. Domain blocks D_i and their range blocks R_i .Fig. 2. Point $(x, y) \in D_i$ is mapped to $f_i(x, y) \in R_i$ by (4), and the intensity value at $f_i(x, y)$: $\mu(f_i(x, y))$ is sampled in the fractal transform.

and

$$|p_i| < 1. \quad (6)$$

We can obtain an image ϕ by calculating

$$\phi = \lim_{n \rightarrow \infty} F^n(\alpha) \quad (7)$$

where α is an arbitrary image and F^n represents the n th iteration of F . The distance $d(\psi, \phi)$ between a given image ψ and ϕ is bounded according to the inequality

$$d(\psi, \phi) \leq \frac{d(\psi, F(\psi))}{1 - s} \quad (8)$$

where s is the contractivity factor of F . This is called the Collage theorem [3].

Fig. 3 is the decoded image derived from the fractal code obtained by encoding the test image Lena (512 × 512 pixels, 8 bits/pixel) shown in Fig. 4.

In the encoding process, the size of the domain block, r_i , θ_i , and p_i were constant as shown in Table I, and the mean-squared error in D_i between the original image and transformed images from R_i was calculated for every combination of s_i , t_i , and q_i . The set of three parameters that provided the minimum error was selected as the code for D_i . The points $f_i(x, y)$ are not normally directly on pixels; therefore, the intensities at $f_i(x, y)$ were calculated by averaging the intensities of the four pixels around $f_i(x, y)$.



Fig. 3. Image decoded by fractal coding.



Fig. 4. Test image "Lena."

TABLE I
CODING PARAMETERS

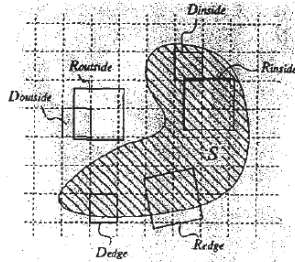
Size of domain block	= 8
r_i	= 2
θ_i	= 0
s_i, t_i	$\in \{-16, -14, -12, \dots, 14, 16\}$
p_i	= 0.9
q_i	$\in \{0.1, 2, \dots, 30\}$

A flat image with a uniform pixel value of 128 was set as the initial image in the decoding process, and F was iterated 30 times. The PSNR of the decoded image was calculated using the

$$\text{PSNR [dB]} = 20 \log \frac{255}{\sqrt{\text{MSE}}} \quad (9)$$

where MSE is the mean-squared error between the original image and the decoded image. A value of 26.2 dB was obtained.

Since the number of combinations of s_i , t_i , and q_i is $17 \times 17 \times 31 = 8959$, which is smaller than $2^{14} = 16384$, 14 bits is a sufficient code length for a domain block. Therefore,

Fig. 5. Example of domain blocks and their range blocks around an object S .

the bit rate is

$$\frac{\text{code length for a domain block}}{\text{number of pixels in a domain block}} = \frac{14}{8 \times 8} = 0.21875 \text{ bits/pixel.} \quad (10)$$

It may be possible to reduce the bit rate by using the mean-separation method [8], [10], [11], the block-type switching method [5], [8], the flexible-block-partitioning method [7], or the variable length coding method [3] in combination.

III. IMAGE REPRESENTATION USING DYNAMICAL SYSTEMS PARAMETERIZED BY FRACTAL CODING

A decoded image can be obtained by the dynamical system defined by (1) parameterized by r_i , θ_i , s_i , t_i , p_i and q_i , and a segmented image and a contour-detected image can be produced from another dynamical system parameterized by r_i , θ_i , s_i , and t_i . In this paper, p_i and q_i are not used in the image representations.

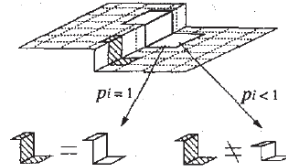
A. Segmentation Using an Enlarging Map W

We propose a discrete dynamical system in which the map

$$W(X) = \bigcup_{i=1}^I \{f_i(x) | x \in \{D_i \cap X\}\} \quad (11)$$

is iterated. Here, $x = (x, y)$ is a point in A , X is a set of x , and it is assumed that each r_i in f_i satisfies (5). Although each f_i enlarges the distance between any two arbitrary points in D_i , the mapped points always lie within A because all R_i are in A . Stretching by f_i and overlapping of R_i generates chaotic loci with the mapped point moving in a complicated manner.

An ideal example of domain blocks and their range blocks around an object $S \subset A$ is illustrated in Fig. 5. There are some texture patterns of intensity, but no distinct edges in S . The background is flat, and its intensity is different from the mean intensity in S . For $q_i = 0$, most of the blocks satisfy the conditions described below because the block which gives the minimum distance from the domain block is selected as its range block. Note that in the experiments, q_i was allowed to

Fig. 6. A range block can be found easily when $p_i = 1$.

be a value other than zero because the objects and background in the test image had a considerable amount of texture.

C1: If a domain block is in S , its range block is also in S . (See D_{inside} and R_{inside} .)

C2: If a domain block is outside S , its range block is also outside S . (See D_{outside} and R_{outside} .)

If the contrast at the edge of S is constant and $p_i = 1$, the following condition also holds.

C3: If a domain block includes a contour of S , its range block also includes the same contour, and the pixel pattern in the range block is similar to that in the domain block. (See D_{edge} and R_{edge} .)

A range block which satisfies C3 can be found easily for a domain block containing a simple edge, such as that shown in Fig. 6. The contrast of a transformed image with $p_i < 1$ is lower than that of the domain block; therefore, a part of another object is often selected as the range block when $p_i < 1$. The decoded images, however, become worse when p_i approaches one, and this is predicted by the Collage theorem. The images are not reconstructed with $p_i = 1$ because the transformation defined by (2) is not contractive in this case. We set $p_i = 0.9$ in our experiments.

Although some range blocks are omitted in Fig. 5 for simplification, it is assumed that there is a range block satisfying conditions C1–C3 for each domain block.

Setting isolated points [shown in Fig. 7(a)] as the initial set of the map W , all points in S are mapped inside S and all points outside S are mapped outside S . If a point is in a domain block such as D_{edge} , the distance between the contour and the mapped point is enlarged step by step in the ratio of the range block size to the domain block size until the point is mapped into a domain block which does not include the contour. Therefore, there is a set T such that

$$\forall n > 0, \quad W^n(T) \subseteq T \quad (12)$$

and

$$\exists N, \exists x \in \bar{T} \text{ such that } \forall n > N, \quad W^n(x) \in T \quad (13)$$

where \bar{T} is the complementary set of T . This is shown in Fig. 7(c). The set T which satisfies (12) and (13) is called the trapping region [12].

Another example of a square object S_{sq} is shown in Fig. 8. All mapped points starting from points in S_{sq} are mapped into the block D_O , which includes no contours, in a finite number of iterations of W even if the points started from domain blocks including contours. The points x_0 , x_1 , x_2 , and x_3 in

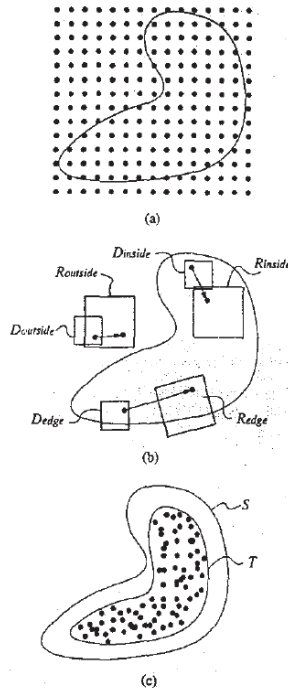


Fig. 7. (a) Initial set of the map W . (b) all points in S are mapped inside S , and (c) there is a trapping region T within the object S .

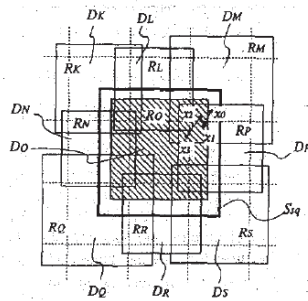


Fig. 8. All mapped points starting from points in S_{aq} are mapped into the block D_O in a finite number of iterations of W .

the figure are such points. For every D_* ($*$ = K, L, \dots, S)

$$W(R_O \cap D_*) \subseteq R_O. \quad (14)$$

Therefore, R_O is a sufficiently large trapping region, and there may be smaller trapping regions within R_O .

Real images are more complicated than the examples; therefore, we cannot show trapping regions for a real image system explicitly at this stage. However, the results of the experiment described in Section IV-A show that the limit set with W forms groups of mapped points.

Since points starting from different objects go to different trapping regions, a segmented image is made available by coloring the initial points according to the group with which the point gets clustered. In other words, the object is equal to a set B , where

$$B = \{x | \exists n \geq 0, W^n(x) \in T\}. \quad (15)$$

The set B which satisfies (15) is referred to as the basin of attraction [13].

The segmentation algorithm is as follows.

- Step 1: Place points x_0^j ($j = 1, \dots, J$, where J is the number of pixels) at the pixel positions.
- Step 2: Perform iteration of the map W η times, and get x_η^j ($j = 1, \dots, J$).
- Step 3: Classify x_η^j into γ clusters.
- Step 4: Put a color on x_0^j for each j , depending on the cluster of x_η^j .

B. Contour Detection Using a Contractive Map M

We also define a map M as

$$M(X) = \bigcup_{i \in E} \{g_i(x) | x \in \{R_i \cap X\}\} \quad (16)$$

where $g_i(x)$ is the inverse mapping of $f_i(x)$ such that

$$g_i \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos \theta_i & \sin \theta_i \\ -\sin \theta_i & \cos \theta_i \end{pmatrix} \begin{pmatrix} x - x_i \\ y - y_i \end{pmatrix} + \begin{pmatrix} -\frac{r_i \cos \theta_i + t_i \sin \theta_i}{r_i} + x_i \\ \frac{s_i \sin \theta_i - t_i \cos \theta_i}{r_i} + y_i \end{pmatrix} \quad (17)$$

and E is the set of indexes i such that D_i includes the contour of an object such as D_{edge} in Fig. 5.

Since g_i is contractive ($r_i > 1$), M^n converges to a limit set when $n \rightarrow \infty$. Fig. 9(a) shows the results of the mapping for a region covering the whole of S in the example of Fig. 5. The result is equal to the union of D_i for $i \in E$, and Fig. 9(b) shows replacement at D_{edge} of Fig. 9(a) by the twice-mapped image. The width of the black area narrows in the ratio of the domain block size to the range block size, and converges to the contour of S shown in Fig. 9(c) as $n \rightarrow \infty$.

We propose an algorithm with a reasonable computational load, in which the number of mapping points is fewer than or equal to the number of pixels in A .

- Step 1: Place points at the pixel positions and call the initial set X_0 .
- Step 2: Iterate the map M κ times and get X_κ . Here, round off the position of each of the mapped points to the position of the nearest pixel in each iteration.

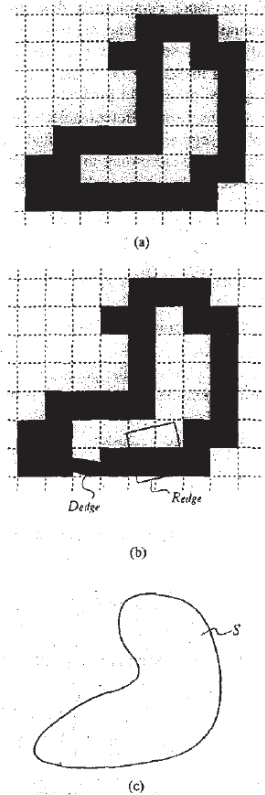


Fig. 9. (a) Result of a mapping, which is equal to the union of D_i for $i \in E$. (b) replacement at D_{edge} by the twice-mapped image. (c) The limit set is equal to the contour of S .

Step 3: Show the pixels of X_k obtained as a result of contour detection.

In general, the number of mapped points increases as a result of the mapping M because there are points which belong to more than one range block. Divergence of the number of mapped points is, however, avoided because some points are unified in Step 2.

The proposed method requires an accurate determination of E because, if E does not include a part of the contour, the limit set might be a broken line. The method for determining E is described in Section IV-B.

IV. EXPERIMENTS

A. Segmentation

The fractal codes used in the encoding and decoding experiments in Section II were also used as parameters of W in

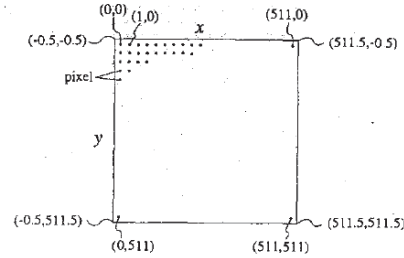


Fig. 10. Coordinates in the experiments. The pixels are placed at positions (x, y) such that x and y are integers.

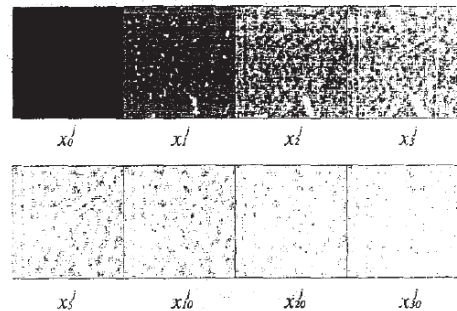


Fig. 11. Iteration of the map W results in the mapped points forming groups.

this experiment. Fig. 10 shows the coordinates on image plane A . The pixels were placed at positions (x, y) such that x and y were integers. When the mapped point was on the border between two blocks, the right side or lower side block was selected as D_i .

Since $r_i = 2$ (Table I) and x_0^j were at positions on pixels

$$x_n^j \in \{(k+0.5, l+0.5) | 0 \leq k, l \leq 510\}, \quad (n \geq 1). \quad (18)$$

The loci of mapped points are limited as described by (18), and are, thus, not chaotic. This state is a result of the pixels being on singular points of the dynamical system when $r_i = 2$ [14]. This is not a problem in this study. When required, the locus can be made chaotic by just adding a perturbation, e.g., $r_i = 1.99999$ because the singular points are unstable. The perturbation allows x_n^j to take various positions.

Fig. 11 shows $x_0^j, x_1^j, \dots, x_{30}^j$. After several iterations, the points form groups. The points move within each group but they never jump to other groups due to the mapping. It seems that there are trapping regions around every group, as is to be expected from Section III-A.

Fig. 12 is a segmented image with $\eta = 30$ and $\gamma = 64$. The value of 30 for η was found to be sufficient by trial and error, iteration greater than 30 times provided almost the same segmented images as Fig. 12. The K -means algorithm [15] was used to classify the mapping points in Step 3. Initially,



Fig. 12. Segmented image with $p_i = 0.5$.

64 cluster-mean points were placed at the same interval in A , and they were replaced eight times.

The hat, the face, and the mirror are segmented well. Although there are many feathers in the decoration of the hat, these regions are recognized as several segments. However, the chin has merged into the shoulder, and there is a substantial amount of grain noise in each segmented region. These errors seem to have been caused by the following factors.

- 1) One of the conditions C1–C3 is broken around the object.
- 2) There are multiple trapping regions in an object.
- 3) There is misclustering in Step 3.

In the case of 1), the mapped points might jump to another object, and this causes merging. Factor 2) causes oversegmentation and grain noise because basins seem to be overly complicated in objects including no edge. Factor 3) is a technical problem which results in multiple groups being judged as a single group and a single group being divided into multiple clusters.

Fig. 13 shows other results with $q_i \in \{0, 1, 2, \dots, 255\}$ and $p_i = 0.7$ and $p_i = 0.5$. The blocks satisfying the conditions are fewer than those in the case of Fig. 12; therefore, the quality is worse, and there is more merging and noise. For segmentation, a p_i value close to 1, e.g., $p_i = 0.9$, is desirable.

B. Contour Detection

In this experiment, a block-type switching method [5], [8] was used to encode the image for the convenient detection of E . Blocks D_i having a smaller variance than the given T_h (T_h was taken as 25 in the experiment) were regarded as blocks without an edge and labeled *type2*, and the other blocks were labeled *type1* [8]. Note that

$$\frac{\sum_{k=1}^N (b_k - a_i)^2}{N} < T_h \quad (19)$$

where N is the number of pixels in D_i , b_k is intensity of the k th pixel, and a_i is the intensity mean of D_i . Jacquin has used more block types and a different block-type detection method in [5].

If a block was labeled *type2*, the intensity mean of the block was only involved in the compressed code. The fractal transformation, therefore, changes to

$$\begin{aligned} (\hat{F}\mu)(x, y) &= \begin{cases} v_i(\mu(f_i(x, y))) & \text{when } \tau_i = \text{type 1 and } (x, y) \in D_i \\ a_i & \text{when } \tau_i = \text{type 2 and } (x, y) \in D_i \end{cases} \end{aligned} \quad (20)$$

Here, τ_i is the block type.

Although the block-type switching method requires bits for τ_i , the total amount of fractal code required is less than that required by the conventional method if there are a sufficient number of blocks of *type2*, because *type2* blocks do not need transformation parameters.

The range blocks for the *type1* blocks were found with the parameters shown in Table I.

X_0 , X_1 , X_2 and X_3 are shown in Fig. 14. Here

$$E = \{\text{block index of type1}\} \quad (21)$$

in (16). The black colored area converges to a limit set, and Fig. 15 shows the result with $\kappa = 4$, that is, for X_4 . X_n with $n > 4$ were almost the same as X_4 . Contours of the hat, the face, the shoulder, and the frame of the mirror were detected. However, the eyes and the decoration of the hat could not be seen. This was because the block size was too large to satisfy the conditions C1–C3 around such small objects. In addition, black regions that are not contours and are a result of inaccuracy in the determination of E can be seen.

There are small pieces of fractal in the image, such as Sierpinski triangles (Fig. 16) in the hat, its decoration, and the hair. The Sierpinski triangles are thought to be the result of an error in the determination of the contour of an object because none of the objects has such a complicated contour. However, they show that the area has a high-contrast texture.

C. Computational Complexity

The CPU time for contour detection was only 1/24 of the decoding time. It is, therefore, obvious that, for the case where the image has been transformed to the fractal code, the detection of contours is faster by the proposed method than by any method that requires decoding of the image. Our method is useful for producing contour images from compressed codes because it reduces the computational complexity. When uncoded images are provided, however, a considerable amount of time is required to encode them.

The CPU time required for segmentation was 442 times that required for contour detection. 90% of the segmentation process was used for the K -means algorithm in Step 3.

V. CONCLUSIONS

A segmentation method and a contour detection method based on dynamical systems parameterized by fractal coding

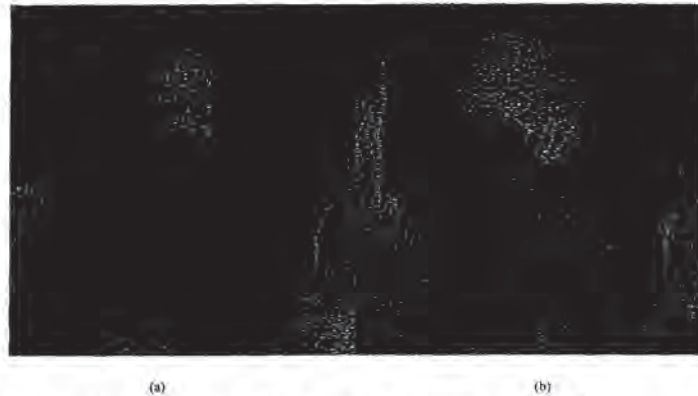


Fig. 13. Segmented image with (a) $p_i = 0.7$ and (b) $p_i = 0.5$. Segmentation quality is worse than with $p_i = 0.9$.

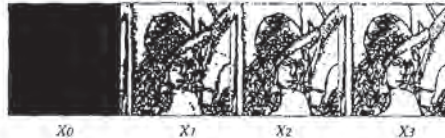


Fig. 14. Converging to a set by iteration of the map Af .



Fig. 15. Contour-detected image.

were proposed. It was shown that an enlarging mapping system produces a segmented image, and that the limit set in the inverse system forms the contours.

Although the quality of segmented and contour-detected images can be improved by adding rotations to f_i or using a smaller block size to satisfy conditions C1–C3, these parameters require more bits in the compressed code. The parameters should be set depending on the use.

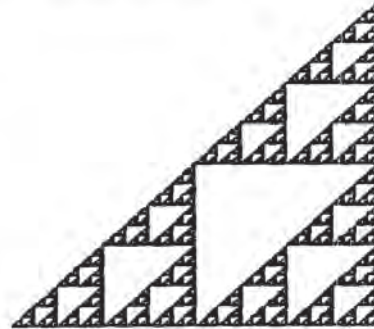


Fig. 16. Sierpinski triangle.

The proposed methods are new applications of fractal coding, and can be considered to be totally new approaches to image representation. A benefit of the proposed methods is that they are done on the level of compressed code. If images compressed by fractal coding are kept on a database, an image can be contour detected in a short time without decompressing it to the gray-scale image. Therefore, the methods are useful for fast browsing to find images, for example. In addition, they will contribute to image processing on compressed codes, which will be critical for the handling of large numbers of images on a machine. This is because image segmentation and contour detection are basic technologies for image editing, recognition, and other types of image processing.

ACKNOWLEDGMENT

The authors would like to thank R. Tokunaga, Assistant Professor, University of Tsukuba, for his valuable suggestions. The authors are also grateful to the reviewers for their many useful comments.

REFERENCES

- [1] M. Barnsley, *Fractals Everywhere*. San Diego, CA: Academic, 1988.
- [2] M. Barnsley and A. D. Sloan, "Method and apparatus for processing digital data," U.S. Patent 5065447, 1991.
- [3] M. Barnsley and L. P. Hurd, *Fractal Image Compression*. Wellesley, MA: AK Peters, 1993.
- [4] A. Jacquin, "A novel fractal block-coding technique for digital images," in *Proc. ICASSP-90*, 1990, pp. 2225-2228.
- [5] ———, "Image coding based on a fractal theory of iterated contractive image transformations," *IEEE Trans. Image Processing*, vol. 1, pp. 18-30, Jan. 1992.
- [6] ———, "Fractal image coding: A review," *Proc. IEEE*, vol. 81, pp. 1451-1465, Oct. 1993.
- [7] Y. Fisher, Ed., *Fractal Image Compression: Theory and Application*. New York: Springer-Verlag, 1994.
- [8] T. Ida and K. Dachiku, "Image compression using iterated transformation theory-based coding" (in Japanese), in *Proc. 5th Karuizawa Workshop Circuit Syst.*, 1992, pp. 137-142.
- [9] T. Ida and Y. Sambonsugi, "Image segmentation using fractal coding," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 5, pp. 567-570, Dec. 1995.
- [10] T. Ida, "Improvement of fractal image coding using new code-decision-criterion," in *Proc. Int. Picture Coding Symp.*, 1994, pp. 444-447.
- [11] S. Lepsoy, G. Gien, and T. Ramstad, "Attractor image compression with a fast noniterative decoding algorithm," in *Proc. ICASSP-93*, 1993, pp. V-337-V-340.
- [12] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. New York: Springer-Verlag, 1983.
- [13] T. Matsumoto, M. Komuro, H. Kokubu, and R. Tokunaga, *Bifurcations*. Tokyo: Springer-Verlag, 1993, p. 309.
- [14] J. Freyland, *Chaos*. London: Institute of Physics, 1992, pp. 31-33.
- [15] A. Tekalp, *Digital Video Processing*. Englewood Cliffs, NJ: Prentice Hall, 1995, pp. 510-512.



Takashi Ida (M'95) received the B.Eng. and M.Eng. degrees in electrical engineering from Waseda University, Tokyo, Japan, in 1987 and 1989, respectively.

Since 1989, he has been with the Research and Development Center, Toshiba Corporation, Kawasaki, Japan. His research interests include image compression and image processing based on fractal theories.

Mr. Ida received the 1994 Young Engineer Award from the Institute of Electronics, Information, and Communication Engineers (IEICE), Japan. He also received awards for young engineers from the Institute of Electrical Engineers of Japan (IEEE) in 1994, and from the Japan Society for Fuzzy Theory and Systems (SOFIT) in 1996, respectively. He is a member of IEICE.



Yoko Sambonsugi received the B.Eng. and M.Eng. degrees in electrical and computer engineering from Yokohama National University, Yokohama, Japan, in 1990 and 1992, respectively.

Since 1992, she has been with the Research and Development Center, Toshiba Corporation, Kawasaki, Japan. Her current research interests include still/motion image processing, coding, and segmentation.

Ms. Sambonsugi is a member of IEICE.

Bidirectional Motion Estimation via Vector Propagation

Showbhik Kalra, *Member, IEEE*, and Man-Nang Chong, *Member, IEEE*

Abstract—A low-complexity vector propagation (VP) algorithm is introduced for the estimation of bidirectional motion vector fields in image sequences. The proposed VP algorithm exploits the strong correlation between forward and backward motion vector fields in image sequences. The performance of the VP algorithm is compared to that of a bidirectional multiresolution block-matching (MRBM) motion estimation (ME) algorithm. Computer simulation results demonstrate that with the VP algorithm, the computational workload of the bidirectional ME is reduced by a factor of nearly 2. The robustness of the VP algorithm is extensively tested using computer generated image sequences and real movies for a motion picture restoration (MPR) system. It is shown that the VP algorithm is robust enough to be used in the computationally demanding MPR algorithm, in which the performance of the novel VP algorithm is close to that of a bidirectional MRBM ME algorithm. With this technique, other video processing systems that desire to take advantage of bidirectional motion estimation can now do so without an excessive increase in computing time.

Index Terms—Bidirectional, image restoration, motion estimation, vector propagation.

1. INTRODUCTION

BIDIRECTIONAL motion estimation solves some fundamental problems in the motion vector field estimation process, that of occlusion and scene changes (motion vector discontinuity). Occlusion refers to the covering/uncovering of a surface due to three-dimensional (3-D) rotation and translation of an object that occupies only part of the field of view [1]. It occurs quite often in real-world images. For example, objects move in front of other objects, objects move toward the camera, cameras zoom, and objects rotate. If only two frames are used at such regions, the motion estimation (ME) algorithm will not be able to find a good estimate of the underlying motion at these regions because there is no corresponding feature in the other frame to match. In addition, this bad motion estimate could affect other surrounding motion vectors due to the spatial smoothness constraints applied in various ME algorithms [15], [20], [21]. Poor motion estimates, in turn, provide bad support to video processing applications such as video coding [16] and motion picture restoration (MPR) [4], and result in visible artifacts in such areas. The simple and well-known solution [15], [17] to this problem is

to use three frames, including one preceding frame and one succeeding frame; this solution is commonly referred to as bidirectional ME. Bidirectional ME approaches also provide a simple and elegant method to handle scene changes in video [9], [18].

As shown in Fig. 1, for each pixel block to be bidirectionally predicted, a forward motion vector (\vec{v}_f) and a backward motion vector (\vec{v}_b) pair is estimated by searching for a matching block from the past and future reference frames [17]. Therefore, bidirectional ME presents a formidable computational challenge. Bidirectional ME forms a major computation bottleneck in video processing applications such as the detection of noise in image sequences [2], [6], interpolation/prediction of missing data in image sequences [3], [4], [18], and deinterlacing of image sequences [9]. The process of finding robust motion vectors using minimal computations is a heavily researched area, and various fast algorithms have been proposed [7], [12]–[14]. Although the multiresolution block-matching (MRBM) algorithm [7] has been regarded as one of the most efficient methods for block-based ME [1], ME still takes up a major part of the processing time in video processing applications. In order to reduce the overall processing time in some video processing applications, the complexity of the bidirectional MRBM ME algorithm is looked into. In this paper, a low-complexity bidirectional ME algorithm using vector propagation (VP) is proposed to alleviate the computational burden of a bidirectional ME algorithm.

As shown in Fig. 1, there exists a strong correlation between \vec{v}_{f1} and \vec{v}_{b2} , and likewise for \vec{v}_{b3} and \vec{v}_{f2} . Motivated by this observation, a novel VP algorithm is proposed. The VP algorithm reduces the computational cost by postprocessing the forward motion vector fields ($\vec{v}_{f1}, \vec{v}_{f2}, \dots, \vec{v}_{f_{n-1}}$) so that the postprocessed forward motion vector fields can be used as the backward motion vector fields, ($\vec{v}_{b2}, \vec{v}_{b3}, \dots, \vec{v}_{b_n}$). The cost of postprocessing the forward motion vector fields is only a small fraction of that of an MRBM ME algorithm.

The problem of vector propagation (VP) is not simply one of reversing the vector field. Due to the presence of noise, occlusions, or nontranslational displacements, multiple motion estimates may match to the same location between frames. Therefore, decisions have to be made as to which motion estimate is the most suitable motion vector in this scenario. Other problematic regions (such as occluded regions) will not be able to find motion estimates from the past motion field. The VP algorithm utilizes the fact that motion vector fields exhibit strong correlation in time and space. Hence, to

Manuscript received August 5, 1997; revised April 29, 1998 and July 20, 1998. This paper was recommended by Associate Editor T. Sikora.

S. Kalra is with the Center for Signal Processing, Nanyang Technological University, 639798 Singapore.

M.-N. Chong is with the School of Applied Science, Nanyang Technological University, 639798 Singapore.

Publisher Item Identifier S 1051-8215(98)09283-0.

1051-8215/98\$10.00 © 1998 IEEE

A Segmentation-Based Lossless Image Coding Method for High-Resolution Medical Image Compression

Liang Shen, Member, IEEE, and Rangaraj M. Rangayyan,* Senior Member, IEEE

Abstract—Lossless compression techniques are essential in archival and communication of medical images. In this paper, a new segmentation-based lossless image coding (SLIC) method is proposed, which is based on a simple but efficient region growing procedure. The embedded region growing procedure produces an adaptive scanning pattern for the image with the help of a very few bits-needed discontinuity index map. Along with this scanning pattern, an error image data part with a very small dynamic range is generated. Both the error image data and the discontinuity index map data parts are then encoded by the Joint Bi-level Image experts Group (JBIG) method. The SLIC method resulted in, on the average, lossless compression to about 1.6 b/pixel from 8 b, and to about 2.9 b/pixel from 10 b with a database of ten high-resolution digitized chest and breast images. In comparison with direct coding by JBIG, Joint Photographic Experts Group (JPEG), hierarchical interpolation (HINT), and two-dimensional Burg prediction plus Huffman error coding methods, the SLIC method performed better by 4% to 28% on the database used.

Index Terms—Lossless compression, medical image coding, region-based image processing, region growing, segmentation.

I. INTRODUCTION

THE trend in medical imaging is increasingly toward direct digital image data acquisition. Currently, many modalities such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), single-photon emission computed tomography (SPECT), and digital subtraction angiography (DSA) produce images directly in digital form. Digital X-ray imaging systems are also being introduced for chest and breast imaging, and various types of film digitizers are being used to convert traditional X-ray films into digital format. The main advantages of digital imaging lie in its ready support of image transfer and archival, and the possibility of manipulating and enhancing diagnostic information. However, digital image transmission and storage are facing major challenges due to the growing size of medical image datasets as well as practical limitations in transmission

bandwidth and storage space. This has led to a search for effective image compression methods.

During the past two decades, various lossless and lossy image coding techniques have been developed. Although lossy coding methods can achieve high compression ratios, e.g., 50:1 or more, they lack the ability to exactly recover the original data. While a lossless coding method will allow exact recovery from the compressed version of the data, the compression ratio achievable is usually only of the order of 2:1 or 3:1. In medical applications, lossless coding methods are required since loss of any information is usually unacceptable.

Most of the recent advances in image coding have taken place in the area of lossy compression, such as vector quantization [1], wavelet and subband coding [2], [3], fractal-based coding [4], neural network-based coding [5], and model-based coding [6]. Lossless coding algorithms have not seen much progress. Theoretically, any lossy coding approach may be modified to be a lossless coding method by appending the coding error; however, such a simple procedure results in low compression ratios because of the large amounts of code needed for encoding the error. Novel techniques are needed to obtain high compression ratios in lossless coding. In recent years, a few new methods have been proposed for lossless compression, such as hierarchical interpolation (HINT)—a multiresolution technique [7], two-dimensional (2-D) linear predictive coding [8], context-based coding [9], 2-D multiplicative autoregressive model-based coding [10], and higher-order entropy coding [11]. Although a detailed comparison of their compression efficiencies is not yet available, most of the methods mentioned above (except HINT) have high complexity of implementation.

In 1982, Kocher and Kunt [12] proposed a contour-texture approach to picture coding; they called such approaches as second-generation image coding techniques. The main idea behind this technique is to first segment the image into nearly homogeneous regions surrounded by contours such that the contours correspond, as much as possible, to those of the objects in the image, and then to encode contour and texture information separately. Because contours can be represented as one-dimensional signals and pixels within a region are highly correlated, such methods are expected to attain high compression ratios. Although the idea seems to be promising, its implementation meets a series of difficulties. A major problem exists at its very important first step—segmentation—which determines the final performance

Manuscript received May 3, 1995; revised September 20, 1996. This work was supported by the Alberta Heritage Foundation for Medical Research, the Alberta Breast Cancer Foundation, and the Natural Sciences and Engineering Research Council of Canada. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. Viergever. Asterisk indicates corresponding author.

L. Shen is with Array Systems Computing Inc., North York, Ont., Canada. *R. M. Rangayyan is with the Department of Electrical and Computer Engineering, The University of Calgary, 2500 University Drive N.W., Calgary, Alta T2N 1N4 Canada. (e-mail: rangaraj@enei.ucalgary.ca).
Publisher Item Identifier S 0278-0062(97)03951-7.

0278-0062/97\$10.00 © 1997 IEEE

of the segmentation-based coding method. It is well recognized that there are no satisfactory segmentation algorithms for application to a wide variety of general images. Most of the available segmentation algorithms are sophisticated and give good performance only for specific types of images.

We propose a segmentation-based lossless image coding (SLIC) technique for radiographic images, especially for digitized or digital mammography and chest radiography. To overcome the problem mentioned above in relation to segmentation, we propose a simple region growing method. Instead of generating a contour set, a discontinuity map is obtained during the region growing procedure. Concurrently, the method also produces a corresponding error image based on the difference between each pixel and its corresponding "center pixel." In our study, it was observed that the international bilevel image compression standard Joint Bi-level Image experts Group (JBIG) [13], [14] is very efficient for coding the discontinuity map and the error image. The following sections provide details of the SLIC algorithm and results of coding digitized X-ray images, which are also compared with those obtained by directly using two international standards for lossless still-image compression, namely JBIG and Joint Photographic Experts Group (JPEG) [15], as well as adaptive Lempel-Ziv (ALZ) [16], 2-D Burg prediction plus Huffman error coding [8], and the HINT methods.

II. SEGMENTATION-BASED IMAGE COMPRESSION

A lossless compression method basically includes two major stages: one is image transformation, with the purpose of data decorrelation; the other is entropy encoding of the transformed data. However, in the SLIC algorithm, image transformation is achieved in both the region growing procedure and, later, in the JBIG coding procedure, while entropy encoding is accomplished within the JBIG algorithm. The JBIG algorithm, defined in the International Telegraph and Telephone Consultative Committee (CCITT) recommendation T.82, uses an adaptive three-dimensional (3-D) coding model followed by an adaptive arithmetic coder [13], [14].

A. Region Growing Algorithm

It should be noted that the aim of segmentation in image compression is not object identification or analysis of features; its aim is instead to group spatially connected pixels lying within a small dynamic gray level range. Our region growing procedure starts with a single pixel, called the seed pixel (#0 in Fig. 1). Each of the seed's four-connected (neighbor) pixels (from #1 to #4 in the exact order as shown in Fig. 1) is checked with a region growing (or inclusion) condition. If the condition is satisfied, the neighbor pixel is included in the region. The four neighbors of the newly added neighbor pixel are then checked for inclusion in the region. This recursive procedure is continued until no spatially connected pixel meets the growing condition. A new region growing procedure is then started with the next pixel of the image which is not already a member of a region; the procedure ends when every pixel in the image has been included in one of the regions grown.

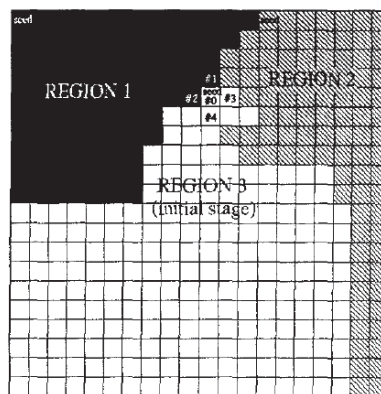


Fig. 1. Demonstration of a seed pixel (#0) and its four-connected neighbors (#1–#4) in region growing.

The region growing conditions used in this work are that: the neighbor pixel not be a member of any of the regions already grown, and that the absolute difference between the neighbor pixel and the corresponding center pixel be less than error level (to be defined later). Fig. 2 demonstrates the relationship between a neighbor pixel and its center pixel: at the specific stage illustrated in the figure, pixel A has become a member of the region (3) being grown and its four neighbors, namely B, C, D, and E, are being checked for inclusion (E is already a member of the region). Under this circumstance, pixel A is the center pixel of neighbor pixels B, C, D, and E. When a new (neighbor) pixel is included in the region being grown, its error-level-shift-up difference with respect to its center pixel is stored as the pixel's "error" value. If only the first of the two region growing conditions is met, the discontinuity index of the pixel is incremented. Therefore, after region growing, a "discontinuity index image data part" and an "error image data part" will be obtained. It is clear that the maximum value of the discontinuity index is four. Most of the previously reported segmentation-based coding algorithms include contour coding and region coding; instead of these steps, we use a discontinuity index map and an error image data part.

The error level is determined by the preselected error bits as

$$\text{error level} = 2^{\text{error bits}-1}. \quad (1)$$

For instance, if the error image is allowed to take up to 5 b/pixel (*error bits* = 5), the corresponding *error level* is 16; the allowed difference range is then $[-15, 15]$. The error value of the seed pixel of each region is defined as the value of its low *error bits* bits; the value of the high ($N - \text{error bits}$) bits of the pixel is stored in a "high-bits seed data part," where N is the number of b/pixel in the original image data.

The above three data parts are used to fully recover the original image during the decoding process. The region growing conditions during decoding are that the neighbor pixel under

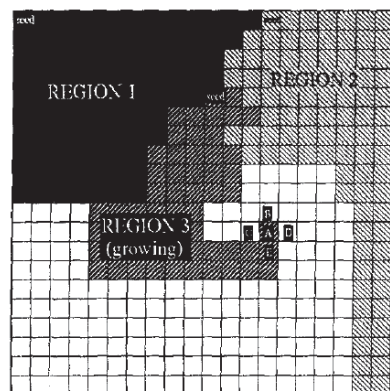


Fig. 2. Demonstration of a pixel being checked for inclusion (B, C, D, or A) and its center pixel (A) during region growing.

consideration for inclusion be not in any of the previously grown regions, and that its discontinuity index equal zero. When the conditions are met for a pixel, its pixel value is restored as the sum of its *error-level-shift-down* error value and its center pixel value (except for the seed pixels of every region). If only the first of the two conditions is satisfied, the discontinuity index of that pixel is decremented. Thus the discontinuity index generated during segmentation is used to guide region growing during decoding. The "high-bits seed data part" is combined with the "error image data part" to recover the seed pixel value of each region.

Fig. 3 is a simple example for illustration of the region growing procedure and its result. The original image is a section of an eye of the 512×512 Lenna image with a size of eight rows by eight columns as shown in Fig. 3(a). The value of *error bits* is set to five in this example. Fig. 3(b) is the result of region growing. The corresponding three data parts, namely discontinuity index image data, error image data, and high-bits seed data, are shown in Fig. 3(c), (d), and (e). The full 512×512 Lenna image and the corresponding discontinuity index image data (scaled) and error image data (scaled) are shown in Fig. 4.

B. The Segmentation-Based Lossless Image Compression Procedure

The complete SLIC procedure is illustrated in Fig. 5. At the encoding end, the original image is transformed into three parts: discontinuity index image data, error image data, and high-bits seed data, by the region growing procedure. The first two data parts are then Gray-coded, broken down into bit planes, and, finally, JBIG-coded. The last data part is stored or transmitted as-is; it needs only $(N - \text{error bits})$ b/region. At the decoding end, the JBIG-coded data files are JBIG-decoded first and then the Gray-coded bit planes are composed back to binary code. Finally, the three parts are combined together by the same region growing procedure to recover the original image.

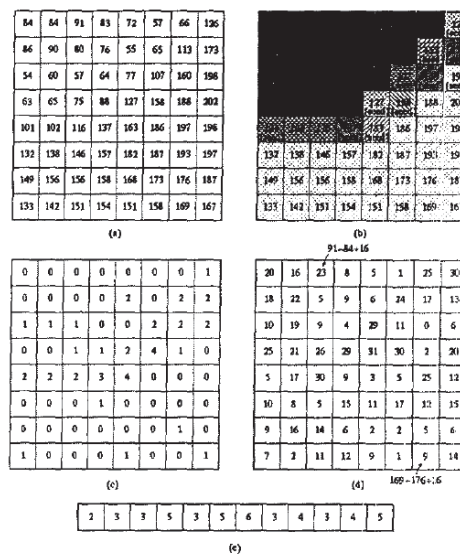


Fig. 3. A simple example of the region growing procedure and its result with *error bits* set to be five. (a) Original image with a size of eight rows by eight columns (a section of the 512×512 Lenna image); (b) result of region growing; (c) corresponding discontinuity index image data part; (d) corresponding error image data part; (e) The corresponding high-bits seed data part (from left to right).

III. EXPERIMENTAL RESULTS

A. Image Data

The X-ray film images used in this study were digitized using an Eikonix-1412 camera capable of digitization with a resolution of up to 4096×4096 , 12-b pixels. To maintain minute structures (such as microcalcifications present in mammograms) visible and detectable after digitization of X-ray films, the pixel size should not exceed $100 \mu\text{m}$ which leads to huge datasets of the order of 10–16 million pixels/image. Efficient archival and transmission of such images calls for lossless coding techniques with relatively high compression ratios.

The five mammograms and five chest radiographs used by Kuduvali and Rangayyan [8] were employed in this study. We first tested our algorithm using 8-b versions of the images which were mapped directly from the original 12-b images. The reason for this step is that most medical images as well as natural images used by other workers in the field have an 8-b dynamic range. Therefore, this step of our investigation would demonstrate the performance of our method with 8-b images in the context of other methods in the field. The SLIC method was also tested with commonly used 8-b nonmedical test images. Finally, we tested the method with 10-b versions of the medical images, which were obtained from the original 12-b versions by discarding the two least-significant bits; this



Fig. 4. Illustration of the 512x512 Lenna image and its segmentation results (*error bits* = 5). (a) Original Lenna image; (b) corresponding discontinuity index image data part (scaled); (c) corresponding error image data part (scaled).

step was adopted as tests with gray scale test images indicated that the 12-b Eikonix camera data were contaminated by noise to the extent of the two least-significant bits [8].

B. Comparison of Results

Two international, lossless, still-image compression standards, namely JBIG [13], [14], and JPEG [15], as well as

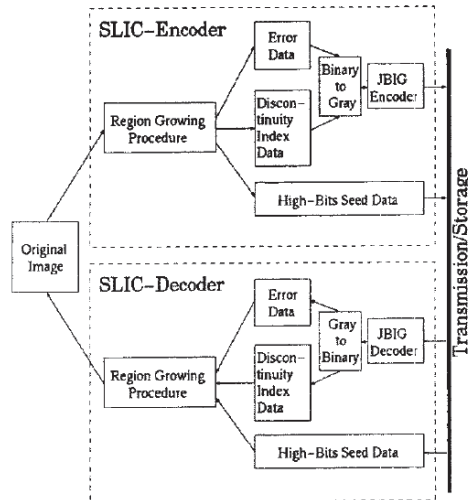


Fig. 5. Illustration of the SLIC procedure.

ALZ coding [16], HINT [7], and 2-D linear predictive coding [8] were used to encode the images for comparison with the SLIC method. In using JBIG (for direct coding of the image or for coding the error image data part and discontinuity index data part), we parameterized the method to use three lines of the image ($NLPS0 = 3$) in the 3-D model and no progressive spatial resolution buildup. The lossless JPEG package we used has features of automatic determination of the best prediction pattern and optimal Huffman table generation. The UNIX utility *compress* was used for ALZ compression. For the 10-b test images, 2-D Burg algorithm-based linear prediction followed by Huffman error coding (2-D-Burg-Huffman or 2DBH), which was previously reported give the best performance among other combinations [8], was utilized to take the place of the JPEG package for 8-b images.

1) *Results for 8-b Images*: The SLIC method has a tunable parameter, which is *error bits*. During this investigation, we found the compression ratio to be almost the same for most of the 8-b medical images by using four or five for *error bits*. Therefore, we fixed *error bits* = 5 for the following experiments with the 8-b versions of the medical images.

The performance of the SLIC method is summarized in Table I along with the results of JBIG, JPEG, ALZ, and HINT compression (Note: *ebits* = *error bits*). It is seen that the SLIC method has outperformed all of the other methods studied with the test image set used, except in the case of one mammogram and one chest radiograph for which JBIG gave negligibly better results. On the average, SLIC improved the bit rate by about 9%, 29%, and 13%, respectively, compared with JBIG, JPEG, and HINT.

While the SLIC procedure performed well with high-resolution medical images, we have observed that its

TABLE I
COMPARISON OF SLIC WITH ALZ, JBIG, JPEG, AND HINT USING FIVE 8-b MAMMOGRAMS
AND FIVE 8-b CHEST RADIOGRAPHS BY b/PIXEL (WITH LOWEST BIT RATES HIGHLIGHTED)

IMAGE (8 Bits)	ENTROPY (1st Order)	ALZ (<i>'compress'</i>)	JBIG	JPEG	HINT	SLIC (<i>ebits</i> = 5)
mammo1 (4096x1990)	6.1316	2.9544	1.8218	2.1389	1.8678	1.7287
mammo2 (4096x1800)	6.0950	2.8570	1.8357	2.1779	1.9169	1.7631
mammo3 (3596x1632)	5.9221	3.0194	2.4025	2.3720	2.1243	2.1173
mammo4 (3580x1696)	5.8955	2.4479	1.9628	1.8907	1.6150	1.4363
chest5 (3536x3184)	7.0495	3.0270	1.6028	1.9238	1.7458	1.4240
chest6 (3904x3648)	7.4611	3.3225	1.8795	2.1478	1.9956	1.7566
chest7 (3120x3632)	5.2970	1.7287	0.6452	1.5997	1.3957	0.9946
chest8 (3744x3328)	7.4535	3.1277	1.4957	1.8908	1.6412	1.5540
mammo9 (4096x2304)	6.0439	2.5998	1.6738	2.1197	1.8415	1.6749
chest10 (3664x3680)	7.1734	2.9457	1.6301	1.9683	1.7280	1.4088
AVERAGE	6.4523	2.8060	1.7250	2.0230	1.7873	1.5709

TABLE II
COMPARISON OF SLIC WITH ALZ, JBIG, AND JPEG USING EIGHT COMMONLY USED 8-b IMAGES BY b/PIXEL

IMAGE (8 Bits)	ENTROPY (1st Order)	ALZ (<i>'compress'</i>)	JBIG	JPEG	SLIC (<i>ebits</i> = 8)	
					(NLPS0 = 3)	(NLPS0 = row)
Airplane (512x512)	6.4863	5.7098	4.2357	4.2040	4.2176	4.1117
Baboon (512x512)	7.1391	7.8441	6.3675	6.1758	6.5498	6.4440
Cameraman (256x256)	7.0097	6.6774	4.9235	4.9522	5.1381	4.9052
Lenna (256x256)	7.5683	7.9074	5.3662	5.0686	5.2595	5.0426
Lenna (512x512)	7.4451	7.0519	4.7996	4.6948	4.7351	4.6299
Peppers (512x512)	7.3744	6.8557	4.8168	4.7616	4.8975	4.7951
Sailboat (512x512)	7.2659	7.0710	5.3360	5.2629	5.4638	5.3594
Tiffany (512x512)	6.3825	5.8805	4.3797	4.3497	4.4341	4.3177
AVERAGE	7.1943	6.8747	5.0281	4.9337	5.0869	4.9507

TABLE III
COMPARISON OF SLIC WITH ALZ, JBIG, 2DBH, AND HINT USING FIVE 10-b MAMMOGRAMS
AND FIVE 10-b CHEST RADIOGRAPHS BY b/PIXEL (WITH LOWEST BIT RATES HIGHLIGHTED)

IMAGE (10 Bits)	ENTROPY (1st Order)	ALZ (<i>'compress'</i>)	JBIG	2DBH	HINT	SLIC (<i>ebits</i> = 6)
mammo1 (4096x1990)	7.2679	5.4499	2.8930	2.9249	2.7482	2.7327
mammo2 (4096x1800)	7.6142	6.0206	3.1879	3.1911	3.1499	3.0764
mammo3 (3596x1632)	6.6763	4.7326	3.1273	2.9662	2.8074	2.8097
mammo4 (3580x1696)	7.2137	5.1437	3.1989	2.7552	2.5844	2.5668
chest5 (3536x3184)	8.9181	7.2134	3.3296	3.3097	3.2825	2.9943
chest6 (3904x3648)	9.4331	7.8388	3.8735	3.8094	3.8862	3.5867
chest7 (3120x3632)	7.0726	4.6658	2.2984	2.5799	2.3439	2.2676
chest8 (3744x3328)	9.2928	7.3961	3.2453	3.1589	3.0232	2.8008
mammo9 (4096x2304)	7.8118	5.9329	3.1846	3.4139	3.2894	3.1680
chest10 (3664x3680)	9.0498	7.5119	3.3472	3.2693	3.1831	3.0680
AVERAGE	8.0350	6.1906	3.1686	3.1379	3.0298	2.9161

performance with low-resolution general images is comparable to the performance of JBIG and JPEG, as shown in Table II. When SLIC uses an optimized JBIG algorithm using the maximum number of lines, i.e., NLPS0 = row, the number of rows of the image (the last column of Table II) in the 3-D model instead of only three lines (NLPS0 = 3 in Table II), its performance is better than that of JBIG and almost as good as that of JPEG.

2) *Results for 10-b Images*: Based on our study, setting *error bits* as six appeared to be a good choice for the 10-b versions of the medical images. Table III lists the results of compression with SLIC, JBIG, ALZ, HINT, and 2DBH coding. The SLIC technique has provided lower bit rates than the other methods. The average bit rate is 2.92 for SLIC, while

JBIG, HINT, and 2DBH have average bit rates of 3.17, 3.03, and 3.14, respectively.

IV. DISCUSSION AND CONCLUSION

Recent issues of journals in the field have published three review papers on medical image compression [17]–[19] and one review on lossless image compression [20]; none of these papers has provided new directions for lossless compression. We have picked up the concept of segmentation-based coding and successfully applied it in conjunction with other existing compression methods to digitized high-resolution X-ray images.

TABLE IV
COMPUTING TIME OF SLIC FOR FIVE MAMMOGRAMS AND FIVE CHEST RADIOGRAPHS ON A SUN-SPARC-STATION2 IN s AND μ s/PIXEL

IMAGE	8-bit Version		10-bit Version	
	Encode (μ s/pixel)	Decode (μ s/pixel)	Encode (μ s/pixel)	Decode (μ s/pixel)
mammo1 (4096x1990)	208.0 (25.52)	209.1 (25.65)	263.8 (32.36)	248.4 (30.48)
mammo2 (4096x1800)	203.9 (27.66)	206.2 (27.97)	264.9 (35.93)	249.8 (33.88)
mammo3 (3596x1632)	171.4 (29.21)	167.6 (28.56)	197.0 (33.57)	186.6 (31.80)
mammo4 (3580x1696)	154.2 (25.40)	157.4 (25.92)	202.0 (33.27)	189.8 (31.26)
chest5 (3536x3184)	273.9 (24.33)	265.9 (23.62)	395.9 (35.16)	379.8 (33.73)
chest6 (3904x3648)	378.6 (26.58)	387.5 (27.21)	491.8 (34.53)	500.7 (35.16)
chest7 (3120x3632)	258.3 (22.79)	272.5 (24.05)	349.4 (30.83)	341.7 (30.15)
chest8 (3744x3328)	306.7 (24.62)	316.7 (25.42)	425.3 (34.13)	406.1 (32.59)
mammo9 (4096x2304)	252.0 (26.70)	270.3 (28.64)	324.6 (34.40)	329.9 (34.96)
chest10 (3664x3680)	350.4 (25.99)	355.3 (26.36)	449.6 (33.34)	450.2 (33.39)
AVERAGE	255.7 (25.88)	260.9 (26.34)	336.4 (33.75)	328.3 (32.74)

Compared with the ALZ method, on which the UNIX utility *compress* is based (*compress* is a very common tool used for routine, lossless compression of image and other data), SLIC needs only about 56% or 47% of the disk space for storing the ten 8-b or the ten 10-b medical images in the present study. One of the international lossless compression standards, JBIG, takes about 109% of the disk space required by SLIC for both of the 8-b and 10-b image sets. For the 8-b medical images, JPEG, another international lossless compression standard, results in 28% more disk space when compared with the SLIC algorithm. Using the 2DBH method instead of SLIC to compress the ten 10-b radiographic images would require 8% more disk space. SLIC has also clearly outperformed HINT: about 2.8 MB or 1.7 MB more disk space will be needed with HINT as the tool for data compression instead of SLIC for the 8-b or 10-b medical image database, respectively.

The success of SLIC is primarily due to its embedded region growing/segmentation procedure. Unlike raster-scanning and Peano-scanning [21] with fixed scanning patterns, SLIC generates an adaptive scanning pattern based on the characteristics of the given image. Since only a simple discontinuity index map is used to guide the scanning pattern, this small extra-byte expense will be of benefit to overall compression performance due to exploitation of the high correlation among neighboring pixels within a region grown (in fact, for coding the discontinuity index data parts, only 0.0030 b/pixel to 0.0103 b/pixel with an average of 0.0059 b/pixel were required for the 10-b medical images). For the X-ray images used in the present study, the number of regions varied from 1 to 9674 for the 8-b images with *error bits* = 5 and from 1 to 11889 for the 10-b images with *error bits* = 6.

It should be also noted that the speed of the SLIC method is very much comparable with the speeds of JBIG and JPEG due to the use of a very efficient region growing method. On a SUN-SPARC-station2, SLIC required about 150–390 s to encode/decode each of the high-resolution 8-b images in the study, and about 180–500 s for each of the 10-b versions (see Table IV). Since SLIC needs about 26 μ s/pixel or 33 μ s/pixel for encoding/decoding 8-b or 10-b images, it only takes about 1.7 s or 2.2 s for coding an 8-b or 10-b 256×256 image.

Most of the previously reported segmentation-based coding techniques [12], [22], [23] involve three procedures: segmentation, contour coding, and texture coding. For segmentation,

a sophisticated procedure is generally employed for extraction of closed contours. In our SLIC method, a simple single-scan neighbor pixel checking algorithm is used. The major problem after image segmentation with other methods lies in the variance of pixel intensity within the regions, which has resulted in the application of such methods to lossy coding instead of lossless coding. To overcome this problem, we use the most-correlated neighbor pixels to generate a low-dynamic-range error image during segmentation. Contour coding is replaced by a discontinuity index map data coding step with a maximum of five levels, while texture coding is turned into coding of a low-dynamic-range error image.

Further investigation on improvement of the performance of SLIC is under way, by developing more efficient coding methods for the error image and discontinuity index data parts, and by modifying the region growing procedure. A detailed test of SLIC will be conducted with a larger database of mammograms from the Alberta Screen Test Program and other images. Extension to compression of 3-D medical images is also under investigation.

ACKNOWLEDGMENT

The authors wish to thank AT&T for their freely distributed JBIG package; K. J. Huang and B. Smith, Department of Computer Science, Cornell University, Ithaca, NY, for their freely distributed lossless JPEG package; and M. A. Viergever, Computer Vision Research Group, University of Utrecht, the Netherlands, for the run-time version of the HINT algorithm. They also thank J. E. L. Desautels, Reference Radiologist, Alberta Screen Test Program, for collaborating in this project.

REFERENCES

- [1] N. M. Nasrabadi and R. A. King, "Image coding using vector quantization: A review," *IEEE Trans. Commun.*, vol. 36, pp. 957–971, Aug. 1988.
- [2] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 205–220, 1992.
- [3] J. W. Woods, *Subband Image Coding*. Boston, MA: Kluwer, 1991.
- [4] Y. Fisher, Ed., *Fractal Compression: Theory and Application to Digital Images*. New York: Springer-Verlag, 1994.
- [5] R. D. Dony and S. Haykin, "Neural network approaches to image compression," *Proc. IEEE*, vol. 83, no. 2, pp. 288–303, 1995.
- [6] K. Aizawa and T. S. Huang, "Model-based image coding: Advanced video coding techniques for very low bit-rate applications," *Proc. IEEE*, vol. 83, no. 2, pp. 259–271, 1995.

- [7] P. Roos, M. A. Viergever, M. C. A. van Dijke, and J. H. Peters, "Reversible intraframe compression of medical images," *IEEE Trans. Med. Imag.*, vol. 7, no. 4, pp. 328-336, 1988.
- [8] G. R. Kuduvali and R. M. Rangayyan, "Performance analysis of reversible image compression techniques for high-resolution digital teradiology," *IEEE Trans. Med. Imag.*, vol. 11, no. 3, pp. 430-445, 1992.
- [9] T. V. Ramabadran and K. S. Chen, "The use of contextual information in the reversible compression of medical images," *IEEE Trans. Med. Imag.*, vol. 11, no. 2, pp. 185-195, 1992.
- [10] M. Das and S. Burgett, "Lossless compression of medical images using two-dimensional multiplicative autoregressive models," *IEEE Trans. Med. Imag.*, vol. 12, no. 4, pp. 721-726, 1993.
- [11] S. S. Yu, M. N. Wernick, and N. P. Galatsanos, "Lossless compression of multi-dimensional medical image data using binary-decomposed high-order entropy coding," in *Proc. IEEE Int. Conf. Image Processing*, Austin, Texas, Nov. 1994, pp. 351-355.
- [12] M. Kunt, A. Ikonomopoulos, and M. Kocher, "Second-generation image-coding techniques," *Proc. IEEE*, vol. PROC-73, no. 4, pp. 549-574, 1985.
- [13] "Progressive bi-level image compression," International Telegraph and Telephone Consultative Committee (CCITT), Recommendation T. 82, 1993.
- [14] International Organization for Standards/International Electrotechnical Commission (ISO/IEC), "Progressive bi-level image compression," International Standard 11544, 1993.
- [15] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34, no. 4, pp. 30-44, Apr. 1991.
- [16] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Trans. Inform. Theory*, vol. 23, pp. 337-343, May 1977.
- [17] M. A. Viergever and P. Roos, "Hierarchical interpolation: An efficient method for reversible compression of images," *IEEE Eng. Med. Biol. Mag.*, vol. 12, no. 1, pp. 48-55, 1993.
- [18] P. C. Cosman, R. M. Gray, and R. A. Olshen, "Evaluating quality of compressed medical images: SNR, subjective rating, and diagnostic accuracy," *Proc. IEEE*, vol. 82, no. 6, pp. 919-932, 1994.
- [19] S. Wong, L. Zaremba, D. Gooden, and H. K. Huang, "Radiologic image compression: A review," *Proc. IEEE*, vol. 83, no. 2, pp. 194-219, 1995.
- [20] R. B. Arps and T. K. Truong, "Comparison of international standards for lossless still image compression," *Proc. IEEE*, vol. 82, no. 6, pp. 889-899, 1994.
- [21] J. A. Provine and R. M. Rangayyan, "Lossless compression of panoscanned images," *J. Electron. Imag.*, vol. 3, no. 2, pp. 176-181, 1994.
- [22] J. Vaisley and A. Gersho, "Image compression with variable block size segmentation," *IEEE Trans. Signal Processing*, vol. 40, no. 8, pp. 2040-2060, Aug. 1992.
- [23] F. C. Leou and Y. C. Chen, "A contour-based image coding technique with its texture information reconstructed by polyline representation," *Signal Processing*, vol. 25, pp. 81-89, 1991.

A Rapid and Automatic Image Registration Algorithm with Subpixel Accuracy

Raymond J. Althof, Marco G. J. Wind, and James T. Dobbins, III*

Abstract—A number of digital imaging techniques in medicine require the combination of multiple images. Using these techniques, it is essential that the images be adequately aligned and registered prior to addition, subtraction, or any other combination of the images. This paper describes an alignment routine developed to register an image of a fixed object containing a global offset error, rotation error, and magnification error relative to a second image. The described routine uses sparsely sampled regional correlation in a novel way to reduce computation time and avoid the use of markers and human interaction. The result is a fast, robust, and automatic alignment algorithm, with accuracy better than about 0.2 pixel in a test with clinical computed radiography images.

Index Terms—Digital imaging, digital radiography, dual-energy imaging, image registration.

I. INTRODUCTION

A number of digital imaging techniques in medicine require the combination of multiple images. In certain cases, such as digital subtraction angiography (DSA), subtraction of sequential images is an inherent part of the imaging procedure. In other cases, such as dual-energy imaging with computed tomography (CT) or computed radiography (CR), subtraction may be used to identify functional or anatomical information not readily accessible in single, unsubtracted images. In each of these cases, it is essential that the image information from the constituent images be adequately aligned and registered prior to combination of the images.

In this paper an algorithm is described that was developed to register multiple CR images of fixed objects. There are several imaging techniques presently being investigated that use multiple CR images, such as dual-energy imaging [1], [2] and improvement of detective quantum efficiency (DQE) by use of multiple plates [3]. The geometric misalignment of

individual CR images, even images acquired with multiple plates in a single cassette, is generally much worse than one pixel. Therefore, it becomes critically important to have a routine to adequately align the CR images prior to use with these multiple-image techniques; otherwise, severe edge artifacts occur in the combined images. While the particular examples given in this paper are for registering CR images, the algorithm is sufficiently general that it may be used with any type of image acquisition apparatus.

II. BACKGROUND

There are several approaches that have been investigated for alignment of multiple digital images. In the case of X-ray imaging, such as with CR, fiducial marks have been successfully used [4]–[6]. The fiducial marks are usually some type of lead marker that is placed in the beam and rigidly fixed with respect to the patient. With CR, these fiducial marks have been mounted on the cassette surface when multiple imaging plates are used in a single cassette [5], [6], or have been mounted in the beam on special cassette holder assemblies when multiple cassettes have been used [4]. The images are then processed to find the locations of the fiducial marks. The center-of-mass (or more accurately the center-of-pixel-density) of the pixels is typically used as the location of the fiducial mark [4]. A variation on the center-of-mass method uses a correlation of pixel values in the multiple images to find the displacement of the fiducial marks with respect to each other [5], [6].

Excellent registration results have been obtained with either two [4] or four [5], [6] fiducial marks being used with both the center-of-mass and correlation methods. There are, however, several problems with the use of fiducial marks in practice. First, the technologist must ensure that the fiducial marks are always imaged in the beam. For many exams this would require special placement of the fiducial marks depending on the collimation used. Second, the lack of transmission through overlying anatomy, or the presence of sharp edges in the anatomy close to the marks, may make finding the location of the fiducial mark difficult. Third, the use of fiducial marks requires specialized equipment, such as cassettes with fiducial marks permanently mounted, and therefore is less convenient than approaches that do not require any special equipment or intervention by the user.

There have been other techniques used in image registration that do not require fiducial marks. These techniques typically employ some type of pixel-by-pixel cross correlation or absolute differences [7], [8]. These methods can be quite

Manuscript received May 3, 1996; revised February 13, 1997. This work was supported in part by the National Institutes of Health under Grant R01 CA 55383, a grant from the Stimuleringsprogramma voor Internationalisering van het Nederlands hoger onderwijs (STIR), and a grant from the Universiteitsfonds and the KIVI (Koninklijk Instituut van Ingenieurs). The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. Vannier. *Asperisk indicates corresponding author.*

R. J. Althof was with the Thoracic Imaging Research Division, Department of Radiology, Duke University Medical Center, Durham, NC 27710 USA. He is now with Philips Medical Systems Nederland B.V., Development Quantitative Analysis Software, Best, The Netherlands.

M. G. J. Wind was with the Thoracic Imaging Research Division, Department of Radiology, Duke University Medical Center, Durham, NC 27710 USA. He is now with Rademaker B.V., Halfweg, The Netherlands.

*J. T. Dobbins, III is with the Thoracic Imaging Research Division, Department of Radiology, Duke University Medical Center, Durham, NC 27710 USA (e-mail: jtd@scott.mc.duke.edu).

Publisher Item Identifier S 0278-0062(97)03950-5.

0278-0062/97\$10.00 © 1997 IEEE

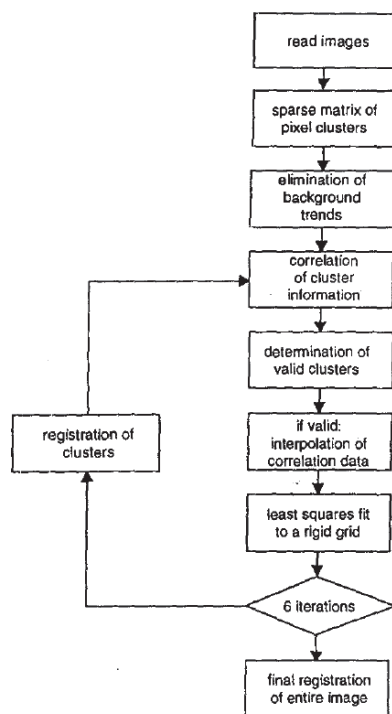


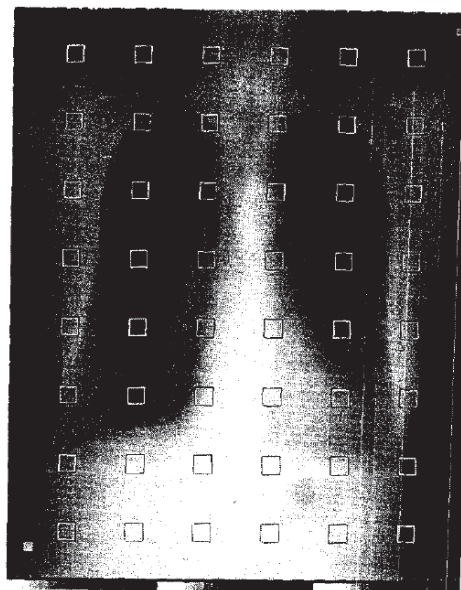
Fig. 1. Flowchart of the registration algorithm.

computationally intensive for large image matrices such as 1760×2140 CR images. Other, more popular, methods are based on structure matching [9]–[11] or control point matching [12]. The first technique will work well if there are high-contrasted structures available in the images. These structures can be found by thresholding or edge-detection routines. The second technique uses edges to define control points. The success of these techniques depends on the ability to isolate the structures of interest satisfactorily in the presence of image noise.

This paper reports on an algorithm that is fast and accurate enough to use for large medical images and does not rely on fiducial marks or specific anatomical landmarks.

III. DESCRIPTION OF THE REGISTRATION ALGORITHM

Fig. 1 is a flowchart of the various components of the alignment algorithm. Local correlation of pixel data between Image I and Image II is computed at specific regions of interest (ROI's) scattered throughout the images. The coordinates of these local correlation peaks are then fit by a least-squares method to a rigid two-dimensional (2-D) grid pattern to deter-


Fig. 2. Location and size of the cluster array in a 1760×2140 digital chest radiographic image.

mine the angle, displacement, and magnification differences between the two images. Each component of the algorithm will be described in detail in this section.

A. Sparse Matrix of Pixel Clusters

The main reduction in computation time with this algorithm is due to the sparse sampling of image data before doing the pixel correlations. We wish to find small ROI's, or clusters, that contain enough local information for reliable correlation, at spacings sufficient to adequately represent the extent of useful information in the image. These clusters need to be distributed over as much of the image as possible to provide the best overall accuracy. If the clusters were all located near the center of the image the error in determining angle and offset would be greater than if they were at the periphery of the images.

Since we desire the algorithm to be as general as possible, it is important to have the clusters distributed uniformly over the image so that the likelihood of identifying significant image information will improve. There is, however, a tradeoff between improved generality of the algorithm with a greater number of clusters and the increase in computation time with more clusters. For the chest CR-images (1760×2140) used in this study, we found that a 6×8 array of clusters, each being 64×64 pixels in size, proved to be sufficiently general and provided an acceptable number of valid clusters.

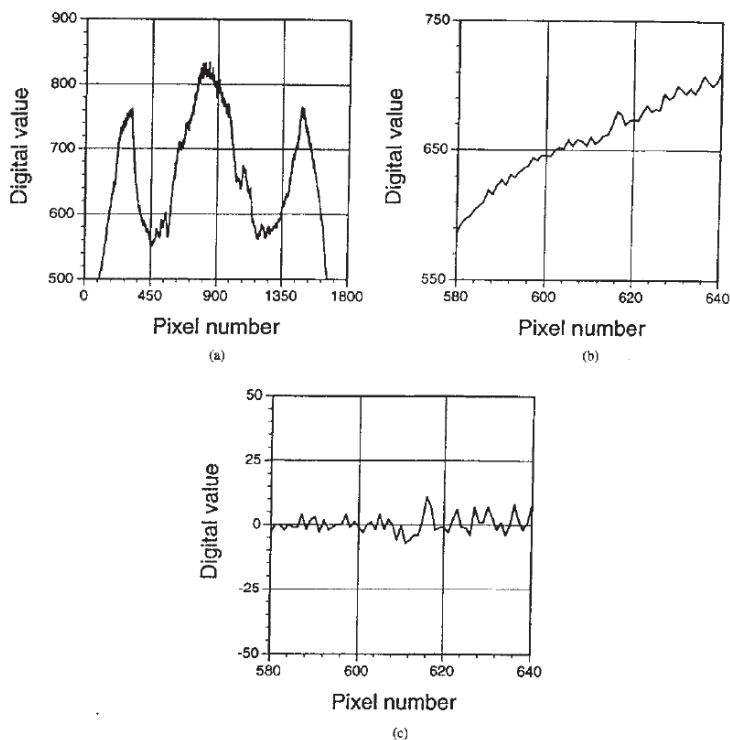


Fig. 3. Elimination of background trends in the cluster data: (a) profile of pixel values across a digital chest radiograph; (b) detail of the profile across the chest radiograph, showing the background trend in the region of one cluster; (c) detail of the profile in the cluster after highpass filtration.

The 64×64 pixel size of each cluster is also a compromise between speed and accuracy. If the clusters are too small, there may not be enough anatomical information to provide reliable correlation data. If the clusters are too large, unnecessary computation time will be required. We determined in initial experiments that a cluster size of 64×64 was a good compromise for chest imaging. The 6×8 array of 64×64 pixel clusters means that only 5% of the total image information is processed in each iteration to determine angle, offset, and magnification values. Fig. 2 shows the location and size of the cluster array used with CR chest images.

B. Elimination of Background Trends

The maximum value of the correlation function for each cluster identifies the approximate displacement between structures in Images I and II at the location of that cluster. However, the correlation function accurately determines the displacement between image structures only if the image background in the clusters is flat. Any nonuniform shading or ramping of pixel values in the background causes the

correlation function to be influenced more by the shading than by the desired anatomical structures.

Fig. 3 illustrates the effect of shading on the results of the image correlation. Fig. 3(a) shows the profile of pixel values across a digital radiographic image of a chest. Note the small-scale anatomical structures and the overall background variation in intensity. For a cluster size which is small relative to the background trend [Fig. 3(b)], the correlation function will maximize the overlap of the background trend rather than the small-scale anatomy. We eliminated the problem of background trend by performing a highpass spatial filtering (11×11 box kernel) of each cluster before calculating the correlation. Fig. 3(c) shows the cluster data after highpass filtering. The algorithm now registers the fine-detail anatomy rather than the background trends, since it is more important to have proper alignment of fine-detail structures.

C. Correlation of Cluster Information

A spatial correlation of pixel data between Image I and Image II was performed for each cluster. The correlation was

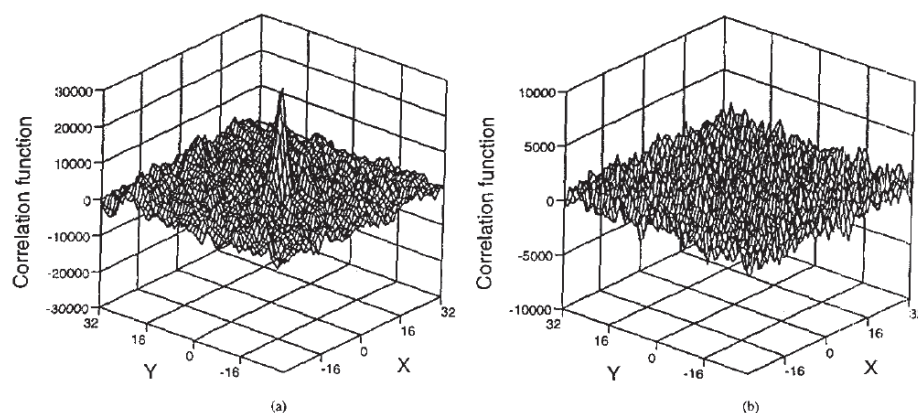


Fig. 4. Correlation function for two clusters: (a) a cluster with very prevalent mid- and high-frequency anatomical information, (b) a cluster with little useful anatomical information (mainly noise). Note that (a) displays a very prominent correlation peak, but (b) does not.

performed in frequency space to save computation time by using simple array multiplication of the Fourier transforms of the clusters

$$A_i(u, v) = \text{FFT}\{a_i(x, y)\} \quad (1)$$

$$B_i(u, v) = \text{FFT}\{b_i(x, y)\} \quad (2)$$

$$\text{CORR}(u, v) = A \cdot B^* \quad (3)$$

$$\text{corr}(x, y) = \text{FFT}^{-1}\{\text{CORR}(u, v)\} \quad (4)$$

where a_i and b_i are the pixel values of cluster i out of Image I and II, respectively. A_i and B_i are the fast Fourier transforms (FFT) of a_i and b_i . The correlation function corresponds to a matrix multiplication of the FFT of one cluster and the complex conjugate of the FFT of the other cluster in frequency space (3). The inverse FFT is then taken to obtain correlation function in Cartesian space (4). The correlation procedure and FFT are adapted from a published source [13].

D. Determination of Valid Clusters

Image noise can give erroneous correlation values in regions where the noise is large compared to the amplitude of the anatomical signal. This was a particular concern for the application of this alignment routine to imaging with multiple CR plates in a single cassette (as in dual-energy imaging). The image noise in the rear plate is considerably worse (two to three times worse) than in the front plate in this case. In regions such as the subdiaphragm where the anatomical information is fairly uniform or where a cluster does not contain any anatomical information (such as outside the patient), there is not a significant correlation peak from the desired signal.

Since the noise is largely Poisson distributed, one would expect that the noise in the two images would be uncorrelated, and would have a uniform (and zero) correlation function in areas without anatomical information. Such an argument would be true for infinite cluster sizes, but the finite and

reasonably small size of the clusters used here did give a correlation function with specific peaks in clusters containing only noise, although the peaks were randomly distributed. In such clusters the size of the local maxima of the correlation function was approximately the same amplitude as the noise in the correlation function, since the pixel values were approximately uncorrelated. Fig. 4 shows the correlation function for two sets of clusters, one with very prevalent anatomical information, and one with essentially only noise.

It was necessary to develop a procedure for evaluating whether the correlation results from a particular cluster were due to reliable image signal or were mainly due to noise. First, the peak value in the correlation function was determined for each cluster. A small region just including the correlation peak was then isolated from the rest of the cluster correlation function. The standard deviation (S.D.) of cluster correlation values outside the isolated region was calculated to determine an approximate measure of uncertainty in the correlation function. The correlation peak was excluded during the calculation so it would not influence the approximation of the uncertainty level. The amplitude of the peak was then compared to the S.D. of correlation values outside of the peak region, and if the ratio were greater than a certain threshold the cluster was considered to have valid, reliable data. Otherwise, the cluster was discarded from further analysis. The determination of the optimum threshold value is discussed in the Results section.

E. Interpolation of Correlation Data

If the cluster was found to be valid, the correlation function was next interpolated up sixteenfold and convolved with a matched filter to allow for partial pixel accuracy. Bilinear interpolation was used. The up-interpolation allowed the use of noninteger sized matched-filter kernels (with respect to the original pixel matrix). Since only the peak of the correlation function contained useful information, we up-interpolated only

an 11×11 pixel region surrounding the peak (interpolated up to 176×176).

We observed that the peaks of the interpolated correlation functions were reasonably described by a specific characteristic shape. We selected a matched filter of about the same size and shape as the average correlation peak. Choosing such a matched filter allowed the signal-to-noise of the correlation peak to be maximized for detection of the peak's location.

F. Least-Squares Fit to a Rigid Grid

The result of the matched filtering of interpolated correlation data is a set of absolute coordinates x_{ij} , y_{ij} of the valid correlation peaks. These coordinates are fitted to a 2-D grid with spacing equal to the distance between cluster centers. Knowledge of the displacements at the grid points allows the rotation and translation to be calculated.

The displacements at all grid values need not be known, since the grid points are rigidly related to one another. For noiseless data, the displacements at only two grid points are sufficient to determine the amount of translation and rotation. However, since our displacement values are obtained from noisy data, better results are obtained by using more than two grid points. A least-squares method was derived to fit the measured x_{ij} , y_{ij} coordinates to the 2-D grid of cluster centers (see Appendix).

A further degree of freedom is the difference in geometric magnification of the two images (such as when imaging two CR plates in a single cassette). The least-squares grid fitting method was adapted to allow for the determination of magnification difference as well as translation and rotation. Determining the magnification was accomplished by allowing the scale of one grid to vary relative to the other grid.

G. Iteration for Improved Accuracy

The results of Section III, parts A-F, are a set of magnification, translation and rotation differences between Images I and II. These results are found to be approximately correct (less than one pixel error), but are not as good as desired (i.e., better than about one-quarter pixel accuracy). By aligning the registered clusters again in successive iterations we achieved much better registration accuracy. It was found that six iterations steps were enough to reach an accuracy better than one quarter pixel.

Because of the closer registration with successive iterations, it is possible that some clusters initially found to be invalid will become valid in subsequent iterations. However, this can be assumed to be true only for clusters that were nearly valid in the first iteration. So, to save computation effort, only the clusters whose correlation signal-to-noise were near or greater than the accepted threshold were checked for validity beyond the first iteration.

Care was taken to minimize cumulative error with the iteration process. If the registered pixel data from one iteration is then re-registered in successive iterations, the accuracy after multiple iterations would likely be dominated by the cumulative error of the interpolative registration routine. Instead, our algorithm always uses the original Image II pixel

data as the starting point for registration, using the improved estimate of angle, offset, and magnification with each iteration. In that way, uncertainties in the registration routine do not accumulate.

Instead of performing the actual registration of the full 1760×2140 images for each iteration, the registration is only performed on the parts of Image II contained within the clusters. Registering only the pixel data in the clusters saves 95% of the computation time required for the full registration.

H. Registration Procedure

Section III, parts A-G, provide the required magnification, offset, and angle values for Image II relative to Image I. Any adequate interpolation can then be employed to do the actual registration of pixel values in Image II to produce the final alignment. A bilinear interpolation scheme was used for the initial evaluation because it is computationally simple. The algorithm calculates for every destination location (x, y) in Image II the fractional source location (x', y') using the values of angle, offset, and magnification. The estimated pixel value at (x', y') is computed from the four nearest neighboring pixel values, using the relative distances to those pixels.

IV. EXPERIMENTAL EVALUATION

The first element of the overall algorithm we evaluated was the threshold value to be used for identifying valid clusters. As we pointed out in Section III, the height of the cluster correlation peak was compared to the value of the background uncertainty in the correlation function. The ratio of correlation peak to correlation uncertainty is indicative of statistical certainty, and a threshold value was chosen to determine whether a given cluster contained reliable data. Several correlation peak-to-noise thresholds were evaluated, based on the range of values found in an initial trial with CR images. The number of valid clusters found and the accuracy of final registration results were measured with each threshold. Only one iteration of the overall algorithm was used for this part of the evaluation.

A similar study was conducted to determine the effect of the shape and size of the convolving function (matched filter) used on the interpolated correlation data. The resulting registration accuracy was measured using both Gaussian and simple box convolving functions. Estimates of the required full-width half-maximum (FWHM) for the Gaussian and box functions were made by examining the shape of a correlation peak acquired using a dual-energy CR image set. FWHM values of 29, 59, and 119 pixels in the interpolated images were evaluated (corresponding to 1.8, 3.7, and 7.4 pixels in the original image data). For this part of the evaluation, a threshold value of six was used for the correlation peak-to-noise as described above.

Next, the convergence properties of the iteration process were measured. A correlation peak-to-uncertainty threshold of six and a convolving box function of 59×59 pixels were used, based on the results of the preceding tests. The angle, magnification and offsets (in x and y) were measured as increasing numbers of iterations were performed.

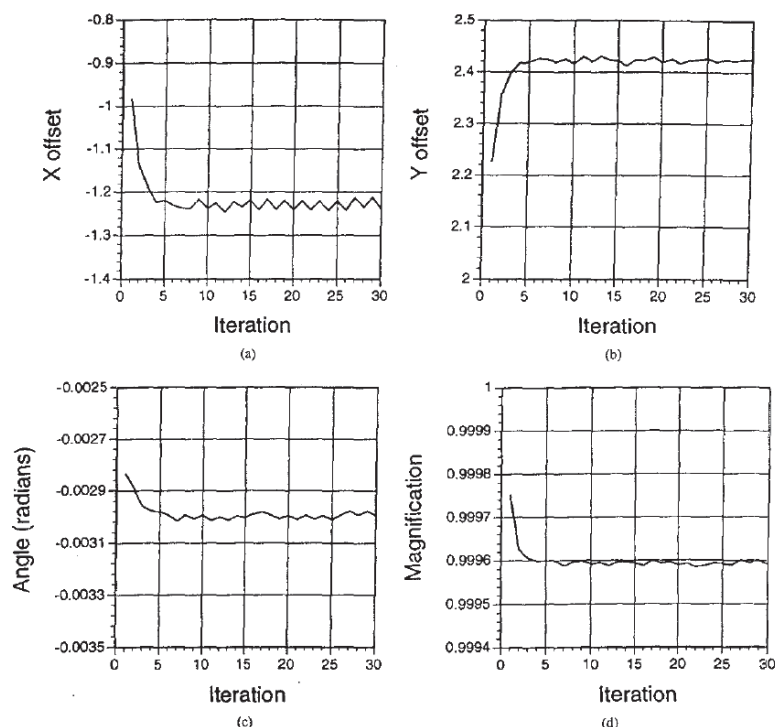


Fig. 5. Convergence trends for (a) x offset, (b) y offset, (c) angle, and (d) magnification with increasing iteration for one clinical image pair. For all images, adequate convergence was reached by the sixth iteration based on the desired level of registration accuracy.

Finally, an estimate of the accuracy of the entire algorithm was made by comparing the results with those from a center-of-mass fiducial mark registration procedure. The registration accuracy was measured using ten patient images acquired for a digital chest radiography study that uses two CR plates in a single cassette [3]. Each image contained two lead fiducial marks placed at opposing corners of the images. (The fiducial marks were not in any of the clusters used in the present algorithm.) The images were registered using the new automated algorithm, and then the center-of-mass algorithm was used to measure the residual misalignment of the two fiducial marks. Residual errors in offset, angle and magnification were then computed.

V. RESULTS

A tradeoff was found with respect to the threshold value for determining valid clusters. Below a threshold of approximately three, many clusters were selected as valid, but some were in fact erroneous. On the other hand, if the threshold value were placed too high, some acceptable clusters were needlessly discarded, occasionally leaving too few clusters for robust

performance (less than three). We found that a threshold value of six (i.e., ratio of correlation peak to correlation S.D.) performed well as a good compromise between sensitivity and specificity for finding valid clusters. After the thresholding step, one is left with a number of presumably valid clusters. A further step is then included in the test for valid clusters. The S.D. of all ϵ_x and ϵ_y values (being the difference between the cluster centers and the correlation peak coordinates) are calculated, and if a given cluster's ϵ_x , ϵ_y values are more than 150% of the S.D. values, then that cluster is regarded as being erroneous. This additional test improved the overall specificity of cluster evaluation.

The evaluation of the size and shape of the matched filter revealed only gradual dependence of registration accuracy on matched filter kernel size, and only a small dependence on filter shape. The rms error in registration accuracy was 0.153, 0.128, and 0.230 pixels for the 29, 59, and 119-pixel FWHM kernels when using the box convolving function, and 0.118, 0.104, and 0.222 pixels error for the comparable FWHM Gaussian functions. As a result, a simple box function of width 59 was adopted. The results with different kernel sizes for the

TABLE I

ERRORS OF THE PRESENT ALGORITHM COMPARED AGAINST A FIDUCIAL MARK CENTER-OF-MASS REGISTRATION ALGORITHM. THE "MAXIMUM OFFSET ERROR" IS THE MAXIMUM DISCREPANCY IN x OR y LOCATION OF THE FIDUCIAL MARKS FOLLOWING THE PRESENT ALGORITHM AND, HENCE, IS A MEASURE OF THE MAXIMUM ERROR OF REGISTRATION IN THE IMAGE. THE FIDUCIAL MARKS USED FOR THE CENTER-OF-MASS ROUTINE WERE IN OPPOSING CORNERS OF THE 1760×2140 IMAGES AND WERE NOT LOCATED IN ANY OF THE CLUSTERS OF THE PRESENT CORRELATION ROUTINE. THE "MAXIMUM DISPLACEMENT ERROR BEFORE REGISTRATION" INDICATES THE MAXIMUM DISCREPANCY BETWEEN CORRESPONDING FIDUCIAL MARK LOCATIONS IN THE IMAGES BEFORE REGISTRATION. ALL ERRORS ARE QUOTED AS ABSOLUTE VALUES.

Clinical Image Pair	Max. displacement error before registration (pixels)	After registration		
		Max. offset Error (pixels)	Angle Error (radians)	Magnification Error
1	1.66	0.10	0.000034	0.000009
2	2.26	0.06	0.000021	0.000018
3	5.25	0.08	0.000001	0.000074
4	4.85	0.07	0.000011	0.000030
5	3.10	0.19	0.000072	0.000078
6	1.20	0.05	0.000006	0.000028
7	0.90	0.21	0.000149	0.000024
8	0.87	0.16	0.000019	0.000090
9	1.65	0.09	0.000012	0.000069
10	1.54	0.16	0.000006	0.000075

matched filter indicates that the kernel size is not very critical. Hence, in retrospect, the up-interpolation may not have been necessary, at least for the type of images investigated here.

The values of angle, magnification and offset each showed strong convergence properties in the ten images evaluated (Fig. 5). In each case, the angle, magnification, and offset errors had satisfactorily reached their asymptotic values within six iterations. (The asymptotic value was determined as the result of 30 iterations.) In each image set evaluated, the convergence was well behaved for each variable (i.e., it converged rapidly with a minimum of oscillation, and it did not change appreciably after the asymptotic value was reached).

The final accuracy of the technique was determined by comparing the results with this new algorithm to a center-of-mass fiducial mark algorithm. While it was difficult to assess the actual accuracy of the fiducial mark technique by itself, fiducial marks were used as the gold standard due to their excellent performance in previous studies. The residual errors in angle, offset, and magnification for the present algorithm were all very small (Table I). The maximum offset error is the larger of the x or y displacement errors between the fiducial marks following registration by the present algorithm; in no case was the error greater than 0.21 pixel. Thus the maximum registration error of any pixel in the image is expected to be <0.21 pixel, since the fiducial marks were near the corners of the image, and the absolute errors nearer the center of the image would be smaller.

The speed of the total algorithm was evaluated on a workstation in both single-processor and dual-processor modes (Sparc-Server 10, Model 512, Sun Microsystems, Mountain View,

CA). For the dual-processor implementation, the algorithm was hand-optimized using multithreaded programming for multiprocessor parallel execution. The evaluation in the single-processor mode was made by simply constraining the code to run on one processor. The algorithm speed was evaluated using two full-size CR images (1760×2140). The execution time for the entire algorithm (including six iterations and bilinear interpolation registration of the full-size image) was 27.6 s for the single-processor and 18.8 s for the dual-processor implementation, including input-output (I/O) time. The actual time used for determining angle, offset, and magnification was approximately 5 s in the dual-processor mode.

VI. DISCUSSION

There are two main factors which limit the overall accuracy of this new algorithm: image noise and matched-filter kernel size. Image noise reduces the number of valid clusters due to stochastic uncertainty in the pixel data set, and therefore increases the uncertainty in the least-squares fit to the 2-D grid. A matched-filter kernel size of 59×59 (with a sixteenfold pixel interpolation interval) was chosen as a reasonable compromise between accuracy and speed, but there was observed some dependence on registration accuracy with kernel size. Fortunately, the variation observed with changes in kernel size was very gradual, and it is not likely that a substantially improved accuracy will result from more precisely identifying the optimum kernel size. The net effect of these two sources of inaccuracy was an error of ≤ 0.21 pixel in the clinical CR images evaluated, which was excellent considering that the two images in each set contained different levels of noise (both were acquired simultaneously in the same cassette; the rear plate received about half of the exposure of the front plate).

The overall robustness of the algorithm appeared excellent for the images tested, since the technique performed well regardless of how misregistered the images were initially. Since no specific anatomical landmarks were used, one would expect that the algorithm would perform equally well with any other type of image, as long as the image contains sufficient medium-to-high spatial frequency information at the locations of the clusters. Furthermore, the presence or absence of antiscatter grids did not seem to make an appreciable difference in registration accuracy, based on further tests of the algorithm on images of an anthropomorphic chest phantom.

The algorithm, as presently implemented, assumes the images are misregistered by only a modest amount (e.g., <10 – 20 pixels). If the images are significantly misregistered, either by angle or offset, then some of the clusters to be correlated in each image may completely miss each other. In such a case, an initial gross alignment of the images may be required before the present algorithm is used to give an extremely accurate final registration. In the clinical CR image data used, however, no such initial gross alignment was required.

One area for further improvement is an automated determination of the number of clusters. If enough valid clusters were not found initially, the cluster matrix could be increased until a sufficient number of valid clusters were obtained. The alteration of cluster matrix is not likely to be necessary if

the algorithm is always used for a single imaging modality (e.g., chest imaging), since the general location of useful image information is somewhat standard. But for a fully automatic algorithm that adjusts to any possible type of image, such an intelligent determination of number of clusters would enhance the generality of the technique.

APPENDIX

Once a number of maxima of correlation have been found, the corresponding locations of the maxima can be used to calculate the angle of rotation θ , the offset, and the factor of magnification difference between the two images. We have developed a way to do a 2-D, least-squares fit of the located maxima to a rigid 2-D grid. Fig. 6 shows how the rigid grid is fit through the set of correlation maxima data points. The error $(\epsilon_{xij}, \epsilon_{yij})$ at each grid point can be described as

$$\begin{pmatrix} \epsilon_{xij} \\ \epsilon_{yij} \end{pmatrix} = \begin{pmatrix} x_{ij} \\ y_{ij} \end{pmatrix} - \begin{pmatrix} x_{g0,0} \\ y_{g0,0} \end{pmatrix} - d \cdot \Psi \cdot \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \cdot \begin{pmatrix} i \\ j \end{pmatrix} \quad (A1)$$

where Ψ is the factor of magnification between the two images, d is the distance between the grid points of the fitting grid, x_{ij} and y_{ij} are the coordinates of the found maxima, and $x_{g0,0}$, $y_{g0,0}$ is the unknown vector of displacement of the $(0,0)$ element of the grid.

Not every found maximum is reliable, as discussed in the text. Therefore, we introduce a factor M_{ij} to describe whether a grid point may be taken into account or not. Making M_{ij} zero for a nonvalid grid point causes the data for that particular cluster to not contribute to the least-squares fitting routine

$$M_{ij} = \begin{cases} 1, & \text{if cluster is valid} \\ 0, & \text{if cluster is not valid.} \end{cases}$$

Applying the least-squares method, we now find the sum of all squared errors $E(\theta, x_{g0,0}, y_{g0,0}, \Psi)$

$$E = \sum_{i=0}^{Q_I-1} \sum_{j=0}^{Q_J-1} M_{ij} \{ [x_{ij} - x_{g0,0} - d \cdot \Psi (i \cos \theta - j \sin \theta)]^2 + [y_{ij} - y_{g0,0} - d \cdot \Psi (i \sin \theta + j \cos \theta)]^2 \} \quad (A2)$$

where Q_I and Q_J are the number of elements in the 2-D grid in the i and j directions, respectively. The minimum sum-of-squares error can be found by calculating the partial derivatives

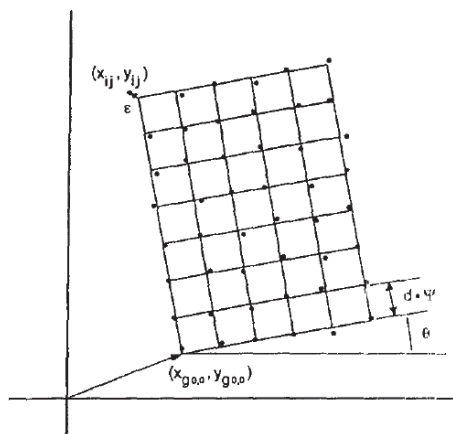


Fig. 6. Rigid grid used for least-squares fit to correlation peaks of individual clusters. For each coordinate (i,j) in the 2-D grid there is a discrepancy (ϵ_x, ϵ_y) between the grid location and the detected correlation peak (x_{ij}, y_{ij}) .

of $E(\theta, x_{g0,0}, y_{g0,0}, \Psi)$ and setting them equal to zero

$$\frac{\partial E}{\partial \theta} = 0; \quad \frac{\partial E}{\partial x_{g0,0}} = 0; \quad \frac{\partial E}{\partial y_{g0,0}} = 0; \quad \frac{\partial E}{\partial \Psi} = 0. \quad (A3)$$

We can now solve the equations for angle, offset, and magnification that minimize $E(\theta, x_{g0,0}, y_{g0,0}, \Psi)$, after introducing some abbreviations for the partial sums in our equations [(A4)–(A6), shown at the bottom of the page], where

$$\begin{aligned} N &= \sum_i \sum_j M_{ij}; & N_I &= \sum_i \sum_j M_{ij} \cdot i; \\ N_J &= \sum_i \sum_j M_{ij} \cdot j; & M_{I^2 J^2} &= \sum_i \sum_j M_{ij} (i^2 + j^2); \\ x_s &= \sum_i \sum_j M_{ij} \cdot x_{ij}; & y_s &= \sum_i \sum_j M_{ij} \cdot y_{ij}; \\ x_I &= \sum_i \sum_j M_{ij} \cdot x_{ij} \cdot i; & x_J &= \sum_i \sum_j M_{ij} \cdot x_{ij} \cdot j; \\ y_I &= \sum_i \sum_j M_{ij} \cdot y_{ij} \cdot i; & y_J &= \sum_i \sum_j M_{ij} \cdot y_{ij} \cdot j. \end{aligned}$$

$$\theta = \tan^{-1} \left(\frac{-x_J + x_s \frac{N_I}{N} + y_I - y_s \frac{N_I}{N}}{x_I - x_s \frac{N_I}{N} + y_J - y_s \frac{N_I}{N}} \right) \quad (A4)$$

$$\Psi = \frac{\sin \theta (-x_J + x_s \frac{N_I}{N} + y_I - y_s \frac{N_I}{N}) + \cos \theta (x_I - x_s \frac{N_I}{N} + y_J - y_s \frac{N_I}{N})}{d [M_{I^2 J^2} - (\frac{N_I^2 + N_J^2}{N})]} \quad (A5)$$

$$\begin{pmatrix} x_{g0,0} \\ y_{g0,0} \end{pmatrix} = \frac{1}{N} \left[\begin{pmatrix} x_s \\ y_s \end{pmatrix} - d \cdot \Psi \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} N_I \\ N_J \end{pmatrix} \right] \quad (A6)$$

ACKNOWLEDGMENT

The authors wish to thank M. Bentum and K. Slump, of the University of Twente, the Netherlands, for many helpful discussions.

REFERENCES

- [1] D. L. Ergun, C. A. Mistretta, D. E. Brown, R. T. Bystritsky, W. K. Sze, F. Kelcz, and D. P. Naidich, "Single-exposure dual-energy computed radiography: Improved detection and processing," *Radiol.*, vol. 174, pp. 243-249, 1990.
- [2] T. Ishigaki, S. Sakuma, and M. Ikeda, "One-shot dual-energy subtraction chest imaging with computed radiography: Clinical evaluation of film images," *Radiol.*, vol. 168, pp. 67-72, 1988.
- [3] D. C. Clark, P. C. Goodman, R. J. Althof, M. G. J. Wind, and J. T. Dobbins, III, "A two-plate technique for improving detective quantum efficiency of computed radiography," presented at 79th Scientific Assembly of Radiological Society of North America, Chicago, Ill., Nov. 28, 1993; *Radiol.*, vol. 189(P), p. 130, 1993.
- [4] J. T. Dobbins, III, J. J. Rice, P. C. Goodman, E. F. Patz, Jr., and C. E. Ravin, "Variable compensation chest radiography performed with a computed radiography system: Design considerations and initial clinical experience," *Radiol.*, vol. 187, pp. 55-63, 1993.
- [5] E. B. Bellers and F. J. de Bruijn, "On the automation of single-exposure dual-energy computed-radiography," Masters thesis, Univ. Twente, Netherlands, 1993.
- [6] R. Arendsen and M. Bentum, "Design and realization of a computer system for high speed single-exposure dual-energy image processing," Masters thesis, University of Twente, Netherlands, 1991.
- [7] W. K. Pratt, "Correlation techniques of image registration," *IEEE Trans. Aerasp. Electron. Syst.*, vol. AES-10, pp. 353-358, 1974.
- [8] D. V. Satish Chandra, J. S. Boland, III, H. S. Ranganath, and W. W. Malcolm, "Multiple image registration using features based on adjacent pixel differences," in *Proc. IEEE Computer Society Conf. Pattern Recognition and Image Processing*, 1982, pp. 524-529.
- [9] A. Della Ventura, A. Rampini, and R. Schettini, "Image registration by recognition of corresponding structures," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, pp. 305-314, 1990.
- [10] G. G. Medioni, "Matching regions in aerial images," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 1983, pp. 364-365.
- [11] A. Goshtasby, G. C. Stockman, and C. V. Page, "A region-based approach to digital image registration with subpixel accuracy," *IEEE Trans. Geosci. Remote Sensing*, vol. 24, pp. 390-399, 1986.
- [12] J. Ton and A. K. Jain, "Registering landsat images by point matching," *IEEE Trans. Geosci. Remote Sensing*, vol. 27, pp. 642-651, 1989.
- [13] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.

MOTION ADAPTIVE CLASSIFIED VECTOR QUANTIZATION FOR ATM VIDEO CODING

Jian Feng,¹ Kwok-Tung Lo² and Hassan Mehrpour³

^{1,2}School of Electrical Engineering, University of New South Wales
Sydney 2052, Australia

³Department of Electronic Engineering
The Hong Kong Polytechnic University, Hong Kong

ABSTRACT

In this paper, a two-layer motion adaptive classified vector quantizer (MACVQ) is proposed for ATM video coding. The proposed method classifies image blocks into three different motion classes: slow-motion, medium-motion and fast-motion. Specified vector quantizers are then designed to encode different motion blocks. A simple edge-based classified vector quantizer is developed to encode the motion-compensated prediction (MCP) error. Compared with the MPEG-type DCT interframe coder, the proposed MACVQ coder has better MSE performance and less accumulation error between consecutive frames.

I. Introduction

Asynchronous Transfer Mode (ATM) has been recommended as the transport vehicle for high speed networks. The ATM technique is versatile and flexible enough to provide a bearer service capable of integrating a wide spectrum of multimedia traffic sources, such as computer data, computer graphic, voice, still image as well as video. With increasing user demand and rapid advance in VLSI technologies, video services like video telephony, video conferencing, video on demand (VOD), video surveillance, and high definition TV (HDTV) will become the dominant traffic in future broadband networks.

For transmission over ATM networks, video data stream is split up into packets. Several problems arise with the introduction of packet video techniques, which mainly result from cell loss problem due to possible network overload. Cell loss problem will cause large degradation in image quality especially when interframe coding schemes are employed, where decoding errors will be accumulated in following frames. To solve this problem, layered coding techniques have been developed [1]-[2]. Most layered coding schemes proposed until now have been mere extensions of well-known video coding standards like H.261 and MPEG-1, which are basically the motion compensated discrete cosine transform (DCT) coding technique. Besides layered

coding techniques, several compensation methods have also been proposed [3]-[4] to tackle cell loss problem.

In this paper, we propose a two-layer motion adaptive classified vector quantization (MACVQ) for interframe coding in ATM environment. Compared with the conventional DCT coder, vector quantization (VQ) [5] [6] has the advantage of simple decoder design which is more suitable for broadcasting services like video on demand (VOD) and HDTV. In the proposed MACVQ method, image blocks are classified into three different motion classes: slow-motion, medium-motion and fast-motion. A simple edge-based classified vector quantizer is then proposed for efficient coding of the motion-compensated prediction (MCP) error.

In the rest of this paper, the proposed two-layer motion adaptive classified vector quantization (MACVQ) technique is first described in section II. Computer simulation results are discussed in section III and some concluding remarks are drawn in section IV.

II. Motion Adaptive Classified Vector Quantization (MACVQ)

The block diagram of the proposed two-layer motion adaptive classified vector quantization (MACVQ) is shown in Figure 1.

In the proposed method, each input frame with size $M \times N$ of an image sequence is first partitioned into subsequent non-overlapping blocks of size $b \times b$. Each block is motion-predicted by applying the full search block matching algorithm [7] into the previous encoded frame. We denote the block of size $b \times b$ as a macroblock. The macroblock is further divided into subsequent $L \times 4$ blocks, where $L = (b/4)^2$. Each block is then classified into one of the three motion classes: slow-motion, medium-motion and fast-motion. When the block is classified as slow-motion block, the MCP error will not be coded and the block will be reconstructed by the motion-predicted block. For the medium-motion block, a edge-based classified vector quantizer is used to encode the MCP error.

Contributed Paper

Manuscript received March 13, 1995

0098-3063/95 \$04.00 © 1995 IEEE

Finally, if the block is a fast-motion block, the pixel elements of the block instead of the MCP error will be encoded using a vector quantizer.

The displaced frame difference (DFD) of the block reflects how large the motion of the block is. If the DFD is large, it means that the motion of the block is large. Therefore, we use the DFD as a criterion to classify the block into different motion classes.

We refer to a block of $m \times m$ ($m=4$) pixels by co-ordinate (k,l) of its upper left corner. For the present frame n , we denote the intensity of the pixel with co-ordinate (i,j) by $F_n(i,j)$. The DFD is then calculated by

$$DFD_{k,l} = \frac{1}{m \times m} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} |F_n(k+i, l+j) - F_{n-1}(k+x+i, l+y+j)|$$

where the offset (x,y) is the motion vector of the macroblock. The block X is classified as a slow-motion block, a medium-motion block or a fast-motion block by inspecting the value of its DFD as follows:

$$X \text{ is } \begin{cases} \text{slow-motion block} & \text{if } DFD \leq T_1 \\ \text{medium-motion block} & \text{if } T_1 < DFD \leq T_2 \\ \text{fast-motion block} & \text{if } DFD > T_2 \end{cases}$$

where T_1 is the lower threshold and T_2 is the upper threshold. When the block is classified as slow-motion block, the MCP error will not be coded and the block will be reconstructed by the motion-predicted block.

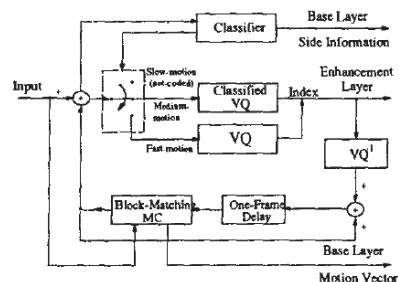


Figure 1 Block diagram of motion adaptive classified vector quantization

A simple classified vector quantization scheme is proposed for coding the MCP error of the medium-motion block. In our proposed method, each block is classified into different types of edge classes by a simple edge classifier. Given a block Y , a two-level bit-map $B(i,j)$ is first generated by checking the MCP error of the block.

$$B(i,j) = \begin{cases} 0 & \text{if } MCP(i,j) < 0 \\ 1 & \text{otherwise} \end{cases}$$

where $MCP(i,j)$ represents the MCP error value at the $(i,j)^{th}$ location. The bit-map reflects the gray level transition information or the edge information of the block. The Euclidean distance between the bit-map of Y and the bit-map of eight prescribed edge classes as shown in Figure 2 is then calculated. Y will be denoted as E_n class when the bit-map of E_n has the minimum distance to the bit-map of Y . Since the bit-map only contains binary information, calculation of the Euclidean distance for two bit-maps only involves exclusive OR operations. Different edge classes will then be vector quantized by different codebooks. Three extra bits are required for each edge class to indicate which codebook is used.

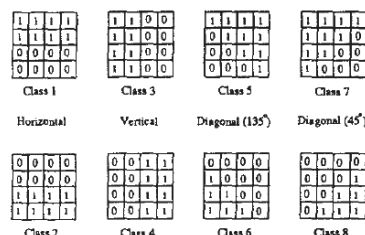


Figure 2 Eight edge classes

The MCP error of the fast-motion block is large, it implies that the motion prediction of the block is not good enough. Therefore, there is little coding gain for coding the MCP error. Instead of the MCP error, the original pixel elements of the block are encoded using a vector quantizer. Extra bits are required to send to the decoder to indicate which class the block is. As most blocks are slow-motion blocks, one bit (0) is used to represent slow-motion class, and two bits are used to represent medium-motion (10) and fast-motion (11) classes.

In the proposed method, the two-layer approach is adopted to minimize the cell loss problem in ATM environment. The image information is divided into two categories, each carrying part of the information needed for full representation of the original image. The motion vector and side information for decoding the MCP error of blocks are packetized and transmitted through the base layer which guarantees no loss. The encoded MCP error data stream is then transmitted in the second or enhancement layer.

III. Simulation Results

Computer simulations have been performed to compare the performance of the proposed MACVQ coder with the MPEG DCT interframe image coder. The test image sequence 'Football' of size 352x240 pixels/frame is used and the simulation is performed on the first 50 frames of the sequence. Although the block size for motion prediction in MPEG is fixed at 16, we perform simulations for motion prediction on the block sizes of 8 and 16. In our experiments, the two thresholds T_1 and T_2 are set to 8 and 20 respectively. For the purpose of fair comparison, the overall bit-rate of the two systems is set to a similar value. The MSE performance for the 'Football' sequence is shown in Figure 3 and the average performance is listed in Table 1.

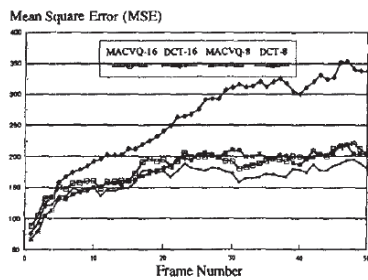


Figure 3 MSE performance of the MACVQ and MPEG DCT interframe coder

Method	MSE	Rate (bpp)
DCT (b=8)	177.19	0.428
MACVQ (b=8)	163.82	0.427
DCT (b=16)	258.47	0.404
MACVQ (b=16)	182.20	0.401

Table 1 Average performance

It is shown that the proposed MACVQ coder has better MSE performance than the MPEG DCT interframe coder for both cases. When the block size is 8, the MACVQ coder performs 8% better than the DCT coder. In the case of block size 16, there is about 30% improvement in MSE performance for the MACVQ coder. It is therefore revealed that the MACVQ can code the MCP error more efficiently than the MPEG DCT interframe coder. It is also noted in Figure 4 that the accumulation of the prediction error in the MPEG DCT coder is more serious than the proposed MACVQ coder especially in the case of block size 16. The MSE performance of the MACVQ coder remains quite constant over a number of consecutive frames. However, the MSE values of the DCT coder become larger when the frame number increases.

Rate Characteristics of the MACVQ Coder

The bit-rate of the proposed two-layer MACVQ coder can be adjusted by varying the lower threshold T_1 . When T_1 becomes larger, more blocks will be classified into slow-motion blocks and need not be encoded. The number of bits required for coding the medium-motion and fast-motion blocks will be decreased. Figures 4 and 5 show the bit-rate characteristics and MSE performance for different settings of T_1 and Table 2 summarizes the average results. In this case, the upper threshold T_2 is fixed at 20 and the block size is 16.

T_1	MSE	$R_{\text{layer 1}}$	$R_{\text{layer 2}}$	R_{total}
8	182.20	0.167	0.234	0.401
14	235.28	0.089	0.115	0.204

Table 2 Average bit-rate (bpp) characteristics

It is noted that the portion of the bit-rate at the base layer and enhancement layer can also be adjusted by the lower threshold T_1 . When it gets larger, more percentage of the information will be packed into the base layer. As shown in Table 2, the bit-rate of the base layer is about 37% and 45% of the total bit-rate when T_1 is set to 8 and 14 respectively.

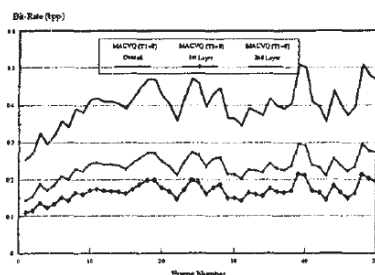
MSE Performance with Cell Loss

The codewords of medium-motion and fast-motion blocks are packetized into fixed-length ATM cells and transmitted via the second layer in ATM networks. When packing into cells, it is important to assure video synchronization in the face of data loss and to localize the impact of loss of encoded data. In our proposed packing method, the following rules are utilized:

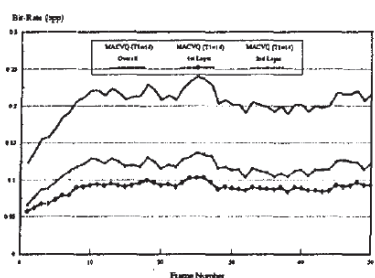
- (1) Block is the basic unit for packing encoded video bit-stream into cells. When the number of bits left in a cell is less than that required for an entire block, the bits are left blank (i.e. putting zeros on them).
- (2) To locate the first block of each cell in the picture. The block location can be identified by two fields: the absolute macroblock address in the picture and its location in the macroblock.

Figure 6 shows our proposed method for packing the encoded video information into AAL-Type 2 information field. There are totally 48 bytes in the information field of an ATM cell. Excluding the overhead information for maintaining cell sequence, error detection and locating the beginning of encoded data, there are 43 bytes left for encoded video information in each ATM cell. Simulations have been performed when there is cell loss occurred in the second layer of the coder. In our experiments, the block size for motion prediction is set to 16 and the two thresholds T_1 and T_2 for the MACVQ coder are

8 and 20 respectively. When the cell loss rate is equal to 10%, the MSE performance is shown in Figure 7 and the average performance is listed in Table 3.



a. $T_1 = 8$



b. $T_1 = 14$

Figure 4 Bit-rate characteristics

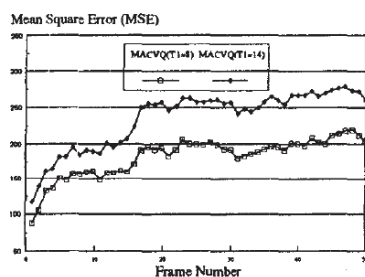
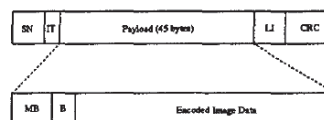


Figure 5 MSE for different thresholds

Method	MSE	Rate (bpp)
DCT	290.58	0.408
MACVQ	206.04	0.405

Table 3 Average MSE with 10% cell loss



SN : sequence number (6 bits)
IT : information type (2 bits)
LI : length indicator (2 bits)
CRC : cyclic redundancy check (10 bits)
MB : address of the beginning macro block (11 bits)
B : address of the beginning block (2 bits)

Figure 6 ATM cell format

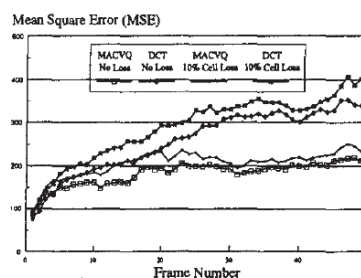


Figure 7 MSE performance with 10% cell loss

It is shown that the MSE increases about 13% when there is 10% cell loss in the second layer. The degradation is not serious since most of the important information is encoded and sent in the first layer.

IV. Concluding Remarks

In this paper, a two-layer MACVQ coder is developed for interframe coding in ATM environment. In our system, image blocks are classified into three motion classes by comparing the DFD of the block with two thresholds. A simple classified vector quantizer is developed to encode the MCP error. Simulation results show that the proposed MACVQ coder is more efficient than the MPEG-type DCT interframe coder.

References

- [1] G. Morrison and D. Beaumont, 'Two-layer video coding for ATM networks,' *Signal Processing : Image Communication*, vol.3, pp.179-185, 1991.
- [2] C. Guillemot, 'Layered coding schemes for video transmission on ATM networks,' *Journal of Visual Communication and Image Representation*, vol.5, no.1, pp.62-74, March 1994.

- [3] Selective Recovery of video packet loss using error concealment,' *IEEE Journal on Selected Areas in Commun.*, vol.7, no.5, pp.807-814, June 1989.
- [4] M. Ghanbari and V. Seferidis, 'Cell-loss concealment in ATM video codecs,' *IEEE Trans. Circuits and Systems for Video Technology*, vol.3, no.3, pp.238-247, June 1993.
- [5] N.M. Nasrabadi, S.E. Lin and Y. Feng, 'Interframe hierarchical vector quantization,' *Proceedings ICASSP'89*, pp.1739-1742, 1989.
- [6] A. Gersho and R. M. Gray, *Vector quantization and signal compression*, Kluwer Academic Publishers, (1992).
- [7] R. Srinivasan and K.R. Rao, 'Predictive coding based on efficient motion estimation,' *IEEE Trans. Commun.*, vol. COM-33, pp.888-896, Aug. 1985.

Jian FENG received her BEng degree with the best student award in Electronic Engineering from Huazhong University of Science and Technology in 1986. She subsequently obtained the MSc degree with the best graduate award in Electronic Engineering from Shanghai Jiao Tong University in

1989. She is now working toward the PhD degree in the school of Electrical Engineering at University of New South Wales. Her research interests are broadband ATM networks, traffic modeling and video communications.

Kwok-Tung LO (M92) was born in Hong Kong in 1964. He obtained his MPhil and PhD degrees in Electronic Engineering from the Chinese University of Hong Kong in 1989 and 1992 respectively. Since 1992, he has been with the Hong Kong Polytechnic University, where he is now an assistant professor at the department of Electronic Engineering. His current research interests include image/video compression, signal processing and multimedia communications systems. Dr. Lo is a member of IEEE and SPIE.

Hassan MEHRPOUR received the BE degree in Electrical Engineering from PTT College in 1971, the MSc degree in Computer Engineering from Boston University in 1980 and the PhD degree in Electrical Engineering from University of New South Wales in 1993. Since 1992, he has been a Lecturer in the department of Engineering at University of New South Wales. His research interests include Video Communications, congestion Control and performance analysis of real-time system, network architecture and system reliability.

PRE-PROCESSING FOR MPEG COMPRESSION USING ADAPTIVE SPATIAL FILTERING

James Goel, David Chan, and Peter Mandl
Genesis Microchip Inc.
200 Town Centre Blvd., Suite 400
Markham, Ontario, Canada L3R 8G5

Abstract - Spatial filtering is an effective method for improving MPEG compression in fixed and variable rate control applications. This paper presents a unique architecture for adaptive spatial filtering based on digital resampling that improves MPEG compression throughput. An overview of a digital resampling and an adaptive spatial filtering system is also presented. Quantitative simulation results indicate that the MPEG compression throughput is greatly improved by digital resampling and spatial filtering without significant loss of image quality.

I. Introduction

The popularity of the Moving Pictures Experts Group (MPEG) digital video compression is fueled by a broad range of increasingly demanding multimedia applications and the rapidly developing HDTV market. Modern computer applications make heavy use of compressed CD-ROM video and future communications systems will take advantage of MPEG compression to provide two-way video conferencing. Compression is the key to ubiquitous digital video.

Much MPEG compression research has been aimed at improving encoding techniques.[3] These systems have attempted to improve compression efficiency by modifying the bit-rate control mechanism to exploit the characteristics of the human visual system. This paper presents two techniques for improving compression throughput using spatial filtering and digital resampling.

The first technique uses spatial filtering to remove redundant high-frequency energy in a video frame prior to compression. A spatially filtered image produces more zero coefficients in the discrete cosine transform stage that are easy to entropy encode. The net result is improved MPEG compression throughput.

The second technique digitally resamples a video frame prior to encoding to reduce the amount of data. Since there is less data, the MPEG compression throughput increases. High-quality digital resampling is required to maintain image fidelity. This paper presents a digital resampling system based on multirate DSP technique. A simple spatial frequency

control mechanism based on a high quality digital resampling system is described followed by a high level system architecture.

Section II presents three specific applications that benefit from enhanced MPEG compression. Section III explains how spatial filtering improves MPEG compression and presents an adaptive spatial filtering architecture based on a multirate digital resampling technique. Section IV describes enhanced MPEG compression system level hardware followed by experimental results in section V. Finally, conclusions are presented in section VI.

II. Three Enhanced MPEG Applications

Video compression development is driven by a need to efficiently transmit and store huge quantities of data in a cost-effective manner. The three following examples illustrate how enhanced MPEG compression can benefit modern applications.

Fixed Bandwidth Satellite Up/Down Link

Consider a digital video broadcasting system consisting of two distant separated groundstations with up link and down link connections to an orbiting satellite. The satellite transmission bandwidth is limited to 10 Mbits/sec and each groundstation transmits 10 compressed video channels occupying 1 Mbit/sec each.

Broadcasters wish to increase the number of video channels handled by the satellite system to maximize revenue. Expanding the satellite's transmission bandwidth is prohibitively expensive because it involves modifying the satellite in space or launching a completely new system. However, spatial filtering techniques can enhance MPEG compression by 10% or more to allow an entirely new video channel in the same 10 Mbits/sec satellite bandwidth. This solution is preferable because a groundstation upgrade can be made quickly and economically. Spatial filtering enhancements require changes to only a few groundstation MPEG encoders and not the many MPEG decoder set-top boxes in the field.

High-Pass Spatial Filtering

High pass spatial filtering is performed in a similar manner. The final image in Frame Store Three produced by the previous technique is subtracted from the original frame in the Input Frame Store with the result stored in Frame Store Two. Frame Store Two now contains the high spatial

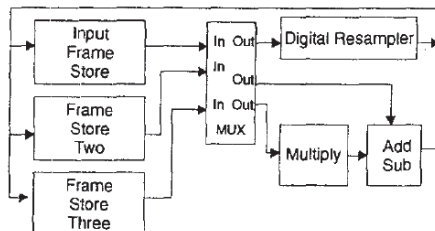


Figure 3 - Adaptive spatial filtering architecture

frequencies previously removed in the low-pass filter stage. These high frequencies can be boosted by multiplying the frame with a fixed value and then adding the data back to the original image. The overall effect of this is to sharpen the edges in the original image.

Bandpass Spatial Filtering

A bandpass filter is implemented by repeating the low-pass operation twice with two different cut-off frequencies and storing the two resulting filtered frames in Frame Store Two and Frame Store Three. The bandpass image is achieved by subtracting the images in Frame Store Two and Frame Store Three. This configuration allows a user to select an exact range of frequencies present in the original image.

This architecture has many benefits. It is flexible and can implement low-pass, high-pass, bandpass and notch spatial filtering without any FIR design or filter coefficients. The amount of spatial frequencies to be removed must be specified. The entire architecture described above relies on a digital resampling technique that removes a precisely specified amount of high spatial frequencies.

Digital Resampling

Real-time digital image resampling independently and arbitrarily changes the number of pixels per line and number of lines per image. Proper image resampling bandlimits the output with digital filters to eliminate visual artifacts. Figure 4 illustrates a multirate digital signal processing system for achieving high-quality digital resampling.

Suppose an original signal $x(n)$ with a sampling rate F' is changed by an arbitrary factor L/M to an output signal $y(m)$ with sampling rate F'' .

$$F'' = \frac{L}{M} \cdot F'$$

Changing the sampling rate is a two stage process. The first **Sample Rate Expander** stage increases the original sampling rate by a factor L to produce an intermediate rate I . This increased sampling rate is implemented by inserting $L-1$ zeros between input samples and then interpolating the zero valued samples with a low-pass digital filter $h_1(k)$.

The **Sample Rate Compressor** stage decreases the intermediate rate I by a factor of M to produce the final sampling rate F'' . The decreased sampling rate is implemented by band-limiting the intermediate signal $s(k)$ by a factor of $1/M$ and discarding $M-1$ samples to prevent aliasing. Typically decimation is implemented by a linear phase finite impulse response (FIR) filter.

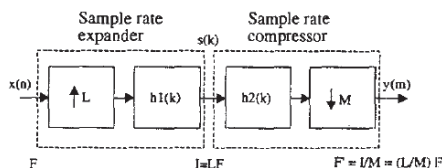


Figure 4 - Changing the sampling rate by a rational factor L/M with a cascade of an interpolator and decimator.

The correct implementation of video resampling is critical to image quality. Figure 5 illustrates how the first repeat spectra of the original signal is shifted to the right as the sampling rate is increased. The shifted spectra leaves behind residual frequencies that must be filtered out to prevent imaging distortions. Figure 6 illustrates how the first repeat spectra of the intermediate image $s(k)$ is shifted to the left as the

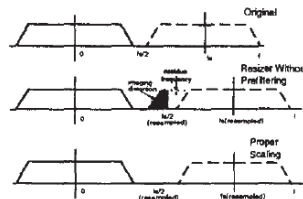


Figure 5 - Increasing the number of samples shifts the repeat spectra to the right leaving behind residual frequencies.

Video Teleconferencing

A second example of enhanced compression requires digital resampling hardware in both the transmitter and receiver. Consider a desktop PC video-conferencing network where each computer has an encoder and decoder to transmit and receive compressed digital video. The encoder can improve compression throughput by digitally resampling to reduce the amount of transmitted data. At the receiving end, the decoder uses the digital resampler to restore the original data rate.

Digital resampling provides a simple compression rate control to allow users to allocate video bandwidth as required. During the course of a meeting, a participant may wish to share information across the same network used to transmit the compressed digital video. Digital resampling prior to compression allows participants to vary the compressed video bandwidth allowing other types of digital data on the same network. After the data has been transferred from one station to the next, the original video bandwidth may be resumed.

HDTV Video Server

Figure 1 illustrates an advanced video server that stores and transmits MPEG-2 compressed data across networks of various capacities. The data is stored on the server as uncompressed video in the maximum HDTV resolution (1920x1080). As each client requests service, digital resampling prior to MPEG-2 compression is used to vary the dimensions of the encoded video to match the resolution supported by the client. This technique ensures that the personal digital assistant has to support only a display resolution of 320x240 while a digital theatre must support full MPEG-2 encoded data resolution of 1920x1080.

III. Spatial Filtering Prior to DCT

The key to the proposed MPEG enhancement lies in adaptively attenuating high spatial frequencies prior to performing the discrete cosine transform (DCT). Figure 2 illustrates how the DCT is performed in an 8x8 pixel grid to produce 64 frequency coefficients. The low-frequency coefficients are positioned in the top left corner and the high-frequency coefficients are positioned in the lower right corner. After the frequency coefficients are quantized, many zero values are produced. These quantized coefficients are entropy encoded in a zig-zag pattern to increase the number of consecutive zeros. Attenuation of high frequencies ensures that many zero valued quantized frequency coefficients are efficiently encoded by subsequent entropy coding stages. These zero valued coefficients are then efficiently compressed and encoded by subsequent entropy coding stages. The following section presents a spatial filtering architecture that adaptively attenuates high spatial frequencies prior to MPEG compression.

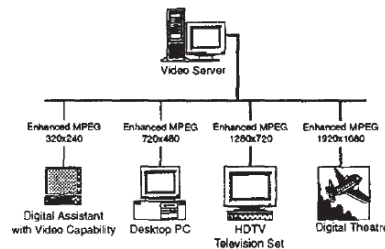


Figure 1 - The digital video server uses enhanced MPEG compression to adapt bandwidth requirements for each client.

Architecture of an Adaptive Spatial Filtering System

The proposed spatial filtering architecture is presented in Figure 3. It is composed of three frame stores, a central multiplexer, digital resampler, multiplier and subtractor. The basic function of this system is to use the digital resampler to remove the specific amount of spatial frequency requested by the user. Digital resampling is the key to the proposed method of effective spatial filtering and is explained in the next section. The multiplexer and frame stores hold intermediate values generated during filter processing. The multiplier and adder/subtractor are used to combine intermediate frames to produce low-pass, high-pass and bandpass spatial filter configurations.

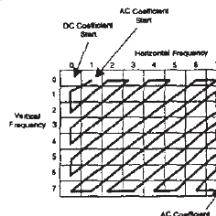


Figure 2 - Spatial filtering prior to the DCT stage increases the number of zero valued coefficients to improve entropy encoding.

Low-Pass Spatial Filtering Configuration

The original input frame is read from the first frame store and digitally resampled to remove high spatial frequencies. The resulting intermediate frame is written to Frame Store Two. This intermediate frame is restored to its original sampling rate by the digital resampler before it is written to Frame Store Three. The combined effect of reducing and then restoring the sampling rate produces a frame that is low-pass spatially filtered to the amount specified by the user.

A COST EFFECTIVE MOTION ESTIMATION PROCESSOR LSI USING A SIMPLE AND EFFICIENT ALGORITHM

Eiji Ogura, Yuuichirou Ikenaga, Yasuhiro Iida,
Yoshitaka Hosoya, Masatoshi Takashima, and Keitaro Yamashita
Sony Corporation
Tokyo, Japan

Abstract

A motion estimation processor LSI has been developed using a simple and efficient algorithm while maintaining the accuracy of full search block matching. Only a single chip is required for performing field and frame real-time motion estimation at half-pel precision. This chip is capable of processing ITU-R601 video sequence for MPEG2 encoding with a 27MHz clock.

1. Introduction

Motion estimation/compensation is an essential technique for today's video compression standards such as MPEG and H.261. The most widely used algorithm is the full search block matching algorithm (FSBMA). A candidate block in the search frame is shifted along all the positions within the search area to find the best match to the reference block of the current frame. The mean absolute error (MAE) is commonly used for the matching criterion and defined as

$$MAE(i,j) = \sum_{x=1}^{M-1} \sum_{y=1}^{N-1} |r(x,y) - s(x+i,y+j)| \quad (1.1)$$

where block size is $M \times N$, $r(x,y)$ is the reference block of the current frame and $s(x+i,y+j)$ is the candidate block within the search area. Motion

vector is the value (i,j) which results in the minimum value of $MAE(i,j)$ within the search area. Extremely high computation is needed to perform this algorithm and the implementation becomes important in a real-time encoding system. Therefore, in order to reduce the total amount of hardware in an encoding system, it is necessary to develop a more efficient motion estimation processor.

In this paper, we will present an efficient and simple algorithm which requires much less hardware than the full search block matching algorithm and describe the motion estimation LSI we have developed using this algorithm.

2. Proposed Algorithm

The integral projection technique[1],[2] is applied to reduce the amount of computation. We are going to denote this first method to be described as the 1p-1p method. Using this method, the number of data used in the calculation for matching at each search point is reduced from 2-dimensional $M \times N$ data of the FSBMA to two 1-dimensional data N and M . Two projections horizontal and vertical are made at each reference and candidate block and used for motion estimation. The horizontal projection(Ph) is made by summing up the whole data within one horizontal line, and the vertical projection(Pv) is made by summing up

the whole data within one vertical line. The horizontal projection (Ph) and the vertical projection (Pv) are defined as follows:

$$Ph = (Ph_0, Ph_1, Ph_2, \dots, Ph_{N-1}), Ph_y = \sum_{x=0}^{N-1} A(x, y) \quad (2.1)$$

: sum of data in one horizontal line

$$Pv = (Pv_0, Pv_1, Pv_2, \dots, Pv_{M-1}), Pv_x = \sum_{y=0}^{M-1} A(x, y) \quad (2.2)$$

: sum of data in one vertical line

where $A(x, y)$ is a pixel within a $M \times N$ block.

Figure 1 shows the two projections made from $M \times N$ block when $M=N=8$. The MAE of the horizontal projection and vertical projection are calculated and used for the matching criterion.

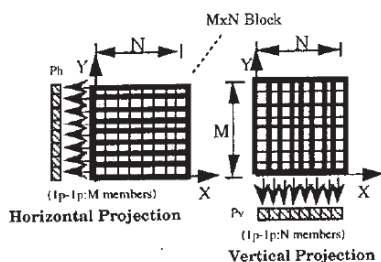


Figure 1: 1p-1p Method ($M=N=8$, 8×8 Block)

In order to reduce the number of data used for matching, adjacent projection data can be summed. We are going to denote these methods as the Sp-Tp methods ($S>1, T>1$) where S adjacent horizontal projection and T adjacent vertical projection data values are summed. In the 1p-1p method, N horizontal projection data and M vertical projection data are used. In the Sp-Tp method, the number of horizontal and vertical projection data is reduced to N/S and M/T respectively. The two projections are

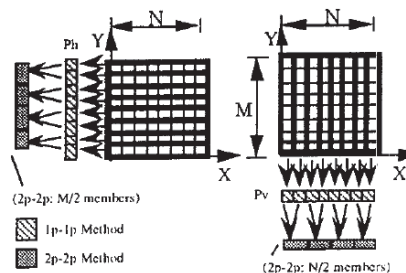


Figure 2: Sp-Tp Method ($M=N=8, S=T=2$)

defined in equations 2.3 and 2.4 and Figure 2 shows an example of the Sp-Tp method when $S=T=2$ and $M=N=8$.

$$Ph = (Ph_0, Ph_1, \dots, Ph_{\frac{N}{S}-1}), Ph_y = \sum_{j=0}^{S-1} \sum_{x=0}^{N-1} A(x, Sy+j) \quad (2.3)$$

: sum of data in S horizontal lines

$$Pv = (Pv_0, Pv_1, \dots, Pv_{\frac{M}{T}-1}), Pv_x = \sum_{i=0}^{T-1} \sum_{y=0}^{M-1} A(Tx+i, y) \quad (2.4)$$

: sum of data in T vertical lines

We performed several simulations to compare the performance between full search block matching and several Sp-Tp methods. Figure 3 compares the loss of luminance(Y) SNR of the 1p-1p method and the several Sp-Tp methods using MPEG2 simulation at 6Mbps, with a 16×16 block. A full exhaustive search and frame based motion vectors were used. Zero dB is the reference value of a full search block matching and each point represents the mean value of seven MPEG test sequences. The figure shows that the coded picture quality decreases as the number of the data used in the matching is reduced. Considering the tradeoff between the estimation accuracy and hardware cost, the 1p-2p method was selected.

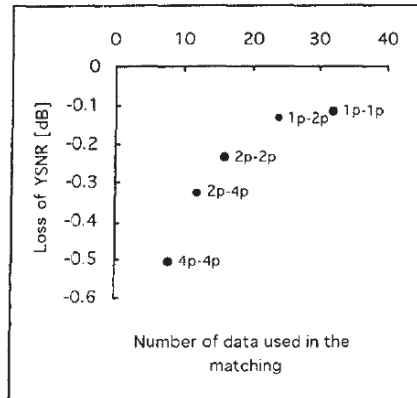


Figure 3 Loss of luminance(Y) SNR vs. number of data used in the matching

3. Chip Architecture

In this section, we will explain the chip architecture which has been developed. Figure 4 shows a simple block diagram of the motion estimation (ME) processor. Both frame and field based motion vectors are estimated in half-pel precision. The motion vectors are estimated in two stages. First, integer-pel precision motion vectors are estimated. Next, the MAE of the eight points around the selected integer-pel precision motion vectors and the integer-pel precision point itself are evaluated. Finally, the vector which has the minimum MAE value is selected as the final motion vector.

3.1 Integer-pel Precision Motion Vector Estimation Block

The 1p-2p method was implemented in the integer-pel precision block. It contains 32 processor elements (PE) to calculate the MAE and operates at 27MHz which is twice of the

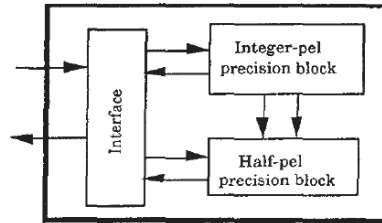


Figure 4 Block diagram of the motion estimation processor

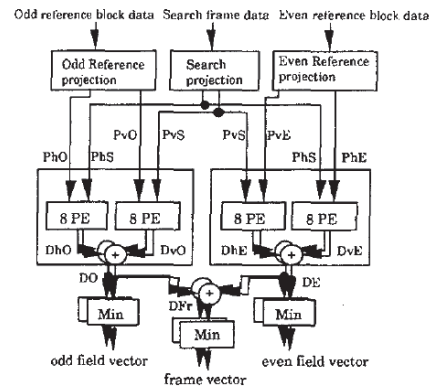


Figure 5(a) Block diagram of integer-pel precision block

pixel rate of ITU-R601 video.

The details of the integer-pel precision block is shown in Figure 5(a). All the calculations are performed on a field basis. The horizontal and vertical projections are calculated for 16x8 odd and even reference blocks. PhO and PvO are horizontal and vertical projections for odd reference block, while PvE and PhE are projections for even reference block. The number of the projection members are eight each. This means that the number of PE is

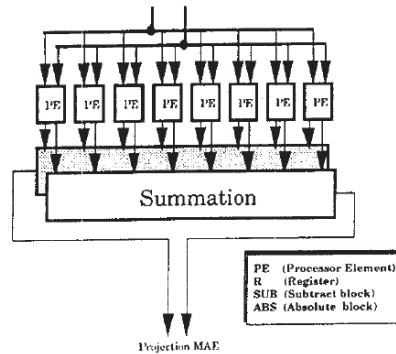


Figure 5 (b) Detail of the 8PE block

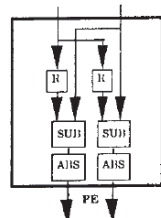


Figure 5 (c) Detail of the PE block

reduced by $1/8(=(8 \times 2)/(16 \times 8))$ compared to the FSBMA. Two projections for both search field are also calculated in a pipe-line manner. The 8PE block contains eight PEs and two summation circuits as shown in Figure 5(b). The detail of the PE block is illustrated in Figure 5(c) and it contains two pairs of subtract (SUB) and absolute (ABS) circuits for processing two adjacent reference data using the same candidate block data. The eight outputs from the PEs are summed at the summation block and each projections MAE are computed. Two MAE for horizontal and vertical projections are obtained for both odd and even fields as shown in equations 3.1 to 3.4.

Next, the sum of both projection's MAE are summed and the MAE of the odd and even reference blocks (DO and DE) are obtained as shown in equations 3.5 and 3.6. DO and DE are used for the matching criteria of the field based motion estimation. DO and DE are then summed and used for the frame motion estimation as shown in equation 3.7. This integer-pel precision ME block is capable of performing motion estimation in the search range of -16 to +15pel in both horizontal and vertical directions corresponding to 1024 search points.

$$DhO = \sum_{i=0}^7 |PhO(i) - PhS(i)| \quad \dots\dots (3.1)$$

$$DvO = \sum_{i=0}^7 |PvO(i) - PvS(i)| \quad \dots\dots (3.2)$$

$$DhE = \sum_{i=0}^7 |PhE(i) - PhS(i)| \quad \dots\dots (3.3)$$

$$DvE = \sum_{i=0}^7 |PvE(i) - PvS(i)| \quad \dots\dots (3.4)$$

$$DO = DhO + DvO \quad \dots\dots (3.5)$$

$$DE = DhE + DvE \quad \dots\dots (3.6)$$

$$DFr = DO + DE \quad \dots\dots (3.7)$$

PhO :Horizontal projection of Odd reference block
PvO :Vertical projection of Odd reference block
PhE :Horizontal projection of Even reference block
PvE :Vertical projection of Even reference block
PhS :Horizontal projection of candidate block
PvS :Vertical projection of candidate block
DhO :MAE of Horizontal projection (Odd field)
DvO :MAE of Vertical projection (Odd field)
DhE :MAE of Horizontal projection (Even field)
DvE :MAE of Vertical projection (Even field)
DO :MAE of Odd reference block
DE :MAE of Even reference block
DFr :MAE of Frame reference block

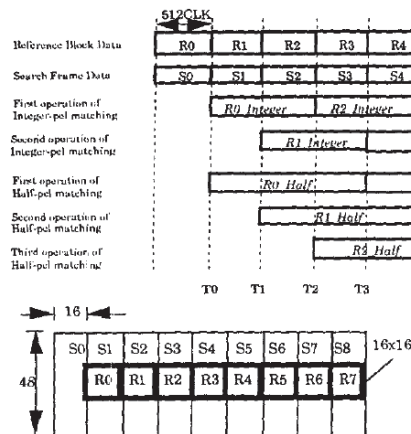


Figure 6 Timing chart of the motion estimation

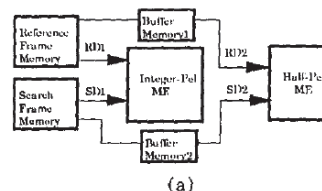
Figure 6 illustrates the timing chart of the motion estimation for both integer-pel and half-pel precision motion vectors. For reference block R0, the search area is S0+S1+S2. For reference block R1, the search area is S1+S2+S3. In the integer-pel ME block, 1024 MAEs have to be calculated and evaluated for each reference block data within the 512 pixel period. Two MAEs for two adjacent reference blocks are computed with two simultaneous operations and evaluated every clock cycle. R0_Integer is the period when the MAE for block R0 is computed and R1_Integer is the period when the MAE for block R1 is computed. As shown in Figure 6, all the circuits are in constant operation and there is no time loss in the pipe-line operation. The half-pel precision motion estimation block is explained in the next section.

3.2 Half-pel Precision Motion Vector Estimation Block

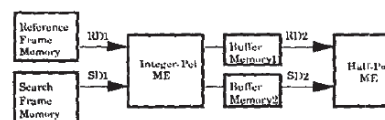
Half-pel motion vectors are obtained using the data surrounding the corresponding integer-pel

motion vectors. First, eight candidate search blocks surrounding the integer-pel motion vector are obtained by interpolation and the MAE of each block is calculated. Finally, the position of the candidate block with the minimum MAE is selected for the half-pel motion vector.

In a conventional configuration as shown in figure 7(a), two kinds of data transfer are necessary. One is the data transfer of reference block data RD1 and candidate search data SD1 to the integer-pel precision ME. The other is the data transfer of reference block data RD2 and candidate interpolating data SD2 to the half-pel precision ME. The ME chip thus has to access t external frame memories twice, which leads to a very high data transfer rate. To overcome this problem, another configuration has been used[3] as shown in figure 7(b). The data RD2 and SD2 used for the half-pel ME block is provided from the integer-pel ME block, so that only RD1 and SD1 data for integer-pel ME block has to be transferred from the external frame memories. However, the SD2 data has to be read out from the buffer memory2 during interpolation.



(a)



(b)

Figure 7 Conventional Configurations

For a block with the dimension of 16×16 , 18×18 data must be read out from the buffer memory². On the other hand, the integer-pel motion vector is obtained within $(16 \times 16 = 256)$ pixel clocks. Therefore, pixel data has to be read out from the buffer memory with clocks of 1.27 times $(18 \times 18 / 16 \times 16 = 1.27)$ of the pixel clock which is the data reading clocks from the external frame memory. However, it is inefficient to operate with clocks that are irrelevant to the pixel clock, thus a better configuration is desirable.

We developed a new efficient architecture which enables the half-pel precision block to operate at the same clock rate as the integer-pel precision block and thus no extra data transfers are needed from the frame memory. The new configuration is illustrated in Figure 8. Integer-pel matching errors (Int_MAE) corresponding to positions in the search area are output from the integer-pel matching circuit in sequence with every clock. When an Int_MAE is output from the integer-pel matching circuit, a pixel necessary for interpolation of its position is output from the FIFO. The delay of the FIFO corresponds to the time necessary for determining the Int_MAE of the first search point after entry of the SD1 data into the projection circuit. The interpolation circuit uses the output data from the FIFO and outputs data of a half-pel position. Computation of half-pel matching error (Half_MAE) begins using the interpolated data and the reference block data RD2. The minimum detecting circuit (Min) detects the minimum MAE. If a new Int_MAE with a smaller value is output from the integer-pel matching circuit, the Min circuit in the integer-pel ME block outputs a control signal to the half-pel matching circuit to initialize it. Then, the calculation of matching error between the data of half-pel

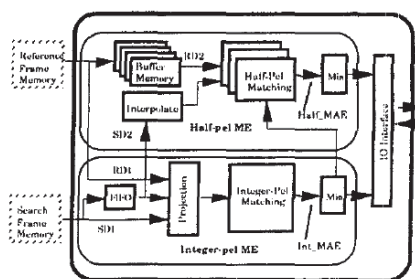


Figure 8 Proposed Configuration

position and the data of the reference block is restarted. In this manner, the integer-pel matching circuit and the half-pel matching circuit operate in conjunction.

As shown in Figure 6, at time T_0 the first Int_MAE of reference block R_0 is obtained and the operation continues until T_2 . This period is labeled as $R_0_Integer$. Half-pel matching operation for obtaining Half_MAE of reference block R_0 also starts at time T_0 and can restart at any time between T_0 and T_2 if a smaller Int_MAE is obtained. The half-pel matching operation may occur between T_0 and T_3 and is denoted as R_0_Half . The period for computing Int_MAE for reference block R_1 is labeled as $R_1_Integer$. As already mentioned, two operations for obtaining two Int_MAEs run in parallel in the integer-pel matching circuit. On the other hand, three operations for obtaining three Half_MAEs run in parallel in the half-pel matching circuit. In summary, by using this proposed configuration for half-pel ME, buffer memory for SD2 is not necessary and both the integer-pel ME and half-pel ME circuits can operate using the same clock and 100% pipe-line efficiency is achievable.

3.3 I/O Interface

The horizontal search area can be easily expanded by cascading one or more ME chips. The specially designed interface allows several chips to be cascaded without the additional hardware required when using conventional ME devices. Figure 9 illustrates the case of using four conventional ME chips to expand the search area in the horizontal direction. The search data has to be delayed between each ME chips using FIFO memories and the several MAE values have to be evaluated at the external minimum circuit (Min). Figure 10 shows the case when four of the proposed ME chips are used. The FIFO memories to delay the search data and a circuit to compare internal MAE and external MAE is included in the interface.

3.4 Characteristic of the ME Chip

The motion estimation processor LSI has been developed as a $0.7\mu\text{m}$ CMOS chip using gate array technology. Table 1 shows the characteristics of the motion estimation processor LSI. The chip size is $14.7\times 14.7\text{mm}$ and has 185K logic gates and 33K bits of RAM.

4. Conclusion

A cost-effective motion estimation processor LSI has been developed using a simple and efficient algorithm while maintaining the estimation accuracy of the commonly used FSBMA. A new efficient architecture for the half-pel precision estimation block is also implemented. Only a single chip is required for performing field and frame real-time motion estimation at half-pel precision in the search range of -16 to $+15$ pel. The search area can be easily expanded by cascading the chip.

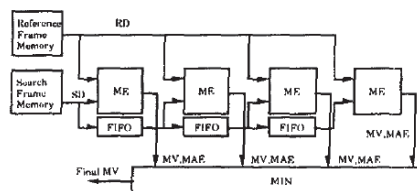


Figure 9. Cascading four conventional ME chips

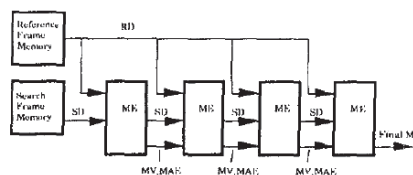


Figure 10. Cascading four new ME chips

Block Size	16x16(Frame), 16x8(Field)
Search Area	-16.5/+15.5 (at half-pel precision)
Process	$0.7\mu\text{m}$ CMOS
Chip size	$14.7\times 14.7\text{mm}$
Frequency	27MHz
Gate count	185K gate of logic, 33K bit of RAM
Power	4.5[w] at 27MHz, 5V

Table 1 Motion estimation processor LSI characteristics

References

- [1] I.H. Lee and R.H. Park, "A Fast Block Matching Algorithm Using Integral Projections" Proc. Tencon '87 Conf. 1987, pp 590-594
- [2] Steve Cain and Ken Sauer, "Efficient Block Motion Estimation Using Integral Projections" IEEE Visual Signal Processing Workshop, Sep. 1992, pp 258-263
- [3] S. Uramoto et al., "A Half-pel Precision Motion Estimation Processor for NTSC-Resolution Video", Proceedings of CICC, pp 11.2.1-11.2.4

Biographies



Eiji Ogura

Assistant Manager of Dept.4, Advanced Development Lab., Consumer A&V Products Company, Sony Corporation.

He received B.E and M.E degree in Electrical Engineering from Tokyo Institute of Technology, Japan in 1987 and 1989, respectively. From 1990 to 1992, he has been a research affiliate at the Media Laboratory of Massachusetts Institute of Technology. He joined Sony Corporation in 1989 and has been engaged in the development of various image processing systems.



Yuuichirou Ikenaga

Manager of Design Engineering Sec., Design Center, Sony Oita Corporation.

He received B.E degree in Electronics from University of Hiroshima, Japan in 1980. He joined Sony Corporation in 1980, and has been engaged in designing of various electrical products, and in 1994 moved to Design Center of Sony Oita Corporation for designing ASICs.



Yasuhiro Iida

Assistant Manager of Hashimoto Signal Processing Lab., Systems & LSI Res.Div., Research Center, Sony Corporation.

He received B.E in Metallurgical Engineering and M.E degree in Electrical Engineering both from The University of Tokyo, Japan in 1980 and 1982, respectively. He joined Sony Corporation in 1982 and has been engaged in research on digital signal processing.

**Yoshitaka Hosoya**

Digital Media Department, TV&Display Products Division,
Consumer A&V Products Company, Sony Corporation.

He received B.E degree Electrical Engineering from Keio Univ.Science and Technology, Japan in 1989. He joined Sony Corporation in 1989 and has been engaged in the development of various digital communication systems.

**Masatoshi Takashima**

Assistant Manager of Dept.4, Advanced Development Lab.,
Consumer A&V Products Company, Sony Corporation.

He received B.E and M.E degree in Electrical Engineering from Musashi Institute of Technology, Japan in 1984 and 1986, respectively. He joined Sony Corporation in 1986 and has been engaged in the development of various image processing systems.

**Keitaro Yamashita**

General Manager of Dept.4, Advanced Development Lab.,
Consumer A&V Products Company, Sony Corporation.

He received his B.S degree in physics from Tokyo University, Japan in 1976. He joined Sony Corporation in 1976 and has been engaged in the development of video recording technology for various consumer VCRs, and various video recording systems.

PREDICTIVE MEAN SEARCH ALGORITHMS FOR FAST VQ ENCODING OF IMAGES

Kwok-Tung Lo
Department of Electronic Engineering
The Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong

Jian Feng
School of Electrical Engineering
University of New South Wales
Sydney, NSW 2052
Australia

ABSTRACT

An improved minimum distortion encoding method called predictive mean search (PMS) algorithm is proposed for image vector quantization in this paper. With the proposed method, the minimum distortion codeword can be found by searching only a portion of the codebook and its relative address to the origin of search is sent instead of absolute address. Two schemes are proposed to transmit the relative address of the minimum distortion codeword. Simulation results show that a significant reduction in both computations and bit-rates is achieved by using the proposed methods.

KEYWORDS : Image coding, vector quantization

1. INTRODUCTION

In the past decade, vector quantization (VQ) [1] [2] has been proven to be effective in image compression. In vector quantization, compression is achieved by matching an input sequence of data samples with the entries (i.e. the codeword) in a pre-classified database known as a codebook. The codeword which produces the minimum distortion is chosen and its index is transmitted to the receiver. The image is then reconstructed at the receiver using a simple table look-up procedure. Compared with other techniques such as transform coding and subband coding, VQ has a very simple decoder which is much suitable for broadcasting video services like video on demand (VOD) and high definition TV (HDTV). However, the large computational load at the encoding process is the main problem to apply VQ in real-time image compression systems.

Many research works have been carried out to reduce the computational complexity of the VQ encoding process. One kind of works is to limit the number of codewords to be searched in the codebook. However, the criterion of minimum distortion can not be guaranteed. Typical examples are tree-search VQ, classified VQ and various hybrid methods [1] [2]. Another kind of works is to improve the computation effectiveness of the full search VQ scheme while maintaining the minimum distortion criterion [3]-[7] by either incorporating a premature exit condition in the search process or re-ordering the codebook according to some criteria.

In this work, an improved minimum distortion encoding algorithm called predictive mean search (PMS) algorithm is proposed for vector quantization of images. The PMS algorithm is based on the fact that the mean value of input block is similar to the mean value of its representative codeword. In other words, the minimum distortion codeword will have similar mean value to the codeword having the closest mean value to the input vector. In the PMS method, the codebook is first re-ordered according to the increasing order of the mean value of the codeword. To exploit high inter-block correlation, the mean value of the input vector is predicted using surrounding pixel elements of neighbour encoded blocks. The codeword having the closest mean value to the predicted mean value will be acted as the origin of the search. A condition is set to restrict the search in a portion of the codebook and maintain the minimum distortion criterion. Once the minimum distortion codeword is found, its relative address to the origin of the search is sent instead of absolute address. Two schemes are proposed to transmit the relative address of the minimum distortion codeword.

In the rest of the paper, the proposed PMS VQ encoding algorithm is presented in section 2. Simulation results using real images are discussed in section 3 and some concluding remarks are drawn in section 4.

2. PREDICTIVE MEAN SEARCH (PMS) ALGORITHMS

2.1 Fast Minimum Distortion VQ Encoding Method

Vector quantization (VQ) is a generalization of scalar quantization to the quantization of a vector. A vector can be used to describe almost any type of pattern simply by forming a vector of samples from the waveform. In vector quantization, an image is divided into subsequent blocks of size $k=N \times N$. Each block is linearly scanned to form a k -dimensional vector. Let $Y = \{y_i, i=1, 2, \dots, M\}$ be a codebook of size M where the codeword $y_i = (y_{i1}, y_{i2}, \dots, y_{ik})$ is a k -dimensional vector. For an input vector $x = (x_1, x_2, \dots, x_k)$, it is required to find the minimum distortion codeword from the codebook Y under the Euclidean distortion measure defined as

Contributed Paper

Manuscript received March 13, 1995

0098 3063/95 \$04.00 © 1995 IEEE

$$d_E(x, y_i) = \sum_{j=1}^k (x_j - y_{ij})^2 \quad (1)$$

It is obvious that the minimum distortion codeword will have similar mean value to the input block, otherwise, there will be large reconstruction error. In other words, the minimum distortion codeword should have similar mean value to the codeword having the closest mean value to the input vector. Therefore, if the search starts with the codeword having the closest mean value and proceeds to the next nearer ones, there is a high probability of arriving at the minimum after only a few steps. To utilize this property to effect a fast searching algorithm, we first sort codewords in the codebook according to the increasing order of the mean value, or equivalently the sum of elements. Consider the squared error of sum of the vector $d_M(x, y_i)$ as below

$$d_M(x, y_i) = \left(\sum_{j=1}^k x_j - \sum_{j=1}^k y_{ij} \right)^2 \quad (2)$$

Once the minimum distortion codeword is found, the following inequality [7] that holds true for any codeword y_i facilitates early termination of the search:

$$d_M(x, y_i) \leq k \cdot d_E(x, y_i) \quad (3)$$

That is, any codeword y_i further in the search whose $d_M(x, y_i)$ is larger than k times the minimum Euclidean distortion $d_{E,min}$ found before can safely be rejected because, from eqn.(3), its own $d_E(x, y_i)$ should be larger than the minimum $d_{E,min}$.

$$d_E(x, y_i) \geq \frac{d_M(x, y_i)}{k} > d_{E,min} \quad (4)$$

Furthermore, the remaining codewords further in the search can also be rejected because their d_E should be even larger. When the q^{th} codeword in the codebook is found to have the closest mean value to the input vector x , it will be acted as the origin of the search for finding the minimum distortion codeword. The search will be performed up and down alternatively as shown in Figure 1. If the condition given in eqn.(3) is valid in either direction, searching will be skipped in this direction and continued in another direction until the best match codeword is found.

2.2 Predictive Mean Search (PMS) Algorithm

In the above method, if the origin of the search can be identified by the decoder, we can send the relative address of the minimum distortion codeword to the starting search point instead of its absolute address. In our proposed PMS algorithm, the mean value of the input block is predicted by averaging nine surrounding pixel elements of previous encoded blocks in the left, upper and upper left directions. This prediction can be performed in both encoder and decoder. The search will then be started at the codeword having the closest mean value to the

predicted mean value. Two schemes are proposed to transmit the relative address of the minimum distortion codeword to the origin of the search.

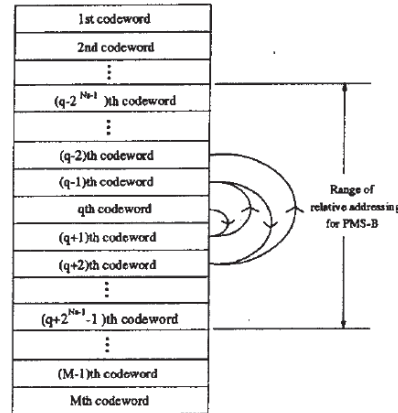


Figure 1 Searching Procedure

Scheme A (PMS-A)

Since most of images are of high correlation, the magnitude of the relative address is non-uniformly distributed and concentrated in some smaller values. Therefore, we can apply a Huffman coder to encode the difference of the address and a number of bits will be saved. However, the complexity of the decoder is increased that it requires to perform the Huffman decoding and the prediction of the mean value beside the table look up procedure.

Scheme B (PMS-B)

In this scheme, the region for relative addressing is fixed to be $[q+2^{N_b-1}-1, q+2^{N_b-1}]$, where q is the origin of the search and N_b is the number of bits to code the relative address and smaller than $\log_2 M$. It is not guaranteed that the minimum distortion codeword is within this area. Therefore, one extra bit is required to tell the decoder whether the minimum distortion codeword is fallen into the range or not. In other words, one bit is to indicate whether the absolute address or the relative address is used. When the minimum distortion codeword is out of the range, absolute address is used. Otherwise, relative address is transmitted to the decoder.

Selection of different values of N_b will affect the overall bit-rate requirement of the system. Two factors are considered: the saving of bits for relative addressing and the proportion of the codewords using relative addressing. Obviously, these two factors are contradicted. When the saving is large (i.e. a small value of N_b), the proportion of the codewords for using relative addressing will be decreased. To

achieve better saving of bits, N_r should be set in order to compromise the two factors. Normally, N_r is set to be 3 or 4 bits less than that required for absolute addressing.

Premature Exit Conditions

The efficiency of the search can be improved by incorporating premature exit conditions. Two premature exit conditions [3] [5] are utilized in our proposed algorithm. Consider an input vector x , let the smallest distortion found before checking y_i be $d_{E,min}$. y_i cannot be a better match than the previous one if

$$|x_j - y_{ij}| > \sqrt{d_{E,min}} \quad \text{for } j \in [1, 2, \dots, k] \quad (5)$$

If the above exit condition [5] is not satisfied, we can perform another test [3]. For any input vector x , if

$$\sum_{i=1}^m (x_i - y_{ij})^2 > d_{E,min} \quad \text{for some } m \in [1, 2, \dots, k] \quad (6)$$

the codeword y_i will not be the best match to x and the remaining elements of the vector need not be calculated.

Obviously, more computations will be saved if those exit conditions valid as early as possible. Therefore, those exit conditions can be applied in the transform domain with further benefits. Since most of the energy of the spatial data is packed into a few low frequency transform coefficients, most of the unlikely codewords can be rejected earlier than in the spatial domain using the above exit conditions. Walsh-Hadamard transform is used in our algorithm for its simplicity. Since the transformation process is only need in the encoding process, no inverse transformation is required at the decoder.

Encoding Process

- Step 1 Perform 2-D Walsh-Hadamard transform;
- Step 2 Predict the mean value of the current block;
- Step 3 Find the codeword y_q having the closest mean value to the predicted mean value;
- Step 4 Start searching from the q^{th} codeword of the codebook. The search will be up and down alternatively.
- Step 5 Skip the search if either of the premature exit conditions is valid;
- Step 6 Obtain the minimum Euclidean distortion codeword y_p ;
- Step 7 Calculate the difference of the index, i.e. $(q-p)$;
- Step 8 Goto step 9, 10 for scheme A, B respectively.
- Step 9 Huffman encoded the index difference $(q-p)$ and send to the decoder. Goto step 13;
- Step 10 If $-2^{N_r-1} \leq (q-p) \leq 2^{N_r-1} - 1$ then goto step 11, otherwise goto step 12.

Step 11 Send the N_r -bits relative address $(q-p)$ preceded by a '1' to the decoder, goto step 13;

Step 12 Send the absolute address p preceded by a '0' to the decoder;

Step 13 Goto next block;

Decoding Process

- Step 1 Predict the mean value of the current block;
- Step 2 Find the codeword y_q having the closest mean value to the predicted mean value;
- Step 3 Goto Step 4 and 6 for Scheme A and B respectively;
- Step 4 Perform Huffman decoding and obtain the relative address $(q-p)$;
- Step 5 Goto Step 7;
- Step 6 if the first bit is 1, obtain the N_r -bits relative address $(q-p)$;
- Step 7 Obtain the absolute address and reconstruct the block by table look-up;
- Step 8 Goto Step 1 (next block);

3. SIMULATION RESULTS

Computer simulations using real images have been carried out to evaluate the efficiency of the proposed VQ encoding algorithms. In our experiments, the codebook is generated by the well-known LBG algorithm [1] with size M equals to 256 and the vector dimension k equals to 16. The training set data consists of eight standard images which are of size 256×256 with 8-bit resolution per pixel. Test images Lena and Sailboat are outside the training set while other test images Peppers and Woman are inside the training set. The Huffman table adopted in the JPEG baseline system to encode the DC difference signal is used to encode the relative address in PMS-A. In our experiments, the number of bits required for coding the relative address of PMS-B (N_r) is set to 4.

We compare the proposed algorithms with several existing encoding methods [3]-[7] in terms of computational complexity and bit-rate requirement. The number of computations per pixel element and the bit-rate requirement of different encoding methods for the four test images are presented in Table 1. All the reported algorithms are satisfied with the minimum distortion criterion, therefore, they result same MSE value for the same picture.

From Table 1, we see that the two PMS algorithms require the least number of multiplication as well as the total computation among other encoding algorithms. It is noted that operations of Huffman table look-up for PMS-A is not taken into account in the table. However, the operations required for the preprocessing, such as Walsh-Hadamard transform,

MSE = 130.38					
Method	N_{Melt}	N_{Add}	N_{Comp}	N_{Total}	Rate(bpp)
FS	256	496	15.94	767.94	0.500
[3]	52.90	89.81	52.90	195.61	0.500
[4]	44.73	73.46	44.73	162.92	0.500
[5]	22.78	95.03	98.03	215.84	0.500
[6]	39.43	62.82	38.43	140.68	0.500
[7]	24.43	37.27	25.08	86.78	0.500
PMS-A	10.45	42.99	31.35	84.79	0.399
PMS-B	10.45	43.12	31.48	85.05	0.434

a. Lena

MSE = 119.06					
Method	N_{Melt}	N_{Add}	N_{Comp}	N_{Total}	Rate(bpp)
FS	256	496	15.94	767.94	0.500
[3]	50.56	85.12	50.56	186.24	0.500
[4]	45.16	74.32	45.16	164.64	0.500
[5]	19.92	86.20	88.82	194.94	0.500
[6]	38.43	62.82	38.43	139.68	0.500
[7]	23.57	35.65	24.12	83.34	0.500
PMS-A	9.97	42.08	30.64	82.69	0.404
PMS-B	9.97	42.21	30.77	82.95	0.432

b. Peppers

MSE = 252.93					
Method	N_{Melt}	N_{Add}	N_{Comp}	N_{Total}	Rate(bpp)
FS	256	496	15.94	767.94	0.500
[3]	56.78	97.56	56.78	211.12	0.500
[4]	59.02	102.03	59.02	220.27	0.500
[5]	24.07	101.41	104.62	230.10	0.500
[6]	48.55	81.08	47.55	177.18	0.500
[7]	27.23	42.23	28.52	97.98	0.500
PMS-A	14.34	49.98	37.04	101.36	0.418
PMS-B	14.34	50.11	37.17	101.62	0.435

c. Sailboat

MSE = 446.35					
Method	N_{Melt}	N_{Add}	N_{Comp}	N_{Total}	Rate(bpp)
FS	256	496	15.94	767.94	0.500
[3]	43.61	71.21	43.61	158.43	0.500
[4]	31.48	46.95	31.48	109.91	0.500
[5]	17.65	74.35	76.71	168.71	0.500
[6]	28.53	41.01	27.53	97.07	0.500
[7]	20.71	30.63	20.57	71.91	0.500
PMS-A	5.78	33.75	23.91	63.44	0.328
PMS-B	5.78	33.88	24.04	63.70	0.385

d. Woman

Table 1 Computational requirement (per pixel element) and bit-rate requirement

prediction of mean value, are all included in the table. Taking the image Lena as an example, the total computations required for PMS-A and PMS-B are 84.79 and 85.05 respectively, which is only about 11% of that required by the full search method.

Since absolute addressing is used in the full search and other encoding methods, the number of bits required to transmit the index for each block is equal to $\log_2 M$ and the bit-rate is fixed at $\log_2 M/k$, i.e. 0.5 bpp in this case. However, the bit-rates (bpp) required for test images (Lena, Peppers, Sailboat, Woman) as shown in Table 1 are (0.399, 0.404, 0.418, 0.328) and (0.434, 0.432, 0.435, 0.385) for PMS-A and PMS-B respectively. It implies that there is about 12 to 35 % reduction in bit-rate requirements when the two PMS algorithms are used. It is noted that PMS-A results the maximum bit-rate reduction among the others. In PMS-A, there is no restriction for the range of the relative address and no need to send an overhead bit to the decoder as PMS-B to indicate whether relative address is used or not. Therefore, PMS-A can achieve the maximum savings in bit-rate requirement. Comparing the system complexity of the two PMS algorithms, PMS-A is the most complicated since Huffman decoding is required at the decoder. When considering both the compression ratio and system complexity, PMS-B is preferred.

4. CONCLUSIONS

Two PMS algorithms are proposed for VQ encoding of images in this paper. In the proposed algorithms, the codebook is first re-ordered according to the increasing order of mean value of the codeword and the search is started at the one having the closest mean value to the predicted mean value of the input block. The relative address of the minimum distortion codeword to the origin of the search is sent to the decoder. Simulation results show that the proposed algorithms are very efficient. They not only require the least number of computations but also achieve a certain amount of bit-rate reduction when compared with several existing VQ encoding algorithms.

5. REFERENCES

- [1] N.M. Nasrabadi and R.A. King, 'Image coding using vector quantization : a review,' *IEEE Trans. Commun.*, vol. COM-36, pp. 957-971, Aug. 1988.
- [2] A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Press, 1992.
- [3] C. Bei and R.M. Gray, 'An improvement of the minimum distortion encoding algorithm for vector quantization', *IEEE Trans. Commun.*, vol. COM-33, pp.1132-1133, Oct. 1985.

- [4] K.K. Paliwal and V. Ramasubramanian, 'Effect of ordering the codebook on the efficiency of the partial distance search algorithm for vector quantization', *IEEE Trans. Commun.*, vol. COM-37, pp.538-540, May 1989.
 - [5] M.R. Soleymani and S.D. Morgera, 'A fast MMSE encoding technique for vector quantization', *IEEE Trans. Commun.*, vol. COM-37, pp.656-659, June 1989.
 - [6] J. Ngwa-Ndifer and T. Ellis, 'Predictive partial search algorithm for vector quantization', *Electronics Letters*, vol.27, pp.1722-1723, Sep. 1991.
 - [7] S.W. Ra and J.K. Kim, 'Fast weight-ordered search algorithm for image vector quantization', *Electronics Letter*, vol.27, pp.2081-2083, Oct. 1991.
- Kwok-Tung LO (M92) was born in Hong Kong in 1964. He obtained his MPhil and PhD degrees in Electronic Engineering from the Chinese University of Hong Kong in 1989 and 1992 respectively. Since 1992, he has been with the Hong Kong Polytechnic University, where he is now an assistant professor at the department of Electronic Engineering. His current research interests include image/video compression, signal processing and multimedia communications systems. Dr. Lo is a member of IEEE and SPIE.
- Jian FENG received her BEng degree with the best student award in Electronic Engineering from Huazhong University of Science and Technology in 1986. She subsequently obtained the MSc degree with the best graduate award in Electronic Engineering from Shanghai Jiao Tong University in 1989. She is now working toward the PhD degree in the school of Electrical Engineering at University of New South Wales. Her research interests are broadband ATM networks, traffic modeling and video communications.

Scene-Context-Dependent Reference-Frame Placement for MPEG Video Coding

Austin Y. Lan, Anthony G. Nguyen, and Jenq-Neng Hwang, *Senior Member, IEEE*

Abstract—The MPEG video-compression standard effectively exploits spatial, temporal, and coding redundancies in the algorithm. In its generic form, however, only a minimal amount of scene adaptation is performed. Video can be further compressed by taking advantage of scenes where the temporal statistics allow larger interreference-frame distances. This paper proposes the use of motion analysis (MA) to adapt to scene content. The actual picture type (intra-coded (I), predicted (P), or bidirectionally coded (B)) decision is made by examining the accumulation of motion measurements since the last reference frame (either I or P) was labeled. The proposed MA-based adaptive-reference-frame-placement scheme outperforms the standard fixed-reference-frame-placement and adaptive schemes based on histogram of difference. When compared with the standard fixed scheme, depending on the video contents, this proposed algorithm can achieve from 2 to 13.9% savings in bits while maintaining similar quality.

Index Terms—Block-matching algorithm (BMA), context-dependent reference frame, mean square prediction error (MSPE), motion analysis (MA), motion vector interpolation, MPEG, scene changes, scene context, variable bit rate (VBR).

I. INTRODUCTION

THE MPEG video-compression standard addresses many of the important issues relating to the exploitation of redundancies in video. Spatial, temporal, and coding redundancies are all taken advantage of in the algorithm. In its generic form, however, only a minimal amount of scene adaptation is performed. Indeed, a fixed arrangement of picture types, or temporal decorrelation structures, is the most widely used encoding method. Encoding in this manner reduces complexity but does not allow for changing temporal statistics. Video can be further compressed by taking advantage of scenes where the statistics allow larger interreference-frame distances. The additional compression over the fixed picture arrangement style comes from the reduction in the number of reference frames spread throughout the coded sequence. Reference frames, by nature, have a more significant allocation of bits in order to provide high quality for prediction.

Scene-context-dependent coding requires sensitivity to the changing motion content and to abrupt changes. Each scene in a video sequence demands a different bit rate and organization

of temporal decorrelation structure according to the amount and magnitude of object and background motion. Intuitively, high motion requires more frequent placement of reference "P" (predicted) frames to uphold quality. Low motion, on the other hand, allows larger distances between references since the correlation remains high over a larger span of frames. The frames in between reference frames are coded as "B" (bidirectionally coded) frames. Abrupt changes in content, or temporal discontinuities, require prediction dependencies to be reset to "I" (intra-coded) frames in order to avoid severe coding errors.

To take advantage of scene content, one needs to perform sequence decomposition based on identifying 1) scene changes and 2) significant changes within a scene that would require reference-frame placement [4]. Scene changes require a reference frame that does not depend on temporal prediction. Since each scene decides the bit rate differently, it is important to be able to effectively identify where a particular scene begins and ends. The typical strategy for handling scene changes is to examine the current frame and the immediate past frame. If there is a significant change in statistics, then this is labeled as a scene change.

The need to track gradual changes within a scene reflects the ever changing levels of motion in all image sequences. Limits must be imposed on consecutive reference-frame distances due to the inverse relationship between temporal prediction accuracy and distance. As the natural progression of images in a sequence reveals significant changes in correlation, reference frames should be set accordingly. The typical strategy is to analyze the current frame and the last reference frame, which could be several frames in the past. If the chosen measure exceeds some threshold, then the current frame is labeled a reference.

A. Existing Methods

Large interreference-frame distances and temporal masking in the human visual system (HVS) are taken advantage of by Lee and Dickinson [2], [4]. Their dynamic, scene-adaptive MPEG encoder defines three picture types as the standard dictates, but each can have either full (I1, P1, B1) or reduced (I2, P2, B2) bit allocation. The reduced-allocation frames exploit the forward and backward temporal masking effect in human vision by coding the first frame of a new scene as a coarse independent I2 reference and the last frame of the previous scene as a coarse causal prediction P2 reference. For scene adaptation, they use a *difference of histograms* (DOH) distance metric on the luminance frame as a measure

Manuscript received October 26, 1996; revised November 1, 1998. This paper was recommended by Associate Editor T. Chen.

A. Y. Lan was with the Information Processing Laboratory, Department of Electrical Engineering, University of Washington, Seattle, WA 98195 USA. He is now with Eastman Kodak Co., Rochester, NY 14620 USA.

A. G. Nguyen and J.-N. Hwang are with the Information Processing Laboratory, Department of Electrical Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: hwang@ee.washington.edu).

Publisher Item Identifier S 1051-8215(99)02951-1.

between server i and server k for stripe j can be expressed as

$$\delta_{i,k,j} = (t_0 + d_i + (v_i + 2 + j)T_F) - (t_0 + d_k + (v_k + 2 + j)T_F). \quad (46)$$

Substituting (45) into (46), we have

$$\delta_{i,k,j} = \left(d_i + \left\lfloor \frac{t_{\text{new}} - (t_0 + d_i)}{T_F} \right\rfloor T_F \right) - \left(d_k + \left\lfloor \frac{t_{\text{new}} - (t_0 + d_k)}{T_F} \right\rfloor T_F \right). \quad (47)$$

Without loss of generality, we can assume $d_i \geq d_k$ and let $H = t_{\text{new}} - (t_0 + d_i)/T_F$. Then we have

$$\begin{aligned} \delta_{i,k,j} &= (d_i + \lfloor H \rfloor T_F) - \left(d_k + \left\lfloor H + \frac{(d_i - d_k)}{T_F} \right\rfloor T_F \right) \\ &= (d_i - d_k) + T_F \left(\lfloor H \rfloor - \left\lfloor H + \frac{(d_i - d_k)}{T_F} \right\rfloor \right). \end{aligned} \quad (48)$$

Noting that $\lfloor x + y \rfloor \geq \lfloor x \rfloor + \lfloor y \rfloor$, we have

$$\begin{aligned} \delta_{i,k,j} &\leq (d_i - d_k) + T_F \left(\lfloor H \rfloor - \left\lfloor H + \frac{(d_i - d_k)}{T_F} \right\rfloor \right) \\ &= (d_i - d_k) - T_F \left\lfloor \frac{(d_i - d_k)}{T_F} \right\rfloor. \end{aligned} \quad (49)$$

Last, making use of the result that $\lfloor x/y \rfloor y \geq (x - y)$, we can then obtain

$$\begin{aligned} \delta_{i,k,j} &= (d_i - d_k) + T_F \left\lfloor \frac{(d_i - d_k)}{T_F} \right\rfloor \\ &\leq (d_i - d_k) - ((d_i - d_k) - T_F) \\ &= T_F \end{aligned} \quad (50)$$

and the result follows. \square

Proof of Theorem 2: Let the new-session request arrive at the admission scheduler at time t during group $v_{\text{new}} = \lfloor tG/T_F \rfloor$. Then due to clock jitter, the current group at other servers can range from $\lfloor (t - \tau)G/T_F \rfloor$ to $\lfloor (t + \tau)G/T_F \rfloor$. To guarantee that the assigned group has not started in any of the servers implies assigning a group larger than the largest current group in any of the servers, i.e., $s_{\text{new}} = \lfloor (t + \tau)G/T_F \rfloor + 1$. Applying the inequality $\lfloor x + y \rfloor \leq \lfloor x \rfloor + \lfloor y \rfloor$, we have

$$s_{\text{new}} \leq \lfloor tG/T_F \rfloor + \lfloor \tau G/T_F \rfloor + 1. \quad (51)$$

Substituting into v_{new} , we have $s_{\text{new}} \leq v_{\text{new}} + \lfloor \tau G/T_F \rfloor + 1$, and the result follows. \square

ACKNOWLEDGMENT

The author would like to thank the reviewers for their constructive comments in improving this paper for its final form.

REFERENCES

- [1] T. C. Chiueh, C. Venkatramani, and M. Vernick, "Design and implementation of the stony brook video server," Computer Science Department, State University of New York at Stony Brook, Tech. Rep. TR-16, Aug. 1995.
- [2] F. A. Tobagi and J. Pang, "StarWorks—A video applications server," in *Proc. IEEE COMPCON Spring'93*, 1993, pp. 4–11.
- [3] R. Buck, "The oracle media server for nCube massively parallel systems," in *Proc. 8th Int. Parallel Processing Symp.*, 1994, pp. 670–673.
- [4] H. Taylor, D. Chin, and S. Knight, "The magic video-on-demand server and real-time simulation system," *IEEE Parallel Distrib. Syst.*, vol. 3, no. 2, pp. 40–51, 1995.
- [5] C. Griwodz, M. Bar, and L. C. Wolf, "Long-term movie popularity models in video-on-demand systems or the life of an on-demand movie," in *Proc. Multimedia'97*, pp. 349–357.
- [6] T. D. C. Little and D. Venkatesh, "Popularity-based assignment of movies to storage devices in a video-on-demand system," *ACM Multimedia Syst.*, vol. 2, no. 6, pp. 280–287, 1995.
- [7] Y. B. Lee, "Parallel video servers—A tutorial," *IEEE Multimedia Mag.*, vol. 5, no. 2, pp. 20–28, 1998.
- [8] D. J. Gemmell, H. M. Vin, D. D. Kandlur, P. V. Rangan, and L. A. Rowe, "Multimedia storage servers: A tutorial," *IEEE Comput. Mag.*, vol. 28, pp. 40–49, May 1995.
- [9] A. L. N. Reddy and J. C. Wyllie, "I/O issues in a multimedia system," *IEEE Comput. Mag.*, vol. 27, pp. 69–74, Mar. 1994.
- [10] D. Mills, "Internet time synchronization: The network time protocol," *IEEE Trans. Commun.*, vol. 39, pp. 1482–1493, Oct. 1991.
- [11] Z. Yang and T. A. Marsland, Eds., *Global States and Time in Distributed Systems*. Los Alamitos, CA: IEEE Computer Society Press, 1994.
- [12] P. S. Yu, M. S. Chen, and D. D. Kandlur, "Grouped sweeping scheduling for DASD-based multimedia storage management," *ACM Multimedia Syst.*, vol. 1, pp. 99–109, 1993.
- [13] J. N. Lloyd and K. S. Samuel, *Urn Models and Their Application*. New York: Wiley, 1997, pp. 125–126.
- [14] Y. B. Lee and P. C. Wong, "VIOLA—Video on local area networks," in *Proc. 2nd ISMM/MASTED Int. Conf. Multimedia Systems and Applications*, Stanford University, Stanford, CA, Aug. 1995, pp. 101–104.
- [15] C. Bernhardt and E. Biersack, "The server array: A scalable video server architecture," *High-Speed Networks for Multimedia Applications*. Norwell, MA: Kluwer, 1996.
- [16] E. Biersack, W. Geyer, and C. Bernhardt, "Intra- and inter-stream synchronization for stored multimedia streams," in *Proc. IEEE Int. Conf. Multimedia Computing and Systems*, Hiroshima, Japan, June 17–23, 1996, pp. 372–381.
- [17] M. M. Buddhikot and G. M. Parulkar, "Efficient data layout, scheduling and playout control in MARS," in *Proc. NOSSDAV'95*, 1995, pp. 318–329.
- [18] Y. B. Lee and P. C. Wong, "A server array approach for video-on-demand service on local area networks," in *Proc. IEEE INFOCOM'96*, San Francisco, CA, Mar. 1996, pp. 27–34.
- [19] P. C. Wong and Y. B. Lee, "Redundant array of inexpensive servers (RAIS) for on-demand multimedia services," in *Proc. ICC 97*, Montreal, Canada, June 8–12, 1997, pp. 787–792.
- [20] A. Reddy, "Scheduling and data distribution in a multiprocessor video server," in *Proc. 2nd IEEE Int. Conf. Multimedia Computing and Systems*, 1995, pp. 256–263.
- [21] R. Tewari, R. Mukherjee, and D. M. Dias, "Real-time issues for clustered multimedia servers," IBM Research Rep. RC20020, June 1995.
- [22] M. Wu and W. Shu, "Scheduling for large-scale parallel video servers," in *Proc. 6th Symp. Frontiers of Massively Parallel Computation*, Oct. 1996, pp. 126–133.

Jack Y. B. Lee received the B.Eng. and Ph.D. degrees from the Department of Information Engineering, Chinese University of Hong Kong (CUHK), in 1993 and 1997, respectively.

He is an Assistant Professor at the Hong Kong University of Science and Technology (HKUST), Hong Kong. Before joining HKUST in 1998, he was a Visiting Assistant Professor at CUHK for one year. His research interests include distributed multimedia systems, fault-tolerant systems, and Internet computing.

for correlation. When the distance from the current frame to the last reference exceeds a threshold, then a full-frame P1 reference is set. Default P1 references are inserted in order to prohibit unusually long distances between reference frames. In a fixed bit-rate (FBR) framework (also called constant bit rate in some literature), the authors found it necessary to limit the number of references within an MPEG group of pictures (GOP). A large number of references was problematic because of the reduced bit allocation for references, which led to reduced quality. Variable-bit-rate (VBR) encoding was discussed as a way to achieve constant picture quality as opposed to constant bit rate.

In the continuation of their work [3], [4], the authors developed additional distance metrics to measure the change in video content. The new measures include 1) histogram of difference (HOD) image, 2) block histogram difference, and 3) block variance difference. For example, in method 1 (HOD), the distance measure is defined as

$$D(f_n, f_m) = \frac{\sum_{i=-q+1}^{q-1} \text{hod}(i)}{\sum_{i=-q+1}^{q-1} \text{hod}(i)} = \frac{1}{N_{\text{pix}}} \sum_{i \in [-q, q]} \text{hod}(i)$$

where $\text{hod}(i)$ is defined to be the histogram at value i histogram $_i(f_n - f_m)$ of the difference between two images f_n and f_m (note that $i \in [-q+1, q-1]$ assuming q gray levels in each image), q is a threshold to determine the closeness of the difference to zero (32 was chosen for q in Lee and Dickinson's simulations and our simulations), and N_{pix} is the total number of pixels. Based on this measure, the more difference pixels distributed away from the zero in the histogram, the more changes between the two images.

Method 2 is similar to the metric used in their original work, but block computations are made in an effort to capture significant local information. Method 3 is similar to method 2 except for the use of variance instead of histograms. HOD is the chosen method for comparative study in this paper on the basis of its monotonically increasing nature as two frames become farther apart. For a fair comparison in this paper, we only use the full (I, P, B) frame assignment without using the reduced one as in [2], [4]. We only experimented with the VBR scheme since it showed better performance than FBR, which also concurs with the simulations of Lee and Dickinson. More specifically, the signal-to-noise ratio (SNR) for VBR is less fluctuating than FBR, which indicates that the perceptual picture quality using VBR is almost constant.

HOD measure is clearly sensitive to image characteristics, and is invariant to rotation and translation. On the other hand, DOH measure is only sensitive to the amount of global motion and cannot serve as a good indicator for adaptive frame placement. There are no considerations in the design to account for both the magnitude of motion and the presence of small object motion. Images with high detail and contrast can easily trigger the placement of a reference frame even though the magnitude of motion is small. This would be unnecessary because motion estimation and compensation (ME/MC) is typically effective over a larger range of object motion.

Sensitivity to global motion is definitely an asset of this system since global motion should dictate how reference frames are placed. However, small object motion still requires attention for coding efficiency, but there are no provisions in this system for assisting motion estimation by local motion tracking. DOH can fail to account for decreases in correlation when objects in consecutive frames undergo significant translation and/or simple rotation, with the axis of rotation being perpendicular to the image plane. For example, it is possible for two video frames to have the same histogram but poor motion-estimation performance due to large amounts of translation. This is because motion estimation is typically effective in only a local region.

An adaptive-size GOP is examined in [5]. The extreme case of allowing only one independent "I" reference frame for every GOP was considered. This type of reference is situated as the first frame of every new scene in order to prevent predictions from being made between two different scenes. This encoding scheme was used for the purpose of indexing video according to scene changes. *Random-access indexing* facilitates video editing and browsing by allowing a user to search through scenes rather than frame by frame for a particular segment. Fast and robust scene-change detection was performed by using MPEG bitstream encoded parameters. The correlation measures developed were based on the 1) prediction error power of "P" pictures, 2) average number of macroblocks using forward or backward prediction, and 3) motion-vector inner products. Frames not classified as a scene change were encoded in a fixed, periodic structure of picture types to reduce bit-rate variations for asynchronous transfer mode (ATM) network scheduling. This encoding technique only adapts to abrupt changes in content and ignores the changes occurring within a scene. An important conclusion from this work was the fact that infrequent "I" reference frames did not cause severe degradation in performance due to error propagation.

B. Scene-Content Adaptation Using Motion Analysis

This paper proposes the use of motion analysis (MA) to adapt to scene content. Motion analysis is a generic term for the analysis of temporal variations encountered in video. This technique is based on the intuitive idea that motion activity is the primary reason why temporal prediction error exists. Indeed, if nothing moved in a sequence, every frame would be the same, barring fade-in and dissolve movie effects. However, objects or background in motion cause adjacent video frames to lose correlation due to natural object transformation, nonuniform illumination, uncovered scenery, video deinterlacing effects, shadowing, etc.

The motion-analysis step computes the motion information between every set of adjacent frames. The simple full-search block-matching algorithm (BMA) is used because it can generate useful statistics for scene-change detection. To take care of scene changes, the thresholding mean square prediction error (MSPE) method, which is similar to that proposed in [4], is adopted. This method measures how well a frame can be predicted from its adjacent neighbor. A scene change will

make the error value very large for easy thresholding.

Significant changes within a scene that require reference-frame placement can be tracked using measures for the amount and the magnitude of motion. The motion vectors yielded by the MA step are thresholded so that at least a certain percentage of them are above the threshold. The image blocks corresponding to motion vectors above the threshold are the motion areas for the current frame. The magnitude of the vectors from motion areas are averaged for the final motion measurement. The actual picture-type decision is made by examining the accumulation of motion measurements since the last reference frame was labeled. The detection of significant changes within a scene in our algorithm differs from [3] in that we actually measure and accumulate the *amount* and *magnitude* of global/local motion between adjacent frames, which cannot be achieved by the HOD analysis [4].

C. Organization of this Paper

Section II focuses in depth on the algorithm and criteria for the proposed adaptive frame-type labeling. Section III describes the VBR simulations and discusses the video data set, performance measures, results on bit savings, and limitations. The last section concludes the paper with a summary of the encoder's key elements and a few words about future extensions of the method.

II. SCENE ADAPTATION USING MOTION ANALYSIS

Motion analysis is the proposed method for adapting to temporal variations found in video. MA is applied to every set of adjacent frames to yield the motion information for the entire sequence. This information is sufficient for the meaningful arrangement of picture types in the context of MPEG video coding. An algorithm based on motion analysis allows an intuitive coding model for video: high motion is addressed by inserting frequent reference frames in order to facilitate the ME/MC process, while low motion is coded with less frequent references since the similarity is high among consecutive frames. The mechanics of the scene-adaptation algorithm developed in this paper are simple. The process begins by performing MA on the current frame. Statistics are gained from this step that can lead to easy scene-change detection, in addition to the flow (motion) vectors from the current frame to its neighbor. A picture-type decision is made by first examining the scene-change statistics for abnormality. Scene changes are handled using I-references. If the current frame is not the first frame of a new scene, the process continues and a motion measurement is taken on the flow vectors that is sensitive to the amount of motion. The accumulated motion measurements since the last reference frame are used as the criterion for labeling the current frame a P-reference. If all conditions fail, the current frame is labeled a B-picture.

A. Design Considerations

This frame-type selection algorithm was developed to take into account several different guidelines. The underlying idea,

however, is still the same: the continuity, amount, and magnitude of motion should dictate how frames are coded. The design considerations are as follows.

- 1) A scene change is modeled as a total change in content. An I-frame should be used.
- 2) Global motion sensitivity is the most important design specification since this type of motion affects the whole image. This includes zooms, pans, and rotations. Segments of video with significant global motion should yield frequent placement of P-reference frames.
- 3) Local motion should also be considered when a "significant" portion of the image is affected. Otherwise, local motion has only limited significance in the context of the entire image. This includes all object motion. Segments of video with little global motion should yield large spacing between consecutive P-reference frames. Local motion requires tracking in order to facilitate ME/MC.

The algorithm is sparing in its use of I- and P-reference frames since they typically require many more bits than B-pictures. Motion activity in video is used to dynamically place these frames to ensure a minimum amount of quality degradation while achieving the bit-rate reduction. The accumulation of motion magnitudes can be used as a good indicator of performance degradation.

B. Context-Dependent Reference-Frame Placement Algorithm

Fig. 1 shows the frame-type selection algorithm embedded in the MPEG encoding scheme. The MPEG elements that are affected by the algorithm are shown as darkened boxes. Fig. 2 is a flowchart of the algorithm. The algorithm operates on the original video frames by estimating motion and then processing the information to determine the picture type. If a reference frame is labeled, the motion information since the last reference is sent to an interpolation procedure for tracking the local motion from the current frame to its reference(s). A suitable quantization parameter for B-pictures is selected based on the interreference-frame distance for the current video segment. A constant quantization parameter is used for I- and P-frames.

1) *Motion Analysis*: In this technique, motion information is obtained from every set of adjacent frames. The BMA is used to estimate motion using a small search range, since movement between consecutive frames is minimal. The MPEG ME/MC step requires both a past and a future prediction for B-pictures, so motion vectors are obtained from the current frame to the immediate past frame and also to the immediate future frame. This is described pictorially in Fig. 3.

Scene-change statistics are also calculated in this step. The MSPE method is adopted [4] as an indication of how well a frame can be predicted from the previous frame. The prediction error will increase dramatically in the event of a total change in content. The MSPE is defined as

$$\text{MSPE} = \frac{1}{N \cdot M} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} [O(i, j) - P(i, j)]^2 \quad (1)$$

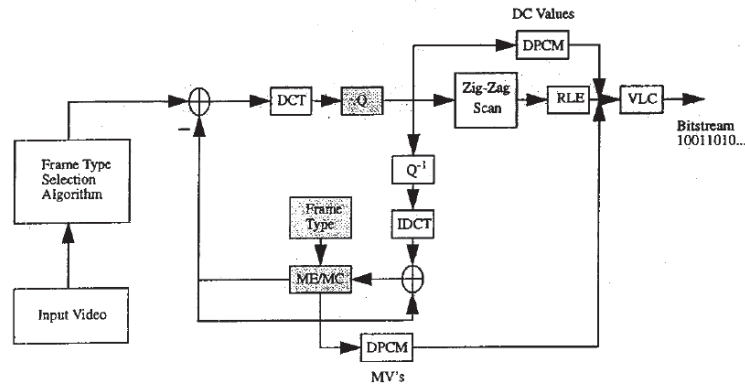


Fig. 1. The frame-type selection algorithm embedded in the MPEG encoding scheme.

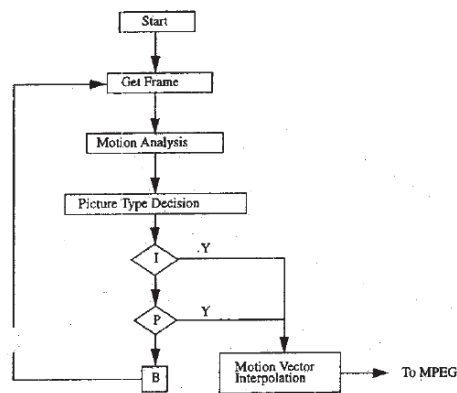


Fig. 2. The flowchart associated with frame-type selection.

where O and P represent the original and predicted (compensated) frames, respectively. N and M are the frame dimensions.

2) **Picture-Type Decision:** The scene-change statistics gathered in the MA stage are thresholded to determine if a change exists. If scene change is detected, the current frame is the first frame of a new scene. This situation is handled by coding the current frame an I-picture and the previous frame a P-picture to reset the prediction dependencies. Coding the previous frame a B-picture would allow the old scene to reference the new scene, which should be forbidden since the content is completely different.

Significant changes within a scene that require placement of a P-reference frame are detected by using the forward motion vectors from the MA stage. The first step involves locating the motion areas of the image. The motion areas are the macroblocks that have a motion-vector magnitude greater

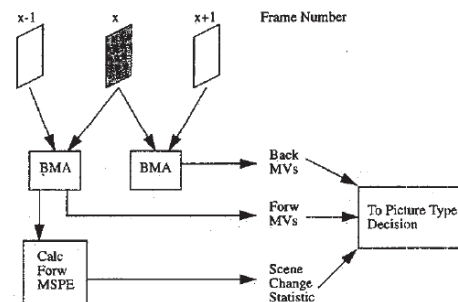


Fig. 3. The motion vectors are obtained from the current frame to the immediate past frame and also to the immediate future frame.

than a threshold. The threshold is adaptively determined for every video frame according to the number of detected motion areas. The number of motion areas must be greater than a user-defined activity threshold for the algorithm to continue. The motion measurement for the frame is calculated as the average motion-vector magnitude from all the motion areas. The decision to code the frame as P or B is based on the accumulated motion measurements since the P- or I-frame. If the accumulated motion exceeds a maximum value, then the current frame is labeled a P-reference.

Fig. 4 demonstrates these principles in flowchart form with MSPE as the chosen scene change statistic. $|MV|$ represents the magnitude of a motion vector. The major algorithm variables are scene-change threshold, accumulated motion, motion threshold, and activity threshold. The scene-change threshold can be easily set to a value suitable for most abrupt changes in content. This is a highly desirable characteristic since the ultimate objective is to have an optimized, autonomous encoder. The accumulated motion is the sum of all motion measurements since the last reference frame. This value is

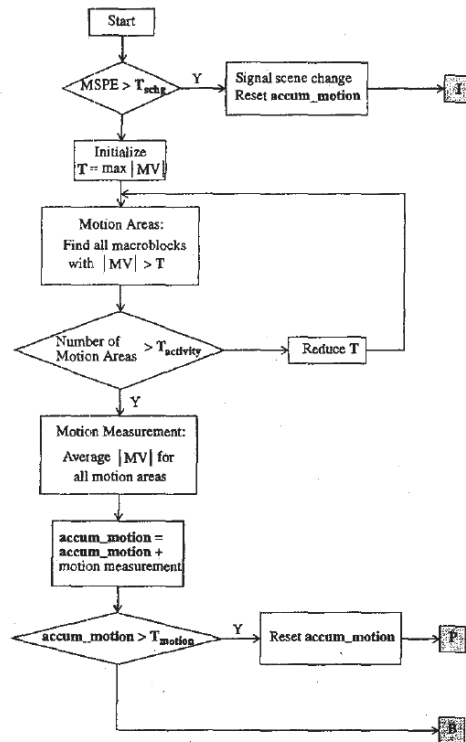


Fig. 4. The principles in flowchart form with MSPE as the chosen scene-change statistic.

compared with a user-defined maximum magnitude of motion, referred to as the motion threshold. The activity threshold affects the sensitivity of the motion measurement to the amount of motion. A high value means that only global motion is measured since the algorithm will loop until the motion areas will cover most of the image. A low value, in general, allots more significance to prominent local motion. The ideal value for full-motion video will place emphasis on global activity and diminish small regions of movement regardless of the magnitude.

3) *Motion-Vector Interpolation*: In addition to picture-type labeling, the motion information that is obtained for every set of adjacent frames is useful for the tracking of object and background motion. This will assist the MPEG ME/MC stage by placing the search center for a macroblock in a reasonable location. Furthermore, the number of calculations performed by the ME routine can be reduced since a large search range is unnecessary when motion tracking is performed.

Normal MPEG motion estimation defines the search center to be the original macroblock spatial coordinates and a search

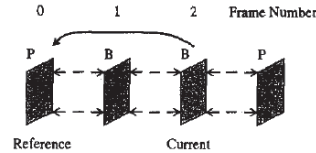


Fig. 5. An example of motion-vector interpolation.

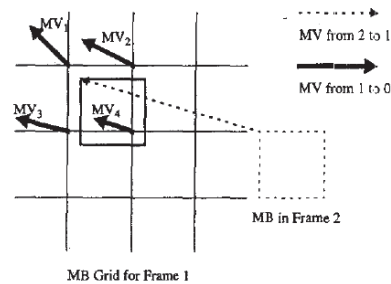


Fig. 6. Interpolation of one macroblock.

range surrounding this location. This implies that the search range must be increased as the current frame is moved farther away from its reference. The typical strategy is to multiply each search-range dimension for adjacent frames by the frame distance. This motion-tracking scheme does not suffer from the same increase in complexity, but it requires repeated interpolations which, may incur some error. Interpolation is required because only the motion between adjacent frames is known. MPEG motion estimation requires the path of motion from the current frame to its designated reference, which could be many frames away. An example is shown in Figs. 5 and 6.

The problem of tracing the motion from frame 2 in Fig. 5 to frame 0 is solved by using repeated applications of bilinear interpolation. Using the notation defined in Fig. 6, bilinear interpolation is defined as

$$MV_{int} = \alpha\beta MV_1 + \alpha(1-\beta)MV_2 + (1-\alpha)\beta MV_3 + (1-\alpha)(1-\beta)MV_4 \quad (2)$$

where the fraction numbers α and β are determined by the pixel distance to MV_1 . In Fig. 6, a macroblock in frame 2 is portrayed using a dashed square. Its projection into frame 1 is shown as the solid square pointed to by the dashed motion vector. Typically, the projection will not coincide with the nonoverlapping grid of macroblocks. Therefore, bilinear interpolation is used to determine the next motion vector that will generate a projection in frame 0. Tracing the motion in this manner will create an extended motion vector for each macroblock in the current frame to its reference. This extended motion vector is used as the search center for the final MPEG motion-estimation stage. The same procedure is used to determine both forward and backward interpolated motion vectors.

4) *Adaptive Quantization*: The quantization of B-pictures is adaptively changed according the interference-frame distance. Long interference-frame distances cause a double coding penalty for B-pictures: the number of bits increases while the SNR decreases. This is due to the degraded prediction performance and reduction in correlation due to increased distance from the reference frames. For these reasons, an alternate quantization parameter Q_{B_n} for B-pictures is used to keep the SNR at adequate levels when reference-frame distances extend farther than a certain length. Otherwise, the normal quantization parameter Q_{B_n} is used. The cost of upholding the SNR in a video segment is an increase in the number of bits for B-pictures. However, a bit-rate reduction can still be attained at a minimal reduction of SNR, as shown in the simulations.

III. SIMULATION RESULTS

This section describes the design and results of the simulations used to demonstrate the effectiveness of the proposed MA-based adaptive MPEG encoder. The simulations in this paper were based on the MPEG-1 standard [6], [7]. The simulation results of a HOD-based adaptive method in placing the reference frames will also be compared with our adaptive method. The principle objective is to compress raw video more than a fixed GOP arrangement while maintaining quality, as measured by SNR and human perception. SNR is defined as the signal variance divided by the variance in the quantization noise. The standard fixed arrangement of picture types is defined to be *IBBPBBPBBP...*. In areas of high motion, the proposed adaptive encoder is defaulted to the fixed arrangement to prevent an unnecessary increase in bit rate. VBR coding is used since the goal is compression. A different constant quantization parameter is defined for each picture type, i.e., Q_I , Q_P , and Q_{B_n} . B-pictures also require alternate value Q_{B_n} for video segments that have a large interference-frame distance.

A. Performance Measures

Since the frame-selection algorithm operates only based upon the luminance (Y) data, SNR is taken from the luminance component for comparison purposes. When comparing the performance of the scene-adaptive encoder to the fixed arrangement of picture types, three measures are scrutinized.

- 1) Total bits spent on coding all video frames (luminance and chroma).
- 2) Average P- and B-picture SNR (dB).
- 3) Average local B-to-reference SNR ratio, which is defined to be

$$\frac{B_{SNR}}{(R_1, SNR + R_2, SNR)/2} \quad (3)$$

where B_{SNR} represents either the minimum B-picture SNR between two references or the average SNR. R_1 and R_2 are the past and future references for the B-pictures. These can be either I or P. A local measure is

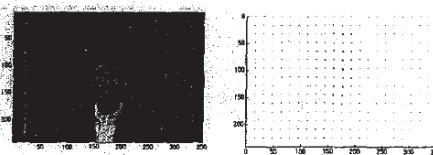


Fig. 7. Claire sequence with motion sample.

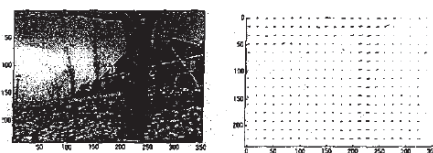


Fig. 8. Garden sequence with motion sample.

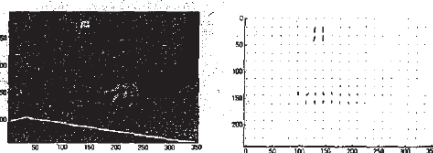


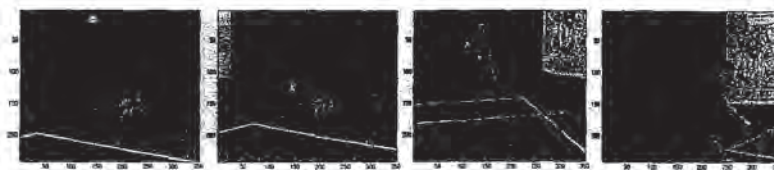
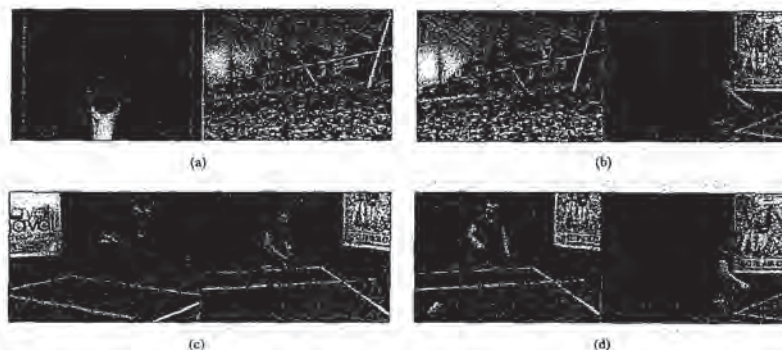
Fig. 9. Tennis sequence with motion sample.

necessary in order to capture the information that is lost in global average measures, such as measure 2).

B. Parameter Adjustment

The motion threshold and alternate quantization parameter for B-pictures must be adjusted to trade off bit savings with SNR. The user-defined maximum magnitude of motion is thresholded against the accumulated motion since the last reference frame, as explained in the algorithm description. Adjusting this parameter yields different interference-frame distances. Long distances mean fewer P-reference frames, and therefore higher compression. As the distance increases, though, the number of bits for the B-pictures in between the references increases while the SNR drops. The alternate quantization parameter can be used to counter the degradation in quality at the cost of increasing the number of bits.

Long distances between references can also cause difficulties due to object occlusion, acquisition noise, and object/background dissimilarities. Therefore, the interference-frame distance should be kept as short as possible, which will facilitate motion-vector interpolation and MPEG motion estimation. The alternate quantization parameter Q_{B_n} should be kept as close to the normal value Q_{B_n} as possible to prevent overcompensation in video segments with small interference-frame distances. This situation can sometimes cause the number of bits to increase dramatically, with the SNR increasing to unnecessarily high levels.

Fig. 10. Four segments of the *Tennis* sequence.Fig. 11. The scene-change frames in the created test sequence and *Tennis* sequence: (a) test sequence, frames 4 and 5; (b) test sequence, frames 13 and 14; (c) *Tennis* sequence, frames 67 and 68; and (d) *Tennis* sequence, frames 97 and 98.

C. Test-Video Description

The video used to test the algorithm comes from the public domain. The three sequences are *Claire*, *Garden*, and *Tennis*. Figs. 7–9 show a sample of each. *Claire* and *Garden* represent video from the two extreme ends of motion. *Claire* is a news-cast, and so it has no global motion, a still background, and low foreground motion. *Garden*, on the other hand, has global panning, moderate object motion, object occlusion/exposure, and high detail. These figures also contain a sample of the motion in the sequences. These motion vectors represent the output from the MA stage. Since *Claire* will require fewer reference frames than *Garden*, it is expected that the bit savings will be more significant.

Tennis is a table-tennis sequence composed of four segments with differing characteristics. A sample of each is shown in Fig. 10. The first segment has moderate local motion, stationary background, and high detail. The second segment has global zooming out, moderate-to-major local motion, and diminishing resolution due to the zooming. Intuitively, there is less correlation to code due to the action of zooming out. We thus have to use many reference frames, which results in an increase in SNR with very little bit-rate saving. The third segment has a scene change, moderate local motion, stationary background, and low detail. The fourth segment also has a scene change, but it has little local motion, stationary background, and high detail. The first, third, and fourth segments will achieve better compression than the second segment since they lack global motion.

D. Presentation and Discussion of Results

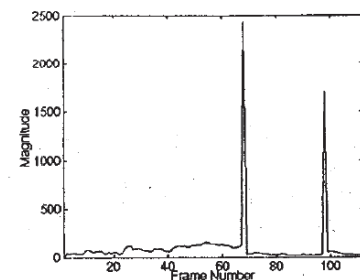
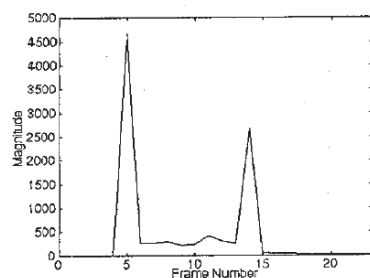
1) *Scene-Change Results*: To test the performance of both adaptive encoders (MA and HOD based) under the condition of a total change in content, a test sequence was created using short segments of *Claire*, *Garden*, and *Tennis*. The full *Tennis* sequence was also used as a test since it has two scene changes. Fig. 11 displays the frames before and after the two scene changes in each of the experiment sequences.

The MSPE measure for scene-change detection, as plotted in Figs. 12 and 13, successfully identifies the two scene changes in each of the experiment sequences by showing strong peaks. From these sequences, it would be easy to determine a fixed threshold for autonomous scene-change detection. The HOD measurement also successfully identifies the two scene changes in each of the experiment sequences by showing strong peaks.

2) *Scene-Adaptive Encoder Results*: Our proposed MA-based adaptive encoders were compared to the HOD-based adaptive encoder and fixed reference-frame placement encoder based on SNR and the savings on the total number of bits. Only one I-picture was used (first frame) for each scene unless otherwise specified. In the *Claire* sequence, the motion threshold was set to 5.0, and the activity threshold was set to four macroblocks. The activity threshold was set low in order to capture the local activity of the head and facial features. Table I shows the numerical results. The total bit-rate savings for our MA-based adaptive method is 13.9% and for the HOD method is 8.54% over these 46 frames. The measure that shows

TABLE I
CLIQUE SEQUENCE RESULTS FOR FIXED REFERENCE-FRAME, HOD-BASED ADAPTIVE ENCODERS AND OUR PROPOSED MA ADAPTIVE ENCODERS

	Quantization	Frames	Ave. P	Ave. B	Ave. Local	Ave. Local	Ave. P	Ave. B	Total
	Q_T, Q_P, Q_{B1}, Q_{B2}	No.	SNR (dB)	SNR (dB)	D/F (min)	D/F (ave)	Size (bits)	Size (bits)	Size (bits)
Pixed	2,3,8,7	46	34.09	32.39	0.944	0.946	68,511	7,145	1,369,920
ROD	2,3,8,7	46	34.26	32.09	0.939	0.942	68,640	8,860	1,282,936
MA	2,3,8,7	46	34.27	32.24	0.938	0.940	70,177	10,628	1,179,860



the most degradation in the MA method and the HOD method is the minimum B-SNR to reference SNR ratio. Nevertheless, these drops are very insignificant to human perception when considering the viewing distance and expected frame rate. Fig. 14(a) displays the fixed and MA-based adaptive frame-type arrangements, along with the motion measurements that are created using the MA-based algorithm. The number of P-pictures is reduced significantly for *Claire* when our proposed MA-based adaptive encoding is used. Fig. 14(b) displays the HOD-based frame-type arrangements, along with the HOD measurements. The number of P-pictures is also reduced significantly.

The *Garden* sequence was tested using 20.0 for the motion threshold and 16 macroblocks for the activity threshold. Table II and Fig. 15 hold the results for this sequence. The bit-rate savings is only 2.0% for an MA-based adaptive encoder and 0.2% for an HOD-based adaptive encoder due to the

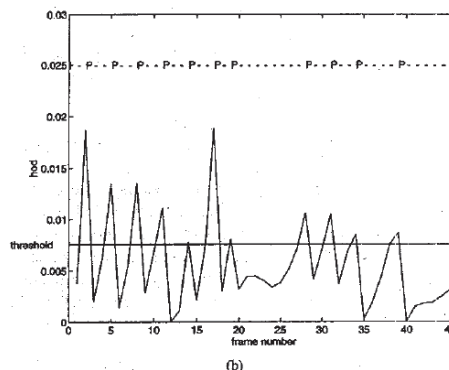
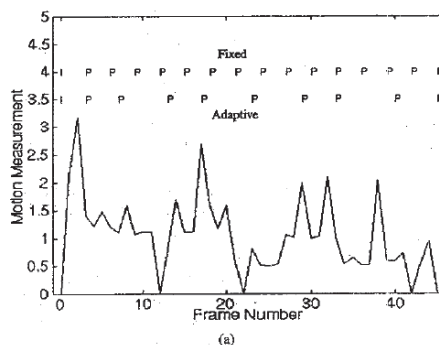


Fig. 14. (a) The fixed and adaptive frame-type arrangements for *Claire* sequence, along with the motion measurements that are made using the MA-based algorithm. (b) HOD of 46 frames from *Claire* sequence.

significant global motion involved. The high image detail also works against the adaptive system. The details contribute to high variance in the prediction residuals, and therefore a higher yield of bits. The number of P-pictures cannot be reduced as much as in *Claire*. The frame bit rates increase more dramatically than in *Claire* as the reference frames are spread farther apart. This can be attributed to the global motion, which affects the entire image rather than a small local region.

In the *Tennis* sequence, the motion threshold was set to 1.4, while the activity threshold was set to 100 macroblocks.

TABLE II
GARDEN SEQUENCE RESULTS FOR REFERENCE-FRAME, HOD-BASED ADAPTIVE ENCODERS AND OUR PROPOSED MA ADAPTIVE ENCODERS

	Quantisation $Q_I, Q_P, Q_{B_n}, Q_{B_d}$	Frames No.	Ave. P SNR (dB)	Ave. B SNR (dB)	Ave. Local D/P (min)	Ave. Local B/P (ave)	Ave. P Size (bits)	Ave. B Size (bits)	Total Size (bits)
Fixed	18,20,35,33	37	18.05	15.93	0.674	0.581	102,682	18,496	1,865,440
HOD	18,20,35,33	37	17.99	15.87	0.671	0.580	104,476	24,127	1,861,416
MA	18,20,35,33	37	18.06	15.91	0.669	0.579	110,955	23,726	1,818,628

The activity threshold at this level measures global motion. Four I-pictures were used in this sequence at frames 1, 23, 68, and 98 (for both MA-based and HOD-based adaptive encoders). Frame 23 is the beginning frame of the global zooming action. Frames 68 and 98 are the first frames of new scenes. This arrangement of I-pictures temporally decomposes the *Tennis* sequence into its component segments, as discussed in Section III-C. Encoding in this fashion yields the results shown in Table III and Fig. 16. The circled picture types represent how scene changes are coded: I-picture for the first frame of the new scene and P-picture for the last frame of the old scene. The bit-rate savings for the MA encoder is 4.88% over the entire sequence, with an average B-picture SNR gain of about 0.01 dB (while the HOD-based adaptive encoder achieves only 1.7% bit savings with an average B-picture SNR degradation of about 0.74 dB). The overall perceptual quality of both schemes, however, remains similar to the fixed arrangement. Fig. 17 shows the savings achieved by the MA-based encoder calculated for each video segment separately. The second segment cannot achieve further compression because of the zooming out of the camera. This global activity causes prediction performance to become significantly worse as reference frames are spread apart. The encoder correctly defaults to the fixed arrangement of picture types in this segment because the measured motion is high in both magnitude and amount. The final two scene-change segments were able to achieve reasonable compression despite the fact that a constant motion threshold was used for the entire sequence. Under ideal conditions, the parameters would be changed to reflect the nonstationary statistics, especially after significant changes in content.

3) *Results for Default I-Pictures*: The need for default I-pictures to facilitate random access is acknowledged as an important requirement for MPEG video coding. Typically, an encoder will force I-frames into the coded sequence at regular intervals. This allows a VCR-like "random search and video halt" feature, which is implemented by simply decoding the next I-picture. Decoding I-pictures when the video is stopped is the most effective implementation since reference-frame reconstruction delay is avoided. Results are presented here using default I-pictures every 15 frames. In the event of a scene change, both adaptive algorithms function by coding an I-picture as usual and then resetting the frame counter for the default I-decision. This is in contrast to coding I-pictures every 15 frames without deviation and also inserting I-pictures for scene changes. Tables IV-VI show the results for *Claire*, *Garden*, and *Tennis*, respectively.

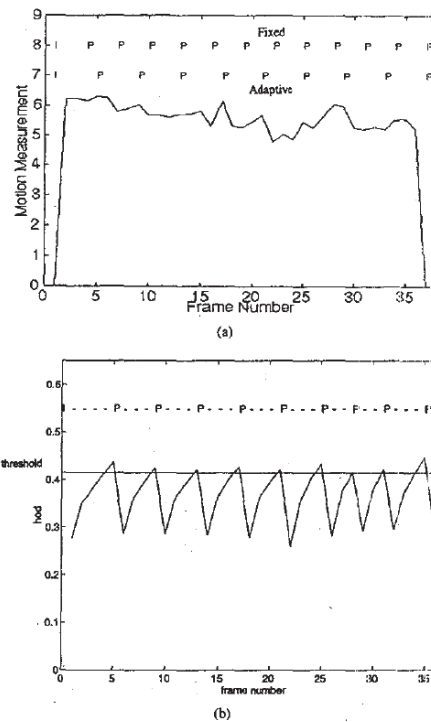
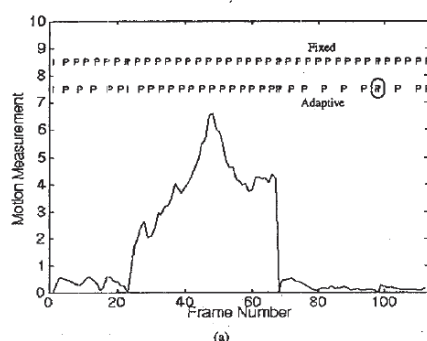


Fig. 15. (a) The fixed and adaptive frame-type arrangements for *Claire* sequence, along with the motion measurements that are made using the MA-based algorithm. (b) HOD of 37 frames from *Garden* sequence.

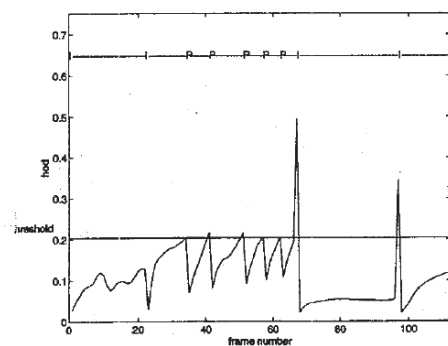
For the *Tennis* sequence, the comparison is made between fixed structure encoding with no consideration for scene changes and adaptive encoding with scene-change detection. The savings on these sequences for our MA-based adaptive encoder are 6.72, 1.44, and 3.66%. The *Claire* sequence has similar SNR performance, but the bit-rate savings is reduced dramatically. *Garden* does not truly save any bits at all. With a 1.44% savings, the B-picture SNR is down 0.3 dB on average. A more acceptable SNR level will make the savings negligible. The savings from the *Tennis* sequence is also reduced as default I-pictures are inserted. However, the adaptive scheme

TABLE III
TENNIS SEQUENCE RESULTS FOR FIXED AND ADAPTIVE FRAME PLACEMENT

	Quantization $Q_I, Q_P, Q_{B_n}, Q_{B_e}$	Frames No.	Ave. P SNR (dB)	Ave. B SNR (dB)	Ave. Local B/P (min)	Ave. Local B/P (ave)	Ave. P Size (bits)	Ave. B Size (bits)	Total Size (bits)
Fixed	7,10,20,16	112	19.82	17.70	0.898	0.902	81,287	14,061	4,828,376
HOD	7,10,20,16	112	20.66	16.86	0.891	0.876	106,236	30,821	4,741,424
MA	7,10,20,16	112	19.60	17.71	0.886	0.893	86,512	16,971	4,687,976



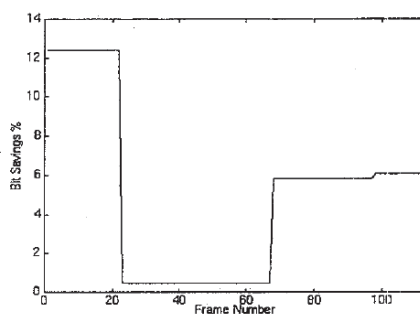
(a)



(b)

Fig. 16. (a) The fixed and adaptive frame-type measurements for *Tennis* sequence, along with the motion measurements that are made using the MA-based algorithm. (b) HOD of 112 frames from *Tennis* sequence.

has better average SNR for P- and B-pictures. The local SNR measures show a small degradation. The fixed arrangement is unable to code the scene changes effectively. Prediction error increases significantly when the scene changes occur. The error propagates to subsequent frames, increasing the number of bits and decreasing the SNR. The savings is reduced in all three test sequences because true scene adaptation is prohibited when using default I-pictures. The entire frame-type arrangement is changed since the accumulated motion is always reset after default I-pictures. When the accumulation of motion is allowed to be cut short, a frame can be defaulted to "I" even

Fig. 17. Bit savings for each segment of the *Tennis* sequence.

though the following frames are still highly correlated. True adaptation would allow the sequence to continue without a reference since the prediction quality is still high. Another effect of inserting default I-frames is a considerable increase in the bit-rate level for the entire sequence. The bit savings is calculated as the difference between the total yield of bits for the fixed and adaptive encoders. When the bit savings is expressed as a percentage of this new bit-rate level, the value is much smaller. *Claire* was affected the most since the bit-rate increase caused by inserting default I-frames was the most dramatic.

It is also interesting to see that the HOD-based adaptive encoder performs better than the fixed arrangement encoder only on the *Tennis* sequence, which contains scene changes that are difficult for a fixed arrangement coder to encode. In summary, the number of total bits is increased 1.3% for the *Claire* sequence and 2.2% for the *Garden* sequence.

The simulation results show that the complexity of our MA-based adaptive encoding is quite similar to that of Lee and Dickinson's HOD method. The CPU time for both methods increases approximately 20% when compared to that of using fixed-reference frame placement with full-search motion estimation.

IV. CONCLUSION

The proposed MA-based method for MPEG video coding has been shown to be an effective scene-content adaptive technique. Further compression over the conventional fixed arrangement of picture types is possible while maintaining similar quality. Our adaptive encoder is also shown to produce

TABLE IV
CLAIRE SEQUENCE RESULTS USING DEFAULT I-PICTURES

	Quantization $Q_I, Q_P, Q_{B_n}, Q_{B_s}$	Frames No.	Ave. F SNR (dB)	Ave. B SNR (dB)	Ave. Local B/P (min)	Ave. Local B/P (ave)	Ave. F Size (bits)	Ave. B Size (bits)	Total Size (bits)
Fixed	2,2,8,7	100	34.36	32.37	0.843	0.944	65,620	7,803	3,803,034
HOD	2,2,8,7	100	34.40	32.38	0.848	0.949	65,089	8,778	3,847,520
MA	2,2,8,7	100	34.55	32.18	0.938	0.980	71,888	10,793	3,268,704

TABLE V
GARDEN SEQUENCE RESULTS USING DEFAULT I-PICTURES

	Quantization $Q_I, Q_P, Q_{B_n}, Q_{B_s}$	Frames No.	Ave. F SNR (dB)	Ave. B SNR (dB)	Ave. Local B/P (min)	Ave. Local B/P (ave)	Ave. F Size (bits)	Ave. B Size (bits)	Total Size (bits)
Fixed	18,20,38,33	100	17.87	18.86	0.877	0.888	108,887	18,983	8,388,380
HOD	18,20,38,33	100	17.87	18.83	0.888	0.874	114,139	23,820	8,818,040
MA	18,20,38,33	100	17.89	18.82	0.884	0.888	118,717	28,807	8,320,748

TABLE VI
TENNIS SEQUENCE RESULTS USING DEFAULT I-PICTURES

	Quantization $Q_I, Q_P, Q_{B_n}, Q_{B_s}$	Frames No.	Ave. F SNR (dB)	Ave. B SNR (dB)	Ave. Local B/P (min)	Ave. Local B/P (ave)	Ave. F Size (bits)	Ave. B Size (bits)	Total Size (bits)
Fixed	7,10,20,18	112	16.89	17.63	0.892	0.898	80,323	14,418	8,237,080
HOD	7,10,20,18	112	20.07	17.39	0.883	0.872	91,661	23,120	8,307,844
MA	7,10,20,18	112	19.88	17.72	0.881	0.894	87,110	18,801	8,038,760

better performance than the HOD-based adaptive encoder on all the test sequences. The algorithm is able to adapt to the temporal variations in video through the use of the block-matching algorithm for motion analysis. The characteristics of motion that are examined are the continuity, amount, and magnitude. Temporal discontinuities in motion, as measured by mean square prediction error, are used to easily detect abrupt scene changes. The level of motion, whether low or high, is a strong indication of how much compression a video segment can sustain by adapting the interreference-frame distances. This system is most sensitive to global motion since the whole frame is affected. Local motion has only limited significance when deciding how to code the entire picture. The magnitude of motion is accumulated to provide the criterion for labeling intrascene reference frames. Motion of objects and background is effectively tracked using bilinear interpolation of the estimated motion from the motion-analysis stage. This serves to facilitate the MPEG motion estimation and compensation step.

A benefit of using MA is the ability to perform local motion tracking. Since motion information is obtained from every set of adjacent frames, one can easily trace the path of motion from the current frame to its reference, which could be many frames away, and thus reduce the search range required during the motion estimation. Tracing the motion vector path over several frames involves repeated interpolations until the stopping frame (either I or P) is

reached. With this special property of efficient motion tracking, even though the proposed MA adaptive encoding introduces additional computational complexity, it is later relieved by the smaller dynamic search range in the motion estimation that accounts for 70% of the overall encoding process. Therefore the CPU time increases only slightly (about 20%) when compared with the fixed reference-frame placement with full-search motion estimation. Furthermore, the complexity of our MA-based adaptive encoding is quite similar to that of Lee and Dickinson's HOD method (from our CPU time estimates). Both methods added longer processing delay and increased the complexity. However, they are very promising approaches for storage media applications such as CD-ROM, where longer encoding delay can be tolerated.

Quality versus bit-rate savings adjustment is the key advantage of this adaptive system over the generic fixed arrangement of picture types. For example, the *Tennis* sequence consists of four scene segments with only one requiring frequent reference frames to elevate the quality. The adaptive encoder is able to take advantage of the scenes where long interreference-frame distances are allowed for increased compression. When the scene with global zooming action is encountered, the adaptive encoder adjusts by reducing the reference-frame intervals. This adaptive encoding scheme is able to satisfy an intuitive coding model: low motion means that fewer reference frames are needed since frame correlation extends over a large span of consecutive images; high motion requires

the frequent placement of references to uphold the quality. Simulations on *Claire* and *Garden*, sequences from the two extreme ends of motion, demonstrate this principle. *Claire*, a newscast sequence with little motion, saves up to 13.9%, while *Garden*, a sequence with considerable global camera panning, saves only 2.0%. These results were obtained using a variable-bit-rate scheme, that the gain is minimal for default I-frame placement.

ACKNOWLEDGMENT

The authors wish to thank all three anonymous reviewers of this paper for their valuable and constructive suggestions, from which the revision of this paper benefited significantly.

REFERENCES

- [1] A. Y. Lan, "Scene context dependent reference frame placement for MPEG video coding," M.S. thesis, University of Washington, Seattle, Mar. 1996.
- [2] J. Lee and B. W. Dickinson, "Scene-adaptive motion interpolation structures based on temporal masking in human visual perception," in *Proc. SPIE Conf. Visual Communication and Image Processing*, Cambridge, Nov. 1993, pp. 499-510.
- [3] —, "Adaptive frame type selection for low bit-rate video coding," in *Proc. SPIE Conf. Visual Communication and Image Processing*, Chicago, IL, Sept. 1994, pp. 1411-1422.
- [4] —, "Temporally adaptive motion interpolation exploiting temporal masking in visual perception," *IEEE Trans. Image Processing*, vol. 3, pp. 513-526, Sept. 1994.
- [5] H. C. Liu and G. Zick, "Automatic determination of scene changes in MPEG compressed video," in *Proc. IEEE Symp. Circuits and Systems*, Seattle, WA, 1995, vol. 1, pp. 764-767.
- [6] Committee Draft of the Standard: ISO 11172-2, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbs," Nov. 1991.
- [7] ISO/MPEG, "Test Model 5," Doc. MPEG93/457, Apr. 1993.



Austin Y. Lan received the B.S. degree from the University of Kansas, Lawrence, in 1993 and the M.S. degree from the University of Washington, Seattle, in 1996, both in electrical engineering. Since 1996, he has been a Development Engineer with Eastman Kodak Co., Rochester, NY. His current research interests include pattern recognition, wavelet-based image compression, and low bit-rate video and audio coding for videophone applications.



Anthony G. Nguyen received the B.S. and M.S. degrees in electrical engineering from the University of Washington, Seattle, in 1985 and 1993, respectively, where he currently is pursuing the Ph.D. degree in the Department of Electrical Engineering.

Since 1985, he has been with the Boeing Co., Seattle, in the Avionics Software Engineering Department, where he is the Lead Engineer in the Environmental Control System Software Engineering Group. His current research interests include video and image processing, analysis, and compression.



Jenq-Neng Hwang (S'82-M'84-SM'96) received the B.S. and M.S. degrees in electrical engineering from National Taiwan University, Taipei, Taiwan, R.O.C., in 1981 and 1983, respectively, and the Ph.D. degree from the Department of Electrical Engineering, University of Southern California, Los Angeles, in 1988.

After two years of obligatory military service, he became a Research Assistant with the Signal and Image Processing Institute, University of Southern California, in 1985. He was a Visiting Student at Princeton University, Princeton, NJ, from 1987 to 1989. In the summer of 1989, he joined the Department of Electrical Engineering, University of Washington, Seattle, where he has been an Associate Professor since 1994. He has published more than 100 journal articles, conference papers, and book chapters in the areas of multimedia signal processing, statistical data analysis, computational neural networks, parallel algorithm design, and VLSI array architecture. He is a member of the editorial board of the *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*.

Dr. Hwang was a corecipient (with S.-R. Lay and A. Lippman) of the 1993 IEEE Signal Processing Society's Annual Best Paper Award in the area of neural networks for signal processing. He was the Secretary of the Neural Systems and Applications Committee of the IEEE Circuits and Systems Society from 1989 to 1991 and was a member of the Design and Implementation of Signal Processing Systems Technical Committee in the IEEE Signal Processing Society. He is a Founding Member of the Multimedia Signal Processing Technical Committee of the IEEE Signal Processing Society. Currently, he is Chairman of the Neural Networks Signal Processing Technical Committee in the IEEE Signal Processing Society and is the society's representative to the IEEE Neural Network Council. He was an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1992 to 1994 and currently is an Associate Editor for IEEE TRANSACTIONS ON NEURAL NETWORKS and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He was the Conference Program Chair of the 1994 IEEE Workshop on Neural Networks for Signal Processing held in Ermioni, Greece, in September 1994. He was the General Cochair of the International Symposium on Artificial Neural Networks held in Hsinchu, Taiwan, R.O.C., in December 1995. He also chaired the Tutorial Committee for the IEEE International Conference on Neural Networks (ICNN'96) held in Washington, DC, in June 1996. He was the Program Cochair of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP) in Seattle 1998.