UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

Unified Patents, LLC,
Petitioner

v.

DigiMedia Tech, LLC
Patent Owner.

Case No. IPR2022-01182
Patent 6,473,532

DECLARATION OF DR. LINA J. KARAM

Table of Contents

I, Lina J. Karam, declare as follows:

I. INTRODUCTION

1. I have been asked to submit this declaration on behalf of Unified Patents Inc. ("Petitioner") in connection with a petition for inter partes review of U.S. Patent No. 6,473,532 ("the '532 patent," Ex. 1001). Specifically, I have been retained as an independent expert consultant by Petitioner to provide my opinions on the technology claimed in, and the patentability or unpatentability, of claims 6 and 16 of the '532 patent ("the challenged claims").

II. QUALIFICATIONS

2. I received a Bachelor of Engineering from the American University of Beirut in Computer and Communications Engineering in 1989, and a M.S. and a Ph.D. from the Georgia Institute of Technology in Electrical Engineering in 1992 and 1995, respectively. A copy of my curriculum vitae, which includes a more detailed summary of my background, experience, qualifications, patents, and publications, may be found at Exhibit 1003.

3. I am currently the Dean of the School of Engineering and a Full Professor in the Department at Electrical & Computer Engineering at the Lebanese American University. I am also an Emerita Professor at Arizona State University (ASU), where I direct the R&D Image, Video, and Usability (IVU) Laboratory at ASU. Before I joined the Lebanese American University, I was a tenured Full

Professor (2010-2019) at ASU, and I have been teaching Electrical Engineering classes since 1995. While at ASU, I also and served as the Computer Engineering Program Chair and, later, as the Computer Engineering Director for Industry Engagement (2016-2018). My research includes such topics as Signal, Image, and Video Processing, Compression, and Transmission; Computer Vision; Machine Learning; Perceptual-based Processing; Visual Attention Models; Automated Quality Assessment and Monitoring; Multidimensional Signal Processing; and Digital Filter Design.

4. I am an Institute of Electrical and Electronics Engineers (IEEE) Fellow, the highest grade level in IEEE conferred each year to no more than one-tenth of 1% of all IEEE voting members, for my contributions in the image and video processing, visual media compression and transmission, and digital filtering areas. The IEEE Signal Processing Society selected me to serve as the Editor-In-Chief of the IEEE Journal on Selected Topics in Signal Processing (IEEE JSTSP), from 2019-2021. I have also served several IEEE Boards and chaired multiple IEEE committees. Currently, I am an expert delegate of the ISO/IEC JTC1/SC29 Committee (Coding of audio, picture, multimedia, and hypermedia information) and participating in JPEG/MPEG standardization activities.

5. I am actively involved in the industry. For example, I contributed to the development of image and video processing and compression at AT&T Bell Labs

(Murray Hill), and multi-dimensional data processing and visualization at Schlumberger. I collaborate in the research and development of computer vision, machine learning, image/video processing, compression, and transmission projects with various industry leaders including Intel, Qualcomm, Google, NTT, Motorola, Freescale, General Dynamics, and NASA. I directed the development of scalable visual compression technologies. The developed codecs were commercialized by General Dynamics. More details can be found in Chien, Sadaka, Abousleman, and Karam, "Region-of-Interest-Based Ultra-Low-Bit-Rate Video Coding," SPIE Symposium on Defense & Security, March 2008. I directed the development of perceptual-based visual compression methods and algorithms. The work on JPEG 2000 Encoding with Perceptual Distortion Control enabled the integration of adaptive perceptual-based visual processing and compression in the JPEG 2000 image coding standard and demonstrated improved performance in terms of visual quality and compression while maintaining full compatibility with the JPEG 2000 standard. For this perceptual-based image compression work, I received a Technical Innovation Award from the US National Aeronautics and Space Administration (NASA). I also am a recipient of the National Science Foundation CAREER Award, the Intel Outstanding Researcher Award, the IEEE SPS Best Journal Paper Award, and the IEEE Phoenix Section Outstanding Faculty Award.

6. I am an inventor, and I have been awarded eight patents in the field of

image and video processing, compression, and transmission: U.S. Patent Nos. 6,154,493 ("The Compression of Color Images Based on a 2-Dimensional Discrete Wavelet Transform Yielding a Perceptually Lossless Image"), 6,124,811 ("Realtime Algorithms and Architectures for Coding Images Compressed by DWT-Based Techniques"), 6,717,990 ("Communication System and Method for Multi-Rate, Channel-Optimized Trellis-Coded Quantization"), 7,551,671 ("System and Method for Transmission of Video Signals using Multiple Channels"), 9,082,164 ("Automatic Cell Migration and Proliferation Analysis"), 9,501,710 ("Systems, Methods, and Media for Identifying Object Characteristics Based on Fixation Points"), 9,704,232 ("Stereo Vision Measurement System and Method"), and 11,0304,85 ("Systems and Methods for Feature Corrections and Regeneration for Robust Sensing, Computer Vision, and Classification"). I have been retained by Unified Patents Inc., and I am submitting this declaration to offer my independent expert opinion concerning certain issues raised in the Petition for inter partes review ("Petition"). My compensation is not based on the substance of the opinions rendered here.

III. MATERIALS CONSIDERED

7. In forming my opinions expressed in this declaration, I have considered, among other things, the following documents. I understand the documents have been given the following exhibit numbers in this proceeding:

Exhibit	Description
Ex. 1001	U.S. Patent No. 6,473,532 to Sheraizin et al. ("the '532 patent")
Ex. 1004	Prosecution History of U.S. Patent No. 6,473,532
Ex. 1005	U.S. Patent No. 6,366,705 to Chiu et al. ("*Chiu*")
Ex. 1006	U.S. Patent No. 7,006,568 to Gu et al. ("*Gu*")
Ex. 1007	US Provisional application No. 60/136,633 ("*Gu*'s Provisional")
Ex. 1008	Arun N. Netravali, Barry G. Haskell, DIGITAL PICTURES: REPRESENTATION, COMPRESSION, AND STANDARDS, Second Edition, 1994 ("*Netravali*")
Ex. 1017	Borko Furht, Raymond Westwater, *Video Presentation and Compression*, *in* HANDBOOK OF MULTIMEDIA COMPUTING (Borko Furht ed., 1998) ("*Furht*")
Ex. 1018	Nikil Jayant, et al., *Signal Compression Based on Models of Human Perception*, 81 IEEE 1385 (1993) ("*Jayant*")
Ex. 1019	Gene K. Wu, Todd R. Reed, *Image Sequence Processing Using Spatiotemporal Segmentation*, 9 IEEE 798 (1999) ("*Wu*")

In forming my opinions, I have also relied on my knowledge and experience.

IV. RELEVANT LEGAL STANDARDS

8. I have been informed by counsel for Petitioner that the following legal principles apply to analysis of patentability based on 35 U.S.C. § 103 for obviousness. I have also been informed that, in an *inter partes* review proceeding such as this proceeding, a patent claim is unpatentable if it is shown by a preponderance of the evidence that the claim would have been anticipated by a prior

art patent or publication, or obvious by one or more properly-combined prior art patents or publications.

A. Obviousness

9. I have been told that under Pre-AIA 35 U.S.C. § 103(a), "[a] patent may not be obtained . . . if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains."

10. When considering the issues of obviousness, I have been told that I am to do the following:

> (a) determine the scope and content of the prior art;
> (b) ascertain the differences between the prior art and the claims at issue;
> (c) resolve the level of ordinary skill in the pertinent art; and
> (d) consider evidence of secondary indicia of non-obviousness such as commercial success, long-felt but unresolved needs, failure of others, etc.

11. I have been told that the relevant time for considering whether a claim would have been obvious to a person of ordinary skill in the art ("POSA") is the time of the alleged invention. I have been asked to assume a priority date for the challenged claims of January 23, 2000.

12. I have been told that a reference may be modified or combined with

other references or with the POSA's own knowledge if the person would have found the modification or combination obvious. A POSA is presumed to know all of the relevant prior art, and the obviousness analysis may consider the inferences that a POSA would employ.

13. In determining whether a prior-art reference could have been combined with another prior-art reference or other information known to a person having ordinary skill in the art, I have been told that the following principles may be considered:

> (a) a combination of familiar elements according to known methods is likely to be obvious if it yields predictable results;
> (b) the substitution of one known element for another is likely to be obvious if it yields predictable results;
> (c) the use of a known technique to improve similar items or methods in the same way is likely to be obvious if it yields predictable results;
> (d) the application of a known technique to a prior art reference that is ready for improvement is likely to be obvious if it yields predictable results;
> (e) any need or problem known in the field and addressed by the prior-art reference can provide a reason for combining the elements in the manner claimed;
> (f) a person of ordinary skill in the art often will be able to fit the teachings of multiple references together like a puzzle; and

(g) the proper analysis of obviousness requires a determination of whether a person of ordinary skill in the art would have a "reasonable expectation of success"—not "absolute predictability" of success—in achieving the claimed invention by combining prior art references.

14. I have been told that whether a prior art reference renders a patent claim unpatentable as obvious is determined from the perspective of a POSA. Further, I have been told that while there is no requirement that the prior art contain an expressed suggestion to combine known elements to achieve the claimed invention, a suggestion to combine known elements to achieve the claimed invention may come from the prior art as a whole or individually, as filtered through the knowledge of one skilled in the art. I have also been told that the inferences a person of ordinary skill in the art would employ are also relevant to the determination of obviousness.

15. I have been told that when prior art is available in one field, design alternatives and other market forces can prompt variations of it, either in the same field or in another. If a POSA can implement a predictable variation and would see the benefit of doing so, that variation is likely to be obvious.

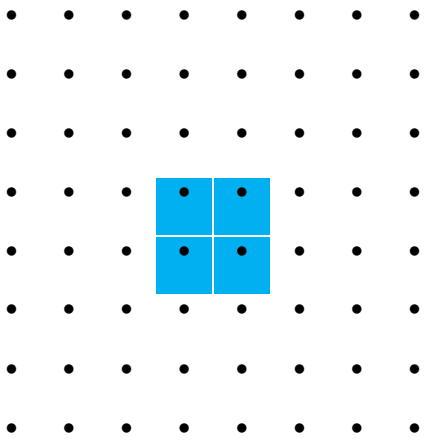
16. I have been told that there is no rigid rule that a prior-art reference or combination of prior-art references must contain a "teaching, suggestion, or motivation" to combine references.

V. TECHNOLOGY BACKGROUND

A. Video Representation Overview

17. Videos are streams of individual pictures or frames that move so quickly that the human eye cannot distinguish between each individual picture much like the flipbooks kids make in school. *See generally, Netravali* at 32-38. When a video is captured, the original image consists of small dots called picture elements, or pixels. *See id.* at 1-8; 60-66. Pixels in a computer monitor are essentially miniature lightbulbs. *See id.* at 1-8; 60-66. When pixels are illuminated, they emit light and produce an image on the screen. *See id.* at 39. Imagine, for example, a picture of a blue square—it may look like this:

18. Now imagine that the blue square is displayed on a screen that is made up of 8x8 pixels, as shown below where each black dot indicates one pixel.



19. Whether a pixel is a certain brightness or, for example, a certain color blue, depends on what signal, or code defined by a series of ones and zeros, is used to represent that pixel value(s). *See id.* at 1-8. Pixel values are comprised of bits (i.e., ones and zeros), which are the most basic form of inform in computer science. Pixels can be comprised of anywhere between 1 to 24 or more bits depending on the complexity of the picture. *See id.* at 4-7, 60-61, 80-81. The complexity of pictures varies from black-and-white to grayscale to colored regions and patterns (e.g., smooth areas, edges, textures), and the more complex, the more bits are required to represent the picture. *Id.* For example, a black and white picture only needs 1 bit, but colored pictures may need 8 or more bits to more accurately represent the color. *Id.* This is where image or video coding come into play.

B. Video Coding Overview

20. Video coding (also referred to as video compression) is the process of

numerically representing pictures for efficient storage and transmission and so that they can be reproduced in real-time. *Netravali* at vii-viii. Each picture contains a lot of information, and thus, a video sequence contains a great deal more. *Id.* at 1-8. To efficiently store and transmit videos, they must be compressed. *See Chiu*, 1:20-2:57. Compression allows for redundant information to be removed within a picture (spatial or intraframe redundancy) or between pictures (temporal or interframe redundancy). *See Furht* at 175-77. There are several different video compression standards. Some of the most widely common being the ones developed by JPEG and MPEG. *Id.* In the 1990s, compression would remove redundancy in a few ways, such as transforming frames or estimating the motion between frames. *See id.* at 181-197.

21. Transforming signals is a part of video coding that aims to reduce the amount of information that needs to be coded or compressed. *See Chiu*, 1:45-62; *Netravali*, 175-77. Transform coding typically attempts to reduce redundancy within a frame (e.g. between pixels) rather than between frames. *See Chiu*, 1:45-62; *Netravali* at 175-76. In compression, commonly used transformations consist of mathematical operations that are applied to an image or video frames and that convert the image or frames from the spatial to the frequency domain. *See Netravali* at 18-21; *Furht* at 196.

22. Motion compensation is the process of estimating or predicting motion

(i.e., temporal change) between successive frames. *See Chiu*, 1:20-2:57; *Netravali* at 339-40. A later frame can be generated using a previous frame (also called as "reference frame") and the associated motion information. *See Chiu*, 3:29-35; *Netravali* at 615-16. Thus, only the motion information, rather than the later frame itself, needs to be stored or transmitted. *See Chiu*, 3:29-35; *Netravali* at 615-16. This way, redundant information between frames—portions of the later frame that do not have motion relative to the previous frame—can be skipped from storage or transmission and encoded as the previous frame. *See Chiu*, 9:27-10:16; *Netravali* at 615-16. As part of compression, some video encoders perform motion estimation and compensation and then apply a transformation on the resulting difference to reduce the redundancy between frames. *See Chiu*, 1:20-2:57. This reduces the amount of information to be coded and stored. *See Chiu*, 1:20-2:57; *Netravali* at 339-40.

C. Perceptual Coding

23. The human visual system cannot always process or perceive all of the information within a frame or between frames. *See, e.g.*, *Chiu*, 8:41-9:18; *Gu*, 5:4-6:35. For example, one pixel may be masked (i.e., less or not noticeable by a human visual system) by the surrounding pixels because they are brighter or have a higher average luminance than the pixel. *See Chiu*, 8:41-8:67; *Gu*, 5:44-65. A portion of a frame may be more or less noticeable because of motion (i.e., change in time)

between frames as well. *Gu*, 5:44-65.

24. Perceptual coding is a form of video coding that seeks to exploit these properties of the human visual system to encode videos more efficiently. *See Chiu*, 6:1-8:40; *Gu*, 6:35-67. Perceptual coding is largely a method of lossless compression—a compression method that recodes videos using fewer bits. *See Chiu*, 6:1-8:40; *Gu*, 6:35-67. It differs from other coding techniques because it assigns pixels fewer bits based on each pixel's importance to human viewers' perception of the visual information. *See Chiu*, 6:1-8:40; *Gu*, 5:44-6:35. The pixel's importance is determined by, for example, whether the human visual system can sense a change, either a spatial similarity between different portions of the same frame or a temporal change between the corresponding pixels in two different frames. *See Chiu*, 7:38-52; *Gu*, 5:40-6:35.

VI. THE '532 PATENT

25. The '532 patent's specification discloses a video encoder and methods of preprocessing frames that involves lossless encoding based on the human visual system. '532 patent, 2:25-28. The specification describes the invention is "for video compression which is generally lossless vis-a-vis what the human eye perceives." *Id.*, 2:22-24. For example, Figure 2 shows the "preferred embodiment" of its "video

compression system having a visual lossless syntactic [VLS] encoder."[1] *Id.*, 4:31-32; *see* Fig. 2 (below).

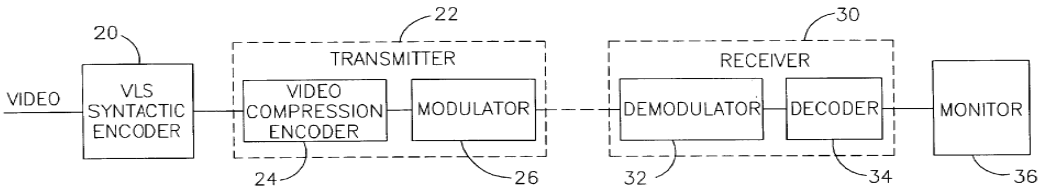


FIG.2

26. A video enters the "VLS syntactic encoder 20 [which] 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." *Id.*, 5:39-44. In other words, the VLS encoder preprocesses a video before it is compressed so that a video takes up less space in terms of size in bits when it is stored or when it is transmitted. The '532 patent is generally limited to discussing what occurs in in this VLS encoder and does not get into great detail describing other aspects of the video encoding process. *See id.*, 5:34-64. Furthermore, the '532 patent discloses one, preferred embodiment for most aspects of its encoders and

[1] A POSA would have understood syntactic to mean syntax or rule-based coding.

methods. *See id.*, 2:25-9:42 (describing the preferred embodiment); 9:43-60, 10:65-11:45 (describing the "alternative embodiments" of the interframe and intraframe processors).

27. The VLS encoder is described in Figure 3 as "a block diagram illustration of the details of the visual lossless syntactic encoder of FIG. 2." *Id.*, 4:35-36; Fig. 3 (annotated below). This figure describes the "video lossless encoder" claimed in claims 1-5. *See id.*, claims 1-5, 12:10-49. This encoder also corresponds to the methods claimed in claims 6-17. *See id.*, claims 6-17, 12:50-14:34; .

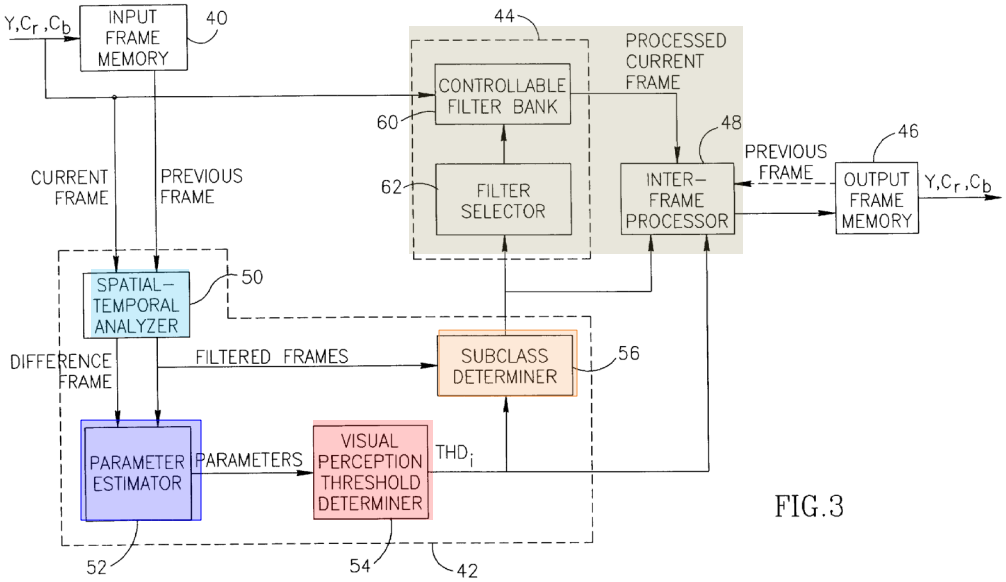


FIG.3

See id., Fig. 3 (annotated).

28. For example, independent claim 6 recites:

> 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 associate subclass.

29. The most relevant parts of the VLS are those in the "frame analyzer 42," the "intra-frame processor 44," and the "inter-frame processor 48." *See id.*, 5:66-6:3. The "frame analyzer 42" (also called the "Analyzer 42") comprises "a spatial-temporal analyzer 50, a parameter estimator 52, a visual perception threshold determiner 54 and a subclass determiner 56." *Id.*, 6:22-24. And as the specification describes, the purpose of the analyzer 42 is to "analyze[] each frame to separate it into subclasses, where subclasses define areas whose pixels cannot be distinguished from each other." *Id.*, 6:3-5. Notably, the specification describes how its invention operates on a per-pixel basis as opposed to a per-detail basis. *See, e.g.*, *id.*, 6:22-9:42. And, as the '532 patent describes, pixels are what ultimately make up details. *See id.*, 8:35-42.

30. The specification describes details as pixels or groups of pixels that are

related based on human visual perception. The "spatial-temporal analyzer 50" performs what the '532 patent describes as "spatially and temporally separating and analyzing details." The spatial-temporal analyzer "divides the video frame into portions having different detail dimensions." *Id.*, 2:35-36. For example, it "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." *Id.*, 6:30-34. "[F]ilters 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 11 pixels)." *Id.*, 6:66-7:2. It also "generates difference frames between the current frame and another, earlier frame." *Id.*, 7:5-6. Put simply, the spatial-temporal analyzer splits pixels within a frame up into different groups based on similar characteristics.

31. The specification states that "[t]he 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)." *Id.*, 5:28-33. It uses Figure 1 to describe the difference between detected and distinguished. *Id.*, 5:22-24 ("there is at least one bird, labeled 14, which the eye can detect but cannot determine all of its relative contrast details.").

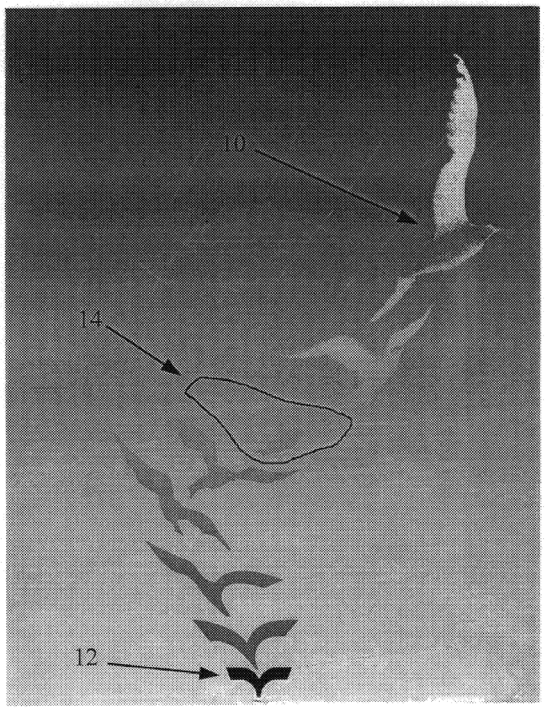


Fig. 1

32. But it is not until the classifying step, where the VLS determines whether the pixels or groups of pixels belong to the same detail, defined by a subclass. *See id.*, 6:3-5 ("subclasses define areas whose pixels cannot be distinguished from each other."); 2:38-40 ("[e]ach subclass includes pixels related to the same detail"). Based on these examples, "detail" at least encompasses a pixel or a group of pixels that is detectable or distinguishable based on the human visual system.

33. Perhaps obviously, the "[p]arameter estimator 52" is what calculates parameters. More precisely, this is where the VLS encoder determines parameters,

or "information content of the current frame." *Id.*, 7:10-17. "The parameters are determined on a pixel-by-pixel basis or on a per frame basis, as relevant." *Id.* The '532 patent lists several examples of what can qualify as a parameter, but it never states what a parameter cannot be. *Id.*, 7:10-8:25.

34. "Visual perception threshold determiner 54 determines the visual perception threshold THD_i per pixel" using the estimated parameters. *Id.*, 8:26-27. Thresholds "define contrast levels above which a human eye can distinguish a pixel from among its neighboring pixels of said frame." *Id.*, 12:10-27. They are then used in by the subclass determiner. The "subclass determiner 56 compares each pixel i of each high pass filtered frame HPF-X to its associated threshold THD, to determine whether or not that pixel is significantly present in each filtered frame. . . . [It] then defines the subclass to which the pixel belongs." *Id.*, 8:35-42. For example, if a pixel is not significantly present in the filtered frames, it belongs to a large detail or one that is "only detected but not distinguished." *Id.*, 8:43-9:13; Table 2. Or if the pixel is present in all of the filtered frames, then it belongs to a very small single detail. *Id.*

35. After the pixels are classified into subclasses (i.e., types of details), intra-frame processor 44 and inter-frame processor 48 then transform the pixels based on their subclasses. *See id.*, 9:16-11:45. "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." *Id.*, 6:3-12. Once the transformation is complete, the video frames then proceed through the rest of the video encoding process.

VII. THE LEVEL OF ORDINARY SKILL IN THE ART

36. A POSA as of the earliest claimed priority date for the '532 patent would have had a Bachelor's Degree in Computer Science, Computer Engineering, or Electrical Engineering, and two or more years of experience working with image processing or visual perception based video coding. Additional education, such as a Master's degree of Computer Science, Computer Engineering, or Electrical Engineering with an emphasis on image processing or computer vision, could substitute for professional experience, and significant work experience could substitute for formal education.

37. I qualify as a POSA as detailed above.

VIII. GROUND 1: *CHIU* IN VIEW OF *GU* AND THE KNOWLEDGE OF A POSA

38. In my opinion, Claims 6 and 16 would have been obvious to a POSA in view of *Chiu* and *Gu*. This is because I believe that a POSA, having knowledge of these two sources, would have been motivated to combine the disclosures of *Chiu*

and *Gu* to create the claimed elements.

A. Overview of *Chiu*

39. *Chiu* discloses "video compression techniques and, more particularly, [] perceptual-based video signal coding techniques resulting in reduced complexity of video coders implementing same." *Chiu*, 1:15-18. It teaches a novel video codec that "includes similar components as the [prior art] video encoder 10 (FIG. 1), with the important exception of the novel inclusion of a perceptual preprocessor 110." *Id.*, 7:55-59. *Chiu*'s perceptual preprocessor relies on "the insensitivity of the human visual system (HVS) to the mild changes in pixel intensity in order to segment video into regions according to perceptibility of picture changes." *Id.*, 8:27-31.

40. The methods detailed in *Chiu* are based on "comparing elements of a portion of a first video image (e.g., pixels of a macroblock of a current frame) with elements of a portion of a second video image (e.g., corresponding pixels of a macroblock of a previous frame)." *Id.*, 4:40-45. In these methods, the intensity difference between those elements is compared to a visual perception threshold through a process called "bush-hogging." *See id.*, 7:14-30. *Chiu* teaches that the result of this process categorizes frames into at least two groups, for example: macroblocks to be motion compensated and macroblocks to not be motion compensated. *See, e.g.*, *id.*, 9:27-45.

B. Overview of *Gu*

41. *Gu* is all about "the compression and encryption of data, and more particularly to compressing data with a human perceptual model." *Gu*, 1:22-24. *Gu* identifies that "[c]ompeting with the need to compress data is the need to maintain the subjective visual quality of an image." *Id.*, 1:37-42. Indeed, this is a problem that most of us in the industry were addressing at this time. The way *Gu* tries to solve this is by using "performance metrics that take the psycho visual properties of the human visual system (HVS) into account are needed." *Id.*, 1:37-42.

42. *Gu* describes that humans are sensitive to changes in, for example, "frequency, brightness, texture, and temporal parameters." *See id.*, 5:40-6:34. "A few human visual models incorporating these psycho-visual properties into their algorithms include: just-noticeable-distortion (JND), . . . [which] provides each pixel with a threshold of error visibility, below which reconstruction errors are rendered imperceptible." *Id.*, 1:43-50. In other words, the JND can be used to "remove the redundancy due to spatial and temporal correlation and the irrelevancy of perceptually insignificant components from video signals." *Id.*, 7:2-6.

43. To analyze details of a frame, *Gu* uses "[a] 3-D wavelet analyzer . . . to generate low and high frequency subbands" of frames. *Id.*, 2:25-28. Then, for each frame, *Gu* teaches how "[a] just-noticeable distortion model generator is used to calculate a profile of the video image based upon local signal properties." *Id.* It

calculates JNDs for each pixel through the use of values such as the "average background luminance and texture masking" and "the average interframe luminance difference." *See id.*, 10:2-11:55. *Gu* summarizes that "[f]or a successful estimation of JND profiles, the subject should not be able to discern the difference between a video sequence and its JND-contaminated version." *Id.*, 7:6-17

44. I understand that *Gu* claims priority to US Provisional application No. 60/136,633, filed May 27, 1999. I understand that for *Gu* to be given the earliest possible effective filing date, and thus qualify as prior art, *Gu*'s Provisional must provide sufficient written description for at least one claim in *Gu*. I have reviewed *Gu* and *Gu*'s Provisional. In view of the limitations recited in *Gu*'s independent claim 1, it is my opinion that *Gu*'s Provisional provides sufficient written description for *Gu* to be entitled to its earliest possible effective filing date and qualify as prior art to the '532 patent.

45. Below, I have summarized some of the ways *Gu*'s Provisional provides such support for claim 1 of *Gu*. In my opinion, *Gu*'s Provisional discloses a method comprising elements (a)-(i).

1. A method, comprising: (a) arranging two successively received video frames into a set;

46. *Gu*'s Provisional relates to "3D WAVELET BASED VIDEO CODEC WITH HUMAN PERCEPTUAL MODEL." Ex. 1007, 30. It generally refers to

Figure 3 to describe this codec and the method it implements. *See id.*, 57, Fig. 3.

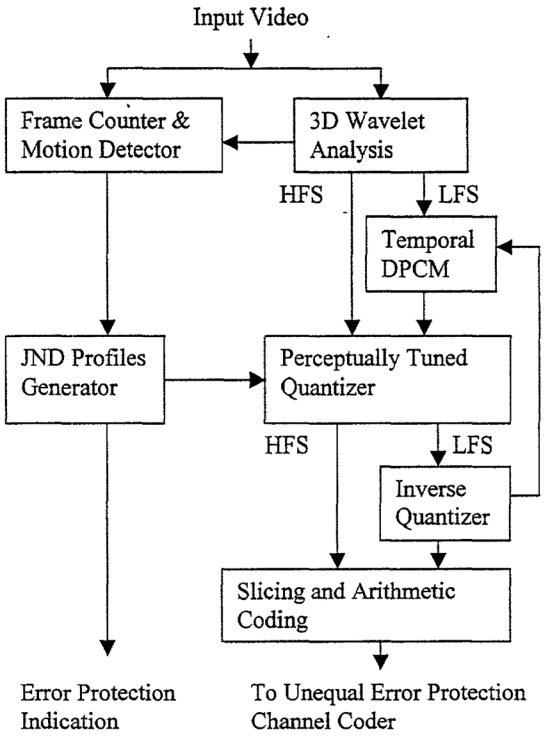


Figure 3 JND Based Video Encoder

47. "The Frame Counter & Motion Detector controls the renewing of the JND profiles from frame counter and drastic movement detection. The JND Model Generators estimate the spatio-temporal JND profile from analyzing local video signals" *Id.*, 56. "The JND profile of a **video sequence** is a function of local signal properties, such as brightness, background texture, luminance changes **between two frames,** and frequency distribution." *Id.*, 39. "In this part, the spatial-temporal JND profile **for each group of two frames** in a video sequence and the JND profiles for subbands are generated sequentially." *Id.*, 64. A POSA would understand that a

JND profile is operating on a set of two successively received video frames based on these disclosures.

2. (b) decomposing said set into a plurality of high frequency and low frequency subbands;

48. In my opinion, *Gu*'s Provisional discloses that "each pair of video frames is decomposed into 11 spatio-temporal subbands as shown in in Figure 2 (the details of such a 3D wavelet decomposition is described in Chapter 3)." *Id.*, 52.

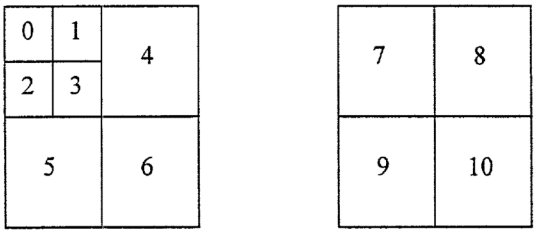


Figure 2 Subbands after 3D Wavelet Decomposition

Ex. 1007, Fig. 2, 53.

49. The 3D Wavelet is described as using "the simple two-tap Haar filter, which essentially separates the signal into temporal low frequency and high frequency parts. Then, the low frequency part is decomposed for two levels with the spatial 2-D wavelet decomposition, while the high frequency part is decomposed for one level. The Antonini (7,9)-filter [21] is used here." *Id.* at 62. A POSA would have understood this to show the 11 spatio-temporal subbands in Figure 5, where LP stands for lowpass and HP stands for highpass. A box labeled with LP would have indicated a low frequency subband, whereas a box with HP would have indicated a

high frequency subband.

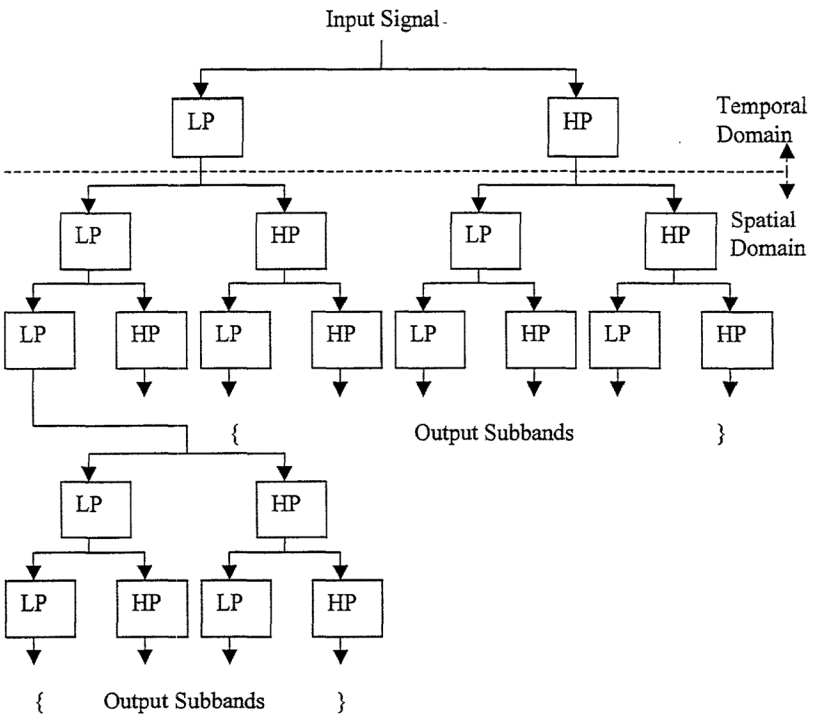


Figure 5 Heterogeneous 3D Wavelet Decomposition

3. (c) generating a human perceptual model for each of the plurality of subbands;

50. Chapter 2 of *Gu*'s Provisional (Ex. 1007) provides for "the Human Perceptual Model of JND." *See id.*, 44-55. It teaches that JND is a human perceptual model. *Id.* As I cited to above, *Gu*'s Provisional makes a JND profile. "From the spatio-temporal JND profile that quantitatively measures the perceptual redundancy inherent in full-band video signals, the JND profiles for each subband are set up with certain distortion allocation [22] for different subbands." *Id.*, 52.

4. (d) encoding said low frequency subbands to produce encoded low frequency subbands;

51. In my opinion, *Gu*'s Provisional also provides support for this limitation. It discloses that "[t]he signal is decomposed into frequency subbands and these subbands are encoded independently or dependently." *Id.* at 39. Therefore, a POSA would understand that the low frequency subbands are encoded. One example that *Gu*'s Provisional gives for this is "[t]he spatial LLLL temporal L subband will be encoded by DPCM." *Id.*, 56. This is also referred to as the "lowest frequency subband." *Id.*, 93. And *Gu*'s Provisional discloses that, when encoded, the subbands are comprised of "wavelet coefficients." *Id.*

5. (dd) inverse quantizing said encoded low frequency subbands:

52. As noted above, Figure 3 discloses the JND based video encoder that implements this method. The Provisional teaches that "Inverse Quantization and 3D Wavelet Synthesis parts implement the inverse operation of the quantization and 3D wavelet decomposition in the video encoder." *Id.* at 56-57. Thus, following Figure 3, a POSA would have understood that the encoded low frequency subbands are inverse quantized.

6. (e) quantizing said high frequency subbands and said low frequency subbands according to said generated human perceptual models to generate bitstream for each of said high frequency subbands and a bitstream for each of said low frequency subbands;

53. Also referencing Figure 3 above, *Gu*'s Provisional discloses that "[t]he Perceptually Tuned Quantizer implements quantization for the wavelet coefficients in each subbands according to their JND profiles. . . . Then the data from all subbands goes through the Slicer and Arithmetic Coding part to do slicing and entropy coding. Afterward we get compressed video signals in 11 bit streams." *Id.* at 56. It teaches that "[t]he perceptually tuned quantizer is the core of the source encoder. With the JND profiles ***for each subband*** at hand, the most important task is to allocate the available bits to certain coefficients obtained from the wavelet decomposition as efficiently as possible." *Id.* at 64. "[O]ne quantization table that leads to the uniform error energy distribution over all subbands is set up for each subband. It is transmitted in the header of the bit stream for this subband." *Id.* at 66. Based on these disclosures alone, a POSA would have understood that *Gu*'s Provisional is teaching quantizing each of the subbands according to human perceptual JND-based models to generate a total of 11 bitstreams, including one bitstream for each low subband and one bitstream for each high subband.

7. (f) redefining said generated bitstreams for each of said high frequency subbands to produce a plurality of redefined high frequency bitstreams;

54. Furthermore, *Gu*'s Provisional also describes how the inventors "derive[d] a new slicing method to truncate the data from each subband into small slices before arithmetic coding." *Id.* at 42. "In the environment of a noisy channel, in order to prevent decoding errors from spreading, a slicing algorithm is derived to truncate the whole subband into short bit streams before arithmetic coding. The idea is to make each such small bit stream carry the same amount of 'distortion sensitivity'. . . . So every time when such a short bit stream is being encoded, a new adaptive statistical model is set up for it. Before the arithmetic encoded output data of these short bit streams are merged into one bit stream, header and trailer symbols are added so that they can be selected out from the received bit stream at the decoder side." *Id.* at 69-70. "First, the data frame [sub band] which will be transmitted is broken into slices $\{\mathbf{s_1}, \mathbf{s_2}, \ldots \mathbf{s_M}\}$. The locations of breaking points are decided by a randomly generated vector v. This vector v is updated after a short time interval. Each slice is AC [arithmetic coding] encoded respectively into bit stream $\{\mathbf{b_1}, \mathbf{b_2}, \ldots b_M\}$. Then, these bit streams are concatenated into one bit stream $\mathbf{b_{total}}$. *Id.* at 92. A POSA would have understood these disclosures to demonstrate how *Gu*'s Provisional teaches redefining the bit stream for every subband—including high frequency and low frequency subbands.

8. (g) redefining said generated bitstreams for each of said low frequency subbands to produce a plurality of redefined low frequency bitstreams;

55. It is my opinion that the disclosures and discussion above in element (f) provide support for this limitation as well.

9. (h) channel coding said plurality of redefined high frequency bitstreams and said plurality of redefined low frequency bitstreams; and

56. *Gu*'s Provisional continues that "[t]hese bit streams are fed into the Unequal Error Protection Channel Coder and an error protection indication from JND Model Generator is also given to the channel coder." *Id.* at 56. A POSA would have understood this to be channel coding the redefined bitstreams. The Provisional provides further support for this limitation in Section "3.6 Perceptual Channel Coding." *Id.* at 70. It teaches that "[i]n our practical video transmission system over a satellite channel, rate compatible punctured convolutional (RCPC) codes [27] are adopted. The advantage of using RCPC codes is that the high rate codes are embedded into the lower rate codes of the family and the same Viterbi decoder can be used for all codes of a family. Reed Solomon code and Ramsay interleaver plus RCPC is used to protect the data from spatial LLLL temporal L subband. Cyclic redundancy check (CRC) codes are combined with RCPC for other less significant subbands to assure acceptable video quality even under bad channel conditions." *Id.* at 70.

10. (i) repeating steps (a)-(h) for a next set until each of said received video frames have been compressed.

57. It is also my opinion that *Gu*'s Provisional provides support for this final limitation. Because this method operates on a video, a POSA would have inherently understood the method to repeat until the entire video has been compressed. The Provisional discloses that "to calculate the spatio-temporal JND profile for each frame in a video sequence, the spatio JND profile of itself and its previous reference frame are required." *Id.* at 51. "The Frame Counter & Motion Detector is designed to control the renew process of the JND. The frame counter is used to count the number of frames being coded. After a certain number of frames (typically 10 or 20 frames) have been coded with the current JND model, this model is refreshed with the updated signal." *Id.* at 63. Therefore, the process is repeated.

C. Rationales for the Combination of *Chiu* and *Gu*

58. *Chiu* and *Gu* are analogous art to the '532 patent. The references are from the same field of endeavor of visual lossless encoding because they both disclose methods of visual lossless encoding. *Chiu* and *Gu* are both at least reasonably pertinent to a problem addressed by the '532 patent—methods of video compression that are "lossless vis-a-vis what the human eye perceives." '532 patent, 2:22-24. For example, *Chiu* discloses a perceptually-lossless video preprocessing method of using JND to determine a visual perception threshold, based on which

frame picture details are classified into subclasses and transformed in accordance with their associate subclasses. And *Gu* discloses a method for generating the JND by spatially and temporally separating and analyzing the picture details. Because of these similarities, a POSA would have been motivated to implement *Gu*'s JND generation method in *Chiu*'s method to compute *Chiu*'s visual perception threshold values. A POSA would have found it obvious to improve *Chiu* with *Gu*'s JND generation method to calculate the JND used in *Chiu*'s threshold value to improve the performance of *Chiu*'s video encoding method.

59. One reason a POSA would have thought to combine *Chiu* and *Gu* because they both seek to solve the same problems—improving methods of video compression using perceptually lossless coding. For example, *Chiu* notes that "[d]epending on picture content, perceptual preprocessing achieves varied degrees of improvement in computational complexity and compression ratio without loss of perceived picture quality." *Chiu*, 4:34-37. *Gu* parallels *Chiu*'s goals, aiming to solve "how to encode these signals with the lowest possible bit rate without exceeding the error visibility threshold," the threshold being used to measure whether an image distortion caused by the encoding would be visually perceivable. *Gu*, 1:54-56.

60. Both references seek to reduce the bitrate without introducing visible distortions. *Chiu* reduces computational complexity and improves the compression ratio (i.e., reduce the bitrate) by "perceptual preprocessing in a video encoder that

takes advantage of the insensitivity of the human visual system (HVS) to mild changes in pixel intensity in order to segment video into regions according to perceptibility of picture changes." *Chiu*, 4:22-30; 4:34-37. *Gu* shares a similar goal: "one problem to be solved for optimizing the delivery of digital data streams is how to locate perceptually important signals in each frequency subband. A second problem is how to encode these signals with the lowest possible bit rate without exceeding the error visibility threshold." *Gu*, 1:51-56.

61. Both references explicitly avoid applying standard non-perceptual error measures or models for quality assessment, such as mean square error (MSE) and mean absolute error (MAE). *See, e.g.*, *Chiu*, 2:24-4:10 (discussing mean absolute error or, alternatively, mean square error as a less desirable coding technique); *Gu*, 5:21-39 (discussing disadvantages of prior non-perceptual systems).

62. Because these references are in the same field of endeavor and share the same goals, a POSA would have looked to improve *Chiu* with the teachings of *Gu* regarding the preprocessing of video frames to achieve perceived visual losslessness.

63. In addition to trying to solve the same problems, *Chiu* and *Gu* also propose similar solutions. As will be discussed in depth in Sections VII.C and VIII.D.2, a POSA would have specifically been motivated to apply *Gu*'s JND generation to compute *Chiu*'s visual perception threshold based on the references'

teachings on at least masking, JNDs, and perceptual redundancy.

64. *Chiu* and *Gu* both disclose using the JND model to study the visual perceptibility and visual redundancy of the details. One of *Chiu*'s methods uses a JND when calculating its visually perceptible threshold value. *Chiu*, 10:30-36. Likewise, *Gu* uses "[t]he JND model . . . [to] provide[] each pixel with a threshold of error visibility, below which reconstruction errors are rendered imperceptible." *Gu*, 1:47-50. In my opinion, this combination would have involved applying a known technique—*Gu*'s JND generation—to prior art that could be improved—*Chiu*'s method of encoding details using JNDs. And the result would have been a predictable variation because *Gu*'s method would merely have improved a similar device (i.e., *Chiu*'s perceptual preprocessor). *Id.* A POSA would have had a reasonable expectation of success for this combination because *Gu*'s JND is the same JND used by *Chiu* for determining its visual perception threshold. *Chiu* even explicitly provides this motivation when it states: "given the teachings of the invention, one of skill in the art will contemplate other methods of modeling the threshold function." *Chiu*, 9:16-18. Furthermore, using perceptual models, or accounting for the human visual system, in video codecs was a standard practice at the time of invention of the '532 patent. *See, e.g.*, *Netravali* at 253-305. Specifically, using perceptual models, such as JND, in video codecs would have been a routine combination for a POSA before the earliest priority date of the '532

patent. *See, e.g.*, *Jayant* at 1404; *Netravali* at 253-305.

D. Independent claim 6

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

65. I understand that what comes before words like "comprising" or "consisting of" is typically called the preamble of the claim. I also understand that preambles are not usually part of the scope of the claim unless they include limitations that are important to understanding the rest of the claim. Based on this understanding, I believe that, to the extent the preamble is limiting, it would have been obvious over *Chiu* in view of the knowledge of a POSA.

66. The objective of the '532 patent "is to provide a method and apparatus for video compression which is generally lossless vis-a-vis what the human eye perceives." '532 patent, 2:22-24. In other words, even if the methods disclosed in the '532 patent may cause loss of certain original video information, "the resultant video displayed on monitor 36 is not visually degraded." *Id*., 5:51-53. "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." *Id*., 5:53-56. "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." *Id*., 5:56-62. "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." *Id*., 5:62-65.

67. The fact that this method is lossless is important to understanding the claim because lossless coding/compression preserves perceptually relevant information, whereas lossy coding/compression can remove such information. *See, e.g.*, *Netravali* at 584, 590-92, 614. The inclusion of *visual* is also vital to understanding the claim because it limits the type of information that is being coded with this method—it signals that this method is focused on the removal of information that will not be perceived by the human visual system. *See* '532 patent, 2:21-24; *Jayant*, 1412-14. It is also important because it provides an antecedent basis for "said frames," which are the subject of the rest of the claim. Therefore, the claimed "method of visual lossless encoding of frames of a video signal" is lossless only with respect to visual information that can be removed without humans being able to notice the difference.

68. Like the '532 patent, *Chiu* discloses a perceptual preprocessor that improves the "computational complexity and compression ratio without loss of perceived picture quality." *Chiu*, 4:22-37. The perceptual preprocessor "takes advantage of the insensitivity of the human visual system (HVS) to mild changes in

pixel intensity in order to segment video into regions according to perceptibility of picture changes." *Id*., 4:25-29. *Chiu* notes that "the invention accurately models the motion in areas with perceptually significant differences to improve the coding quality. Depending on picture content, perceptual preprocessing achieves varied degrees of improvement in computational complexity and compression ratio without loss of perceived picture quality." *Id.*, 4:32-37; *see also id*., 6:64-66 ("The ideal JND provides each pixel being coded with a threshold level below which discrepancies are perceptually distortion-free"), 8:17-22 (". . . at no loss of perceived quality"), 13:9-12 (". . . without a perceived loss in picture quality"). *Chiu* also discloses that the visual lossless video encoding method is performed on frames of a video signal. *Id.*, 10:19-21 ("In Step 402, the current frame F_k and the previously reconstructed frame $\hat{F}_{k-1}$ are input to the preprocessor 110.").

69. A POSA would have understood *Chiu* to be focused on visual lossless coding as well based on these disclosures. They demonstrate how *Chiu* is centered on removing information in a manner that the human visual system will not notice it is gone. A POSA would have understood that *Chiu*'s JND-based perceptual preprocessor is lossless—or programmed to perform a video encoding method that does not cause any loss of information perceived, i.e., distinguishable, by a human eye. *See Jayant* at 1413-14.

70. Therefore, *Chiu* in view of the knowledge of a POSA teaches a "method

of visual lossless encoding of frames of a video signal," as recited in the preamble of claim 6.

2. [6.b]: spatially and temporally separating and analyzing details of said frames;

71. *Chiu* in view of *Gu* and the knowledge of a POSA teaches this limitation. As I detail below, *Chiu*'s perceptual preprocessing method uses a visually perceptible threshold value, modeled by a JND, to analyze the degree of visual perception of the details. *Gu* discloses that the JND value is generated by spatially and temporally separating and analyzing details. A POSA would have found it obvious to implement *Gu*'s JND generation method in *Chiu*'s method to compute *Chiu*'s visually perceptible threshold value.

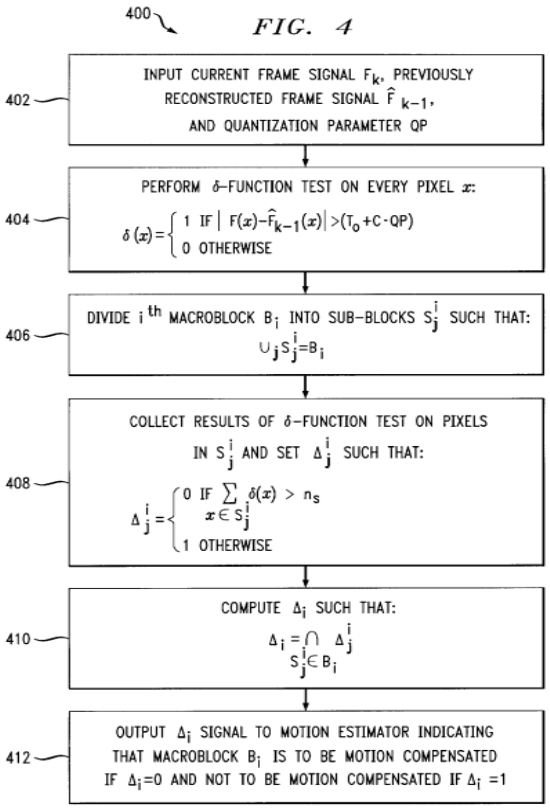
72. The '532 patent describes spatially and temporally separating and analyzing each frame by using a "[s]patial-temporal analyzer 50 [to] generate[] a plurality of filtered frames from the current frame, each filtered through a different high pass filter (HPF), where each high pass filter retains a different range of frequencies therein." *See* '532 patent, 6:30-34. "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." *Id.*, 6:51-54. And the spatial-temporal analyzer also creates difference frames. *See id.*, 7:5-9. Based on these disclosures, a POSA would have understood spatially and temporally separating and analyzing to be

about figuring out how the neighboring pixels relate to one another, both within the frame to which they belong and between frames.

a. *Chiu* teaches a perceptual preprocessing method for encoding "details"

73. As described in Section VI, "details" in the '532 patent claims refer to at least pixels or groups of pixels that are related based on human visual perception (e.g., detectable or distinguishable by a human eye).

74. *Chiu*'s video encoding method is used to process pixels or groups of pixels that are detectable or distinguishable by a human eye. Figure 4 of *Chiu* provides an overview of this method, "perceptual preprocessing method 400:"



Chiu, 10:17-19, Fig. 4. In particular, "in step 404, a δ-function test **is performed on every pixel x** such that:

$$\delta(x) = \begin{cases} 1 & \text{if } \left|F_k(x) - \hat{F}_{k-1}(x)\right| > T \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

where the visually perceptible threshold value T is modeled by a fixed value as follows:

$$T=T_0+\cdot QP. \quad (9)$$

T_0 is the just-noticeable-distortion (JND) value, QP is the quantization parameter, and c is a constant."[2] *Id*., 10:23-36. A "δ-function test" or thresholding is a way of segmenting pixels into groups—for example, based on whether it is imperceptible or distinguishable from its neighbors. *See, e.g.*, *Jayant* at 1404; *Chiu*, 7:14-30.

75. Regarding the JND value in the above equation, *Chiu* discloses that an "ideal JND provides **each pixel** being coded with a threshold level below which discrepancies are perceptually distortion-free." *Chiu*, 6:64-66. *Chiu* also discloses that "JND provides each signal being coded with a threshold level below which reconstruction error is imperceptible in the sense that a human can not [sic] perceive

[2] Equation (9) in *Chiu* appears to have a typo by missing the constant "C." The equation is correctly written in Figure 4, reproduced above.

any impairment in the video if the discrepancies are below the JND value." *Id*., 7:19-22. In view of *Chiu*'s disclosures, a POSA would have understood that the JND is used to measure whether a pixel is perceived, i.e., detectable or distinguishable, by a human eye. A POSA would have had this understanding because *Chiu*'s methods are, for example, about whether the human visual system can "distinguish the randomness of pixel locations" or "detect" difference between pixels. *See id.*, 4:66-5:4; 12:7-11. Because *Chiu* discloses that the δ-function test uses the JND and is performed on every pixel, a POSA would have understood the perceptual preprocessing method 400 is used to encoding perceivable pixels, i.e., "details."

b. *Gu* teaches generating the JND by "spatially and temporally separating and analyzing details of said frames"

76. *Gu* similarly uses JND as a threshold in its methods. For example, the reference discloses "[t]he JND model provides each signal a threshold of visible distortion, below which reconstruction errors are rendered imperceptible." *Gu*, 9:52-54. *Gu* further discloses a method for generating a "spatial-temporal JND profile for [a] set of two frames" consisting of a current frame and a previous frame. *Id*., 9:54-59. This JND generation spatially and temporally separates and analyzes details as it divides and describes pixels based on the local characteristics of the neighborhood associated with each pixel using spatial texture characteristics due to

local luminance changes, spatial average background luminance, and temporal characteristics due to intensity changes between frames.

77. *Gu*'s JND generation can be summarized by two steps. *Id*., 9:65-66. The first step separates and analyzes pixels spatially, meaning how pixels relate to one another within the same frame. The second step separates and analyzes pixels temporally, meaning how pixels relate to one another between frames.

78. First, "the perceptual redundancy inherent in the spatial domain is quantitatively measured as a 2D profile by a perceptual model that incorporates the visibility thresholds due to average background luminance and texture masking." *Id*., 9:66-10:3. "The following expression yields:

$$JND_S(x, y)=\max\{f_1(bg(x, y), mg(x, y)), f_2(bg(x, y))\},$$

$$\text{for } 0\leq x<W, 0\leq y<H \qquad (1)$$

where f_1 represents the error visibility threshold due to texture masking, $f_{[2]}$ the visibility threshold due to average background luminance; H and W denote respectively the height and width of the image; mg(x, y) denotes the maximal weighted average of luminance gradients around the pixel at (x, y); bg (x, y) is the average background luminance." *Id*., 10:3-16. "The value of mg(x, y) across the pixel at (x, y) is determined by calculating the weighted average of luminance changes around the pixel in four directions." *Id*., 10:35-37. "Four operators, $G_k(I,j)$,

for k=1, . . . 4, and I, j=1, . . . 5, are employed to perform the calculation. . . ." *Id*., 10:37-41. "The operators $G_k(I, j)$ are

$$G_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 8 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & -3 & -8 & -3 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, G_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 8 & 3 & 0 & 0 \\ 1 & 3 & 0 & -3 & -1 \\ 0 & 0 & -3 & -8 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix},$$

$$G_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 8 & 0 \\ -1 & -3 & 0 & 3 & 1 \\ 0 & -8 & -3 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix}, G_4 = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 \\ 0 & 3 & 0 & -3 & 0 \\ 0 & 8 & 0 & -8 & 0 \\ 0 & 3 & 0 & -3 & 0 \\ 0 & 1 & 0 & -1 & 0 \end{bmatrix},$$

."

Id., 10:56-11:10. Moreover, "[t]he average background luminance, bg(x, y), is calculated by a weighted lowpass operator, B(I, j), 1, j=1 . . . 5." *Id*., 11:11-12.

$$bg(x, y) = \frac{1}{32}\sum_{i=1}^{5}\sum_{j=1}^{5} p(x-3+i, y-3+j)\cdot B(i, j) \qquad (5)$$

$$\text{where } B = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 2 & 0 & 2 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Id., 11:15-28.

79. Second, *Gu* addresses how pixels relate to one another between frames. *Gu* discloses "the JND profile representing the error visibility threshold in the spatio-

temporal domain is determined according to the following expression:

$$JND_{S\text{-}T}(x,y,n)=f_3(ild(x,y,n))\cdot JND_S(x,y,n) \quad (6)$$

where ild(x, y, n) denotes the average interframe luminance difference between the n th and (n−1) th frame." *Id*., 11:29-36. "Thus, to calculate the spatio-temporal JND profile for each frame in a video sequence, the spatio JND profile of itself and its previous reference frame are required.

$$ild(x,y,n)=0.5\cdot[p(x,y,n)-p(x,y,n-1)+bg(x,y,n)-bg(x,y,n-1)]$$

f_3 represents the error visibility threshold due to motion." *Id*., 11:36-44.

i. *Gu* teaches "spatially and temporally separating . . . details of said frames"

80. As in the '532 patent, these steps in *Gu* are about figuring out how the neighboring pixels relate to one another, both within the frame to which they belong and between frames. A POSA would have understood the above steps to teach the claimed "spatially and temporally separating . . . details of said frames" because they transform the pixel information in each video frame into:

- mg(x,y): spatial texture characteristics due to local luminance changes;
- bg(x,y): spatial average background luminance;
- ild(x,y,n): temporal characteristics due to motion or due to temporal intensity changes.

81. Among them, mg(x,y) quantifies the "spatial nonuniformity of the background luminance," also known as local texture characteristics. *Id.*, 5:61-63; 10:13-15. mg(x,y) is computed using four directional high-pass filters or operators "$G_k(I,j)$, for k=1, . . . 4, and I, j=1, . . . 5." *Id.*, 10:34-11:10 ("The value of mg(x, y) across the pixel at (x, y) is determined by calculating the weighted average of luminance changes around the pixel in four directions."). A POSA would have understood the $G_k(I,j)$ operators to be highpass filters because they measure the spatial nonuniformity and thus, when applied to a pixel surrounded by a region having a constant or non-varying intensity (i.e., zero frequency), the output value is zero. *See Wu* at 802-04. Similarly, when applied to a region with a low-varying intensity (i.e., low frequency), $G_k(I,j)$ results in a very small value, i.e., it attenuates low-varying intensities corresponding to low frequencies. *See Jayant* at 1395; *see also Wu* at 802-04. Therefore, $G_k(I,j)$ preserves pixel information associated with large spatial nonuniformity while filtering out (removing/attenuating) pixel information associated with uniform spatial regions. Because mg(x,y) relies on

$G_k(I,j)$, a POSA would have understood it to be spatially separating the details based on high-frequency and low-frequency spatial information.

82. bg(x,y) quantifies the local "average value of background luminance" in the local region (i.e., neighborhood) associated with pixel (x,y). *Gu*, 5:49; 10:15; 11:11-28. bg(x,y) "is calculated by a weighted lowpass operator B(I, j), I[sic], j=1 . . . 5." *Id*., 11:11-12. Lowpass operators are filters that retain pixel information associated with uniform spatial regions while filtering out (removing/attenuating) pixel information associated with large spatial nonuniformity. *See Wu* at 802-04. Generally, lowpass filtering is associated with areas of small changes or difference. *See Wu* at 802-04. Therefore, a POSA would have understood bg(x,y) to keep pixel information associated with a small change in background luminance. And because the pixels are being filtered out or preserved based on background luminance, a POSA would have further understood this to be another form of spatially separating the details.

83. ild(x,y,n) quantifies "temporal[] [luminance] chang[es] . . . as an object moves in a scene." *Gu*, 6:6-12. The comparison of a pixel between frames defines whether there was any movement between frames—i.e., whether a detail moved from one pixel to another. *Id*, 6:5-17. ild(x,y,n) is calculated as the local "average interframe luminance difference between the n th frame (n-1) th frame" at pixel (x,y). *Id*., 11:34-43. Basically, what ild(x,y,n) is doing is creating a variation of a

difference frame because it is comparing the current frame ($p(x, y, \mathrm{n})$) to a previous frame ($p(x, y, n-1)$). ild(x,y,n) outputs a zero or small value when the two frames are the same or only have a small temporal difference (i.e., when the temporal frequency is zero or low), while it outputs a large value when the temporal difference is large. What this means is that ild(x,y,n) is acting as a highpass filer and is keeping pixel information associated with a large temporal change while attenuating/removing pixel information associated with a small temporal change. And because the pixels are being filtered out or preserved based on temporal change, a POSA would have further understood this to be a form of temporally separating the details.

84. To summarize, mg(x,y) and bg(x,y) filter the pixel information in each video frame with respect to its visual characteristics "in the spatial domain," while ild(x,y,n) filters the pixel information in each video frame with respect to its temporal characteristics. *Id*., 10:13-15; 10:34-11:28; 11:34-43. A POSA would have understood these functions to be forms of separating details because they relate neighboring pixels to one another, both within the frame to which they belong and between frames. Therefore, *Gu* teaches the claimed "spatially and temporally separating . . . details of said frames" by disclosing calculating, respectively, mg(x,y) (which involves spatial highpass filtering) and bg(x,y) (which involves spatial lowpass filtering) in the spatial domain and ild(x,y,n) (which involves creating a

difference frame) in the temporal domain.

ii. *Gu* teaches "spatially and temporally analyzing . . . details of said frames"

85. In the above steps for generating the JND model, *Gu* also teaches the claimed "spatially and temporally . . . analyzing details of said frames" by disclosing the mathematical operations for calculating the spatial (i.e., mg(x,y) and bg(x,y)) and temporal (i.e., ild(x,y,n)) components. *Gu*'s use of the operators $G_k(I, j)$ and B(I, j) not only is a part of the separation but of the analysis of details as well. *Gu*, 10:56-11:25. A POSA would have understood that the highpass operators, such as $G_k(I, j)$, and weighted lowpass operator, such as B(I, j), all in the form of matrices, are spatial filters that analyze how one pixel relates to another. *See Jayant* at 1392-1395; *see also Wu* at 802-04. Similar to the disclosures in *Gu*, the '532 patent discloses the use of filters, or operators, as one way to do spatial analysis. *See, e.g.*, '532 patent, 6:30-34 ("spatial-temporal analyzer 50 generates a plurality of filtered frames from the current frame"). And filtering was a form of analysis known in the art at this time. *See Jayant* at 1395. Moreover, in generating the temporal component, *Gu* further performs analysis by calculating "the average interframe luminance difference between the nth and (n−1)th frame." *Gu*, 11:34-43; '532 patent, 7:5-6 ("[spatial-temporal] analyzer 50 also generates difference frames between the current frame and another, earlier frame"). As described above, a POSA

would also have understood that the generation of the temporal component (i.e., ild(x,y,n)) as taught by *Gu*, is performed by using highpass difference operations.

86. In sum, *Gu* teaches "spatially and temporally separating" details by disclosing separating pixels or groups of pixels based on spatial components (average background luminance and texture) and temporal change components. And *Gu* teaches "spatially and temporally . . . analyzing" details by disclosing performing spatial filtering (calculations using high and low pass operators) and temporal filtering (motion analysis) of the visually perceivable pixels or groups of pixels, to determine, respectively, the spatial (i.e., mg(x,y) and bg(x,y)) and temporal (i.e., ild(x,y,n)) components.

c. A POSA would have found it obvious to implement *Gu*'s JND generation method in *Chiu*'s method to compute *Chiu*'s visually perceptible threshold value

87. One of *Chiu*'s methods relies on a threshold defined by a JND. *Chiu*, 10:30-36. A POSA would have found it obvious to implement *Gu*'s JND generation method in *Chiu*'s method to calculate the JND used in *Chiu*'s threshold value because the references are in the same field of endeavor and the combination would yield predictable results.

88. First, a POSA would have been motivated to make such combination because *Chiu* and *Gu* are directed to the same field of endeavor. As discussed in depth in Sections VIII.C and VIII.D.2, *Chiu* and *Gu* aim to solve the same problems

in visual lossless encoding (e.g., balancing bitrate reduction against picture quality). And they both offer solutions to these problems using JNDs. In my opinion, this alone shows that the references are in the same field of endeavor.

89. But the similarities between *Chiu* and *Gu*'s teaching go beyond those previously discussed. Important to JND generation, both references also teach how important spatial masking and temporal masking are to perceptual coding. *Chiu* discloses that "the determination of T is very complicated for video because video involves temporal masking and spatial masking at the same time." *Chiu*, 6:50-53. *Gu* similarly teaches how "the derivation of JND profiles must take both spatial and temporal masking effects into consideration." *Gu*, 7:12-14.

90. Moreover, the JND model in both references is used to study the visual perceptibility and visual redundancy of the details. For example, *Chiu*'s method 400 uses the JND to determine the visually perceptible threshold value, which is then used to determine what details should be recoded with fewer bits. *Chiu*, 10:30-36;7:18-22 ("JND provides each signal being coded with a threshold level below which reconstruction error is imperceptible in the sense that a human cannot perceive any impairment in the video if the discrepancies are below the JND value."); 12:7-24. Likewise, *Gu* uses "[t]he JND model . . . [to] provide[] each pixel with a threshold of error visibility, below which reconstruction errors are rendered imperceptible." *Gu*, 1:47-50. Thus, *Chiu* and *Gu* are both using JNDs to solve the

same problem (i.e., using a visually perceptible threshold to locate or identify perceptually important signals). Because *Chiu* does not explicitly provide how to calculate a JND, a POSA would have been motivated to look to *Gu* to understand how to calculate it. A POSA therefore would have been motivated to combine *Chiu*'s teachings of how JNDs can be used inthreshold determination with *Gu*'s teachings of how to calculate JND values for each detail.

91. Such a combination would have required minimal modifications to *Chiu*'s method. *Chiu*'s perceptual preprocessor already determines thresholds on a per-pixel/per-detail basis by using factors such as "(i) average spatial and temporal background luminance level against which the pixel is presented; (ii) supra-threshold luminance changes adjacent in space and time to the test pixel; and (iii) spatio-temporal edge effects." *Chiu*, 7:38-44. These factors are similar to how *Gu* determines its JND using mg(x,y) (luminance changes), bg(x,y) (background luminance), and ild(x,y,n) (temporal effects).

92. A POSA would have also appreciated that combining the teachings of *Chiu* and *Gu* would yield a predictable result. This is because *Chiu* and *Gu* both disclose the same kind of JND. For example, *Chiu* discloses that "JND provides each signal being coded with a threshold level below which reconstruction error is imperceptible in the sense that a human cannot perceive any impairment in the video if the discrepancies are below the JND value." *Chiu*, 7:18-22. *Gu* mirrors this

disclosure, stating the "JND provides each signal to be coded with a visibility threshold of distortion, below which reconstruction errors are rendered imperceptible." *Gu*, 7:6-9. *Chiu* describes "[t]he information for updating such pixels can be considered to be perceptual redundancy." *Chiu*, 7:22-24. *Gu* discloses its JND model is used "[t]o remove . . . the irrelevancy of perceptually insignificant components from video signals." *Gu*, 7:2-6. *Chiu*'s JND can be modeled "[i]n accordance with Weber-Fechner theory." *Chiu*, 7:14-16. And *Gu*'s JND model is also modeled after the Weber-Fechner theory. *Id*., 5:49-53. Given these teachings, a POSA would have appreciated that the JNDs described in *Chiu* and *Gu* are the same values (i.e., a threshold level below which reconstruction error is imperceptible by human eye) used for quantifying the same physical phenomena (i.e., visual masking).

93. All of these reasons would have given a POSA a reasonable expectation of success for using the JND generation method taught by *Gu* to compute the JND required by *Chiu*. The state of the art further supports this expectation. For example, *Jayant* discloses how JND was being incorporated into different forms of image and video coding around this time. *See generally, Jayant* at 1404, 1411-17.

3. [6.c]: estimating parameters of said details;

94. *Gu*'s JND generation includes the estimation of parameters. Therefore, *Chiu* in view of *Gu* and the knowledge of a POSA renders obvious this limitation.

95. The ’532 patent describes that the parameters “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.” ’532 patent, 7:10-17. The specification lists at least nine examples of parameters. *See id.*, 3:17-49. Such as “determining $N\Delta_i$, a per-pixel signal intensity change between a current frame and a previous frame, normalized by a maximum intensity” and “generating NY_i, a normalized intensity value per-pixel.” *See id.*, 3:17-49. The “[n]ormalized $N\Delta_i$” “measures the change Δ_i, per pixel i, from the current frame to its previous frame.” $N\Delta_i$ is a parameter measuring temporal change of the details; it is basically a difference frame. *Id.*, 7:31-34. “Y_i is the luminance level of a pixel in the current frame,” which tells us that NY_i is a parameter describing the intensity of a pixel in a frame. *Id.*, 7:56-57. None of these parameters “have to be calculated to great accuracy as they are used in combination to determine a per pixel, visual perception threshold THD.” *Id.*, 7:14-16. And the specification notes that “[a]t least some of the following parameters are determined,” implying that the ’532 patent has a broader definition of parameter than the nine members of the list. A POSA would have interpreted this to mean that other parameters may be used in the ’532 patent’s methods or that not all of the listed parameters must be used. *Id.*

96. While the calculations involved in *Gu*’s JND generation demonstrate spatial and temporal separation and analysis, how the resulting values are used teach

the estimation of parameters. The final values of mg(x,y) (luminance changes), bg(x,y) (average luminance), ild(x,y,n) (temporal changes), which are calculated on a per pixel basis, each describe information about the current frame just as the '532 patent teaches. *Gu*, 10:3-16; 11:29-36. These components are then used to compute the spatial and temporal masking visibility threshold components of the JND:

- f_1: error visibility threshold due to texture masking;
- f_2: error visibility threshold due to average background luminance;
- f_3: error visibility threshold due to motion or temporal masking.

See Gu, 10:3-11:55. *Gu* describes masking as "the ability of one signal to hinder the perception of another within a certain time or frequency range." *Id.*, 5:25-28.

97. f_1, f_2, and f_3 are all components that "describe the information content of the current frame" as dictated in the '532 patent. *See* '532 patent, 7:10-17; *Gu*, 10:3-11:55. f_1 quantifies "[t]he reduction in the visibility of stimuli due to the increase in spatial nonuniformity of the background luminance," also "known as texture masking." *Gu*, 5:61-63. It is estimated by combining mg(x,y), bg(x,y) values. *Id.*, 10:10-20; Equation (2). f_2 quantifies the human eye's sensitivity "to luminance contrast rather than absolute luminance values." *Id.*, 5:45-46. It is estimated using bg(x,y) and constants. *Id.*, 10:10-25; Equation (3). f_3 quantifies the "temporal masking" phenomena—"the losses of spatial resolution and magnitude resolution as an object moves in a scene." *Id.*, 6:8-12. In other words, f_3 accounts

for the visual masking "due to motion," i.e., due to changes between frames or temporal change. *Id*., 9:65-11:44. A POSA would have considered masking to be "information content of the current frame" because masking dictates how pictures or details within pictures are perceived. *See, e.g.*, *Netravali* at 269-78. Therefore, a POSA would have understood *Gu* to teach that f_1, f_2, and f_3 are the claimed "estimating parameters of said details" because they describe the content of the current frame.

98. Additionally, a POSA would have considered *Gu's* $(p(x,y,\mathrm{n}) - p(x,y,n-1))$ to be a parameter. $(p(x,y,\mathrm{n}) - p(x,y,n-1))$ is a difference frame used to calculate ild(x,y,n) and in turn f_3. *See Gu*, 11:34-43. Although not identical to $N\Delta_i$ in the '532 patent, $(p(x,y,\mathrm{n}) - p(x,y,n-1))$ is also a measure of a detail's temporal change much like $N\Delta_i$ and corresponds to "a per-pixel signal intensity change between a current frame and a previous frame." '532 patent, 3:20-21. It is also describing the information content of the frames much like $N\Delta_i$. Thus, a POSA would have considered $(p(x,y,\mathrm{n}) - p(x,y,n-1))$ a parameter as well.

99. A POSA would have therefore understood *Gu*'s JND generation to also include the estimation of parameters. This limitation would have been obvious over *Chiu* in view of *Gu* and the knowledge of a POSA for the same reasons discussed in Section VIII.D.2.c because the estimation of parameters is another part of the JND generation.

4. [6.d]: defining a visual perception threshold for each of said details in accordance with said estimated detail parameters;

100. Throughout this declaration, I have described how *Chiu* discloses defining a visual perception threshold, based on a JND, for each of said details. *Chiu*, 7:38-52; 10:17-38. I have further demonstrated that *Gu*'s JND generation involves parameters, such as f_1, f_2, and f_3. These disclosures would have motivated a POSA to combine, with a reasonable expectation of success, used *Gu*'s parameters f_1, f_2, and f_3 to generate the JND to be used in *Chiu*'s visual perception threshold. And therefore, this limitation would have been obvious over *Chiu* in view of *Gu* and the knowledge of a POSA.

101. As discussed in Section VIII.D.2, a POSA would have found it obvious to use *Gu*'s JND generation method to calculate the JND in *Chiu*. *Gu* discloses a spatio-temporal JND determined by equations (1) and (6):

$$JND_S(x, y)=\max\{f_1(bg(x, y), mg(x, y)),f_2(bg(x, y))\},$$

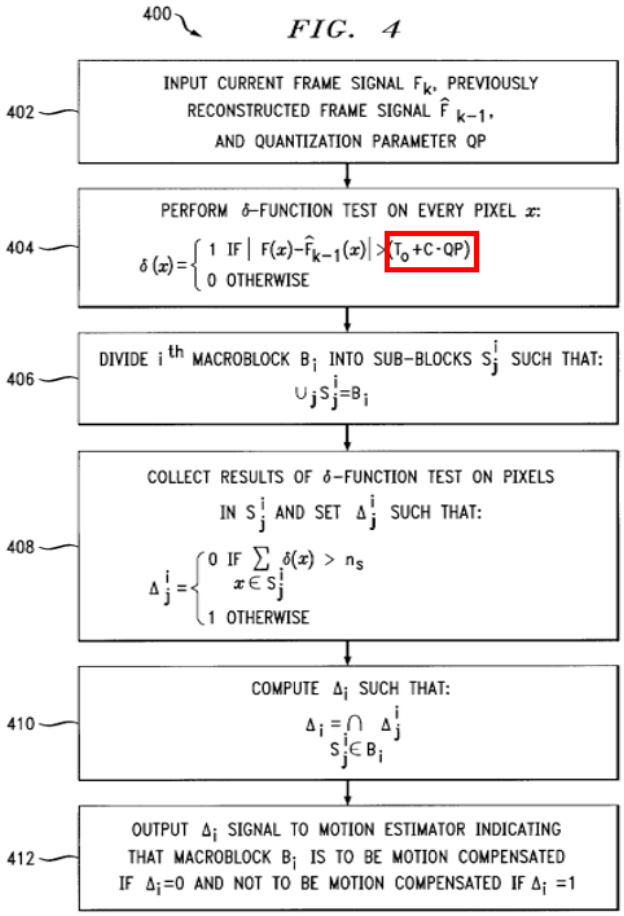
$$\text{for } 0\leqq x<W,0\leqq y<H \qquad (1)$$

$$JND_{S\text{-}T}(x,y,n)=f_3(ild(x,y,n))\cdot JND_S(x,y,n) \quad (6)$$

where f_1, f_2, and f_3 are estimated detail parameters as described in Section VIII.D.3. *Gu*, 9:65-11:55. A POSA would have understood that these equations demonstrate how *Gu*'s JND is based on estimated detail parameters, f_1, f_2, and f_3.

102. *Chiu*'s perceptual preprocessing method (400 in Figure 4) defines a

"visually perceptible threshold value T" for each pixel as $T = \mathrm{To} + \mathrm{C} * \mathrm{QP}$, where T_O is a JND, QP is a quantization parameter, and C is a constant. *Chiu*, 10:30-36. This method operates on a pixel-by-pixel basis. *Id.*, 7:31-53.



Chiu, Fig. 4 (annotated).

103. Because a POSA would have found it obvious to use *Gu*'s JND generation method to calculate the JND in *Chiu*'s visual perception threshold, a POSA would have understood that the resulting visual perception threshold would have been defined by *Gu*'s estimated detail parameters, f_1, f_2, and f_3. *See Chiu*,

10:23-34; *Gu*, 10:3-11:55. To illustrate this principle, *Chiu*'s thresholding equation, as modified by *Gu*, would read:

$$\delta(x) = \begin{cases} 1 & if\ |p(x,y,\mathrm{n}) - p(x,y,n-1)| > (JND_{S-T}(x,y,n) + C * QP) \\ 0 & otherwise \end{cases}.$$

104. Moreover, because both *Gu*'s JND and *Chiu*'s threshold are computed on a per pixel basis, they are both consistently calculated for each detail. *Chiu*, 7:14-52; *Gu*, 10:3-11:55. A POSA would understand this to be true because of how *Gu*'s JND generation conducts spatial and temporal separation and analysis to figure out the relationship between neighboring pixels within a frame and between frames. *Gu*, 10:3-11:55. Accordingly, *Chiu* in view of *Gu* and the knowledge of a POSA renders obvious "defining a visual perception threshold for each of said details in accordance with said estimated detail parameters," as claimed.

5. [6.e]: classifying said frame picture details into subclasses in accordance with said visual perception thresholds and said detail parameters; and

105. *Chiu* in view of *Gu* and the knowledge of a POSA renders this limitation obvious. Specifically, *Chiu*'s visual lossless encoding method uses the visually perceptible threshold value T, as improved by the proposed combination of by *Gu*'s JND and estimated parameters, to determine whether the details in a macroblock (i.e., pixels or groups of pixels perceivable by a human eye) are randomly distributed or not, and skips motion estimation encoding the macroblock

if the details are random. *Chiu,* 10:17-11:50; *Gu*, 9:50-11:55. A POSA would have found it obvious that the proposed combination would classify the details in the macroblocks into small, random details (subclass 1) and large, connected details (subclass 2).

106. ***Chiu*** in view of ***Gu*** and the knowledge of a POSA teaches classifying "details into subclasses in accordance with said visual perception thresholds and said detail parameters" The "preferred embodiment" of the '532 patent is a description of this classifying step. *See, e.g.*, '532 patent, 3:1-3:11; 3:65-4:4; 4:31-60 (Figs. 2, 3, & 11); 8:35-9:10. The '532 patent describes how its subclass determiner completes this classifying step. *See id.*, 4:31-60 (Figs. 2, 3, & 11); 8:35-9:10. The specification states that "[s]ubclass 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." *Id.*, 8:35-42. "'[S]ignificantly 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)." *Id.* The pixels are then assigned to a subclass, which are described as different types of details. *Id.* For example, a pixel that is not significantly present in the filtered frames belongs to a large detail or one that is "only detected but not distinguished." *Id.*, 8:43-9:13; Table 2. Or if the pixel is present in all of the filtered frames, then it belongs to a very small, single detail. *Id.*

107. The language of the "preferred embodiment" is very similar to the claims that concern classifying "details into subclasses in accordance with said visual perception thresholds and said detail parameters" and related limitations. *Compare id.*, 3:1-4:24 (preferred embodiment language) *with id.*, 12:50-14:34 (claim language). In particular, the preferred embodiment language is the same as claims 6 and 11 (which purports to narrow the classifying step).[3] *Id.*

108. Throughout the specification and in claims 1-5 and 11, the classifying step is based on the threshold and detail dimensions, rather than the parameters. *See, e.g.*, 2:36-41 ("the threshold levels and the detail dimensions"); 3:34-42 ("'significantly present" is defined by the threshold level and by the 'detail dimension'"); claim 1, 12:9-37 ("utilizing said threshold levels and said detail dimensions, for associating said pixels"). Based on the specification's use of parameters and detail dimensions, a POSA would understand them to be two separate components. *See id.*, claims 1, 5, and 6, 12:9-63. Therefore, a POSA would have understood classifying to occur "in accordance with said visual perception

[3] Claim 11 and the preferred embodiment identically narrow the classifying step as including "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." *See id.*, 3:1-4:4; claim 11.

thresholds as previously defined in accordance with the detail parameters."

109. *Chiu*'s perceptual preprocessing method in Figure 4 describes classifying details into at least two classes. *See Chiu*, Fig. 4. After the current frame F_k and the previously reconstructed picture $\hat{F}_{k-1}$ are received by the perceptual preprocessor 110 (step 402), *Chiu* discloses that "in step 404, a δ-function test is performed on every pixel x such that" the visually perceptible threshold value T, pixel values in the current frame F_k and in the previously reconstructed picture $\hat{F}_{k-1}$ are used to assign a value of "1" or "0" to each pixel:

$$\delta(x) = \begin{cases} 1 & if \ |F_k(\mathbf{x}) - \hat{F}_{k-1}(\mathbf{x})| > (T_0 + C * QP) \\ 0 & otherwise \end{cases}$$

Chiu, 10:23-34 (reproduced), where **x** refers to pixel location (x, y).

110. As altered by *Gu*, *Chiu*'s method discloses:

$$\delta(x) = \begin{cases} 1 & if \ |p(x, y, \mathrm{n}) - p(x, y, n-1)| > (JND_{S-T}(x, y, n) + C * QP) \\ 0 & otherwise \end{cases}$$

See Chiu, 10:23-34; *Gu*, 10:3-11:55. Recall that $JND_{S-T}(x, y, k)$ is a function of *Gu*'s estimated parameters, f_1, f_2, and f_3, meaning that the detail parameters are being relied on in this process. *Gu*, 10:3-11:55. A POSA would have understood *Chiu*'s $F_k(\mathbf{x}) - \hat{F}_{k-1}(\mathbf{x})$ to be equivalent to *Gu's* $p(x, y, \mathrm{n}) - p(x, y, n-1)$ because both represent a pixel of a current frame and a pixel of a previous frame. *See Netravali* at 7 (describing nomenclature).

111. For each pixel, the δ-function test compares the value of the difference

frame ($|p(x,y,\mathrm{n}) - p(x,y,n-1)|$) to the visually perceptible threshold value T ($JND_{S-T}(x,y,n) + C * QP$). *Chiu*, 9:19-21. If the value of the difference frame is above the threshold, it is assigned a 1; and if it is below the threshold, it is assigned a 0. *Id.*, 9:21-26. *Chiu* further discloses that "the number of surviving pixels" in a particular macroblock, i.e., the number of pixels assigned with the value 1, "are compared with the decision value n to determine whether [the] particular macroblock is to be encoded or not." *Id.*, 11:1-4. "However, the location of some of the surviving pixels can be random, that is, surviving pixels [within a particular macroblock] may be connected or isolated." *Id.*, 11:4-6. "Given this fact, . . . the picture discrepancy due to the elimination of isolated pixels, referred to as 'perceptual thinning,' is difficult for the HVS to detect." *Id.*, 11:6-9. "Thus, in order to take advantage of this," as shown in step 406 of the following annotated Figure 4, "each macroblock may be divided into" multiple sub-blocks, such as "four 4×4 sub-blocks." *Id.*, 11:9-15. "Then, in step 408, the results of the δ-function tests (step 404) are collected for the pixels in $S^i{}_j$ and a variable $\Delta^i{}_j$ associated with each sub-block is assigned a value as follows:

$$\Delta_j^i = \begin{cases} 0 & \text{if } \sum_{x \in S_j^i} \delta(x) > n_s \\ 1 & \text{otherwise} \end{cases} \qquad (10)$$

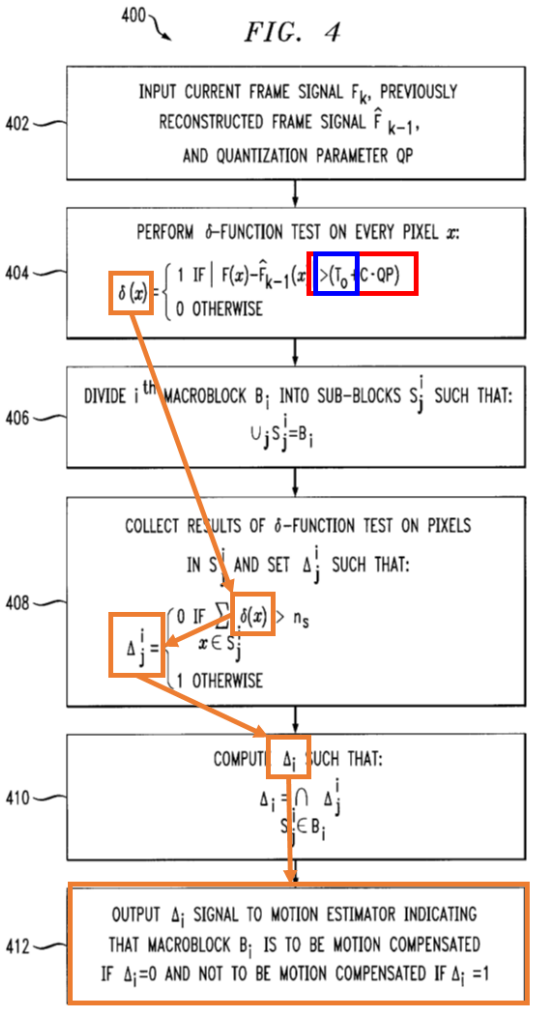
where n_s is a sub-block decision value." *Id.*, 11:15-25. "In step 410, the skip-

information-bit Δ_i for the macroblock is determined as follows:

$$\Delta_i = \bigcap_{x \in Bi} \Delta_j^i \qquad (11)$$

where ∩ is the logic 'AND' operator." *Id.*, 11:25-33. "Accordingly, if Δ_i is equal to zero (0) for a given macroblock, that macroblock is motion compensated." *Id.*, 11:33-35. "If Δ_i is equal to one (1) for a given macroblock, that macroblock is skipped." *Id.*, 11:35-36. "Lastly, in step 412, a Δ_i signal is sent to the motion estimator 112 instructing it whether or not the macroblock is to be motion compensated." *Id.*, 11:36-38.

112. The above process is illustrated in the following annotated Figure 4 of *Chiu*:



Id., Fig. 4 (annotated).

113. A POSA would have understood that steps 408 and 410 are where the classification occurs. As shown above, step 408 is where the thresholds (annotated red box), defined in accordance with the detail parameters (annotated blue box), are used to determine whether the number of surviving pixels (i.e., details) in each sub-block exceeds a decision value n_s—whether the surviving pixels (i.e., details) in each sub-block are likely to be connected. A POSA would have understood surviving

pixels to be likely connected when their combined value exceeds n_s, because n_s is a threshold value measuring how many of pixels in a subblock are visually noticeable. When the number of noticeable pixels in the subblock exceeds n_s, they collectively occupy a majority area of the subblock and thus are more likely to be connected; conversely, when number of noticeable pixels in the subblock exceeds n_s, they are scarce and are more likely to be scattered randomly in the subblock. Step 410 determines whether all of the sub-blocks in a particular macroblock have likely connected surviving pixels (i.e., details). As annotated in the orange boxes and as understood by a POSA, if all the sub-blocks in a particular macroblock have likely connected surviving pixels (i.e., details), the surviving pixels (i.e., details) in the macroblock form large, connected details; otherwise, the surviving pixels (i.e., details) in the macroblock are small, random details. *Chiu*, 11:15-25 ("the location of some of the surviving pixels can be random, that is, surviving pixels may be connected or isolated"), 12:7-11 ("Accordingly, the illustrative alternative embodiment of the invention enhances perceptual preprocessing by taking advantage of the inability of the HVS to distinguish the randomness of pixel locations in the difference frame in order to skip more macroblocks."). A POSA would have understood *Chiu* to disclose these categories because a POSA would have only wanted to perform motion compensation on large, connected details because these are the ones that the human visual system would perceive. A POSA

would not have wanted to motion compensate small, random details because they do not impact the overall picture quality.

114. As such, in *Chiu*'s method 400, as improved by *Gu* in the proposed combination, it would have been obvious to a POSA to use the pixel values and the visually perceptible threshold value T, defined by *Gu*'s JND and parameters, to classify the details in each macroblock into the following subclasses:

- Small, random details;
- Large, connected details.

Chiu also discloses that the "'block' as used herein is intended to include not only macroblocks as defined in the above-noted compression standards, but more generally any grouping of pixel elements in a video frame or field." *Chiu*, 5:61-65. Thus, a POSA would have understood that *Chiu*'s teachings regarding the classifying are not limited to the details in macroblocks, they but are applicable to any related groups of pixels (i.e., details) in a frame. Accordingly, *Chiu* in view of *Gu* and the knowledge of a POSA renders obvious "classifying said frame picture details into subclasses in accordance with said visual perception thresholds and said detail parameters," as claimed.

a. To the extent that claim 6 requires "parameters" to be separately used to classify details in combination with the thresholds, *Chiu* in view of *Gu* and the knowledge of a POSA also teaches this interpretation.

115. As noted above, *Chiu* as modified by *Gu* can be modeled by

$$\delta(x) = \begin{cases} 1 & if\ |p(x,y,\mathrm{n}) - p(x,y,n-1)| > (JND_{S-T}(x,y,k) + C * QP) \\ 0 & otherwise \end{cases}.$$

See Chiu, 10:17-11:50; *Gu*, 9:65-11:55. And as discussed in Section VIII.D.3, $p(x,y,\mathrm{n}) - p(x,y,n-1)$ would have been considered a parameter. A POSA would have therefore understood that *Chiu*'s method as modified by *Gu* involves a parameter ($p(x,y,\mathrm{n}) - p(x,y,n-1)$) and the threshold, defined by *Gu*'s JND.

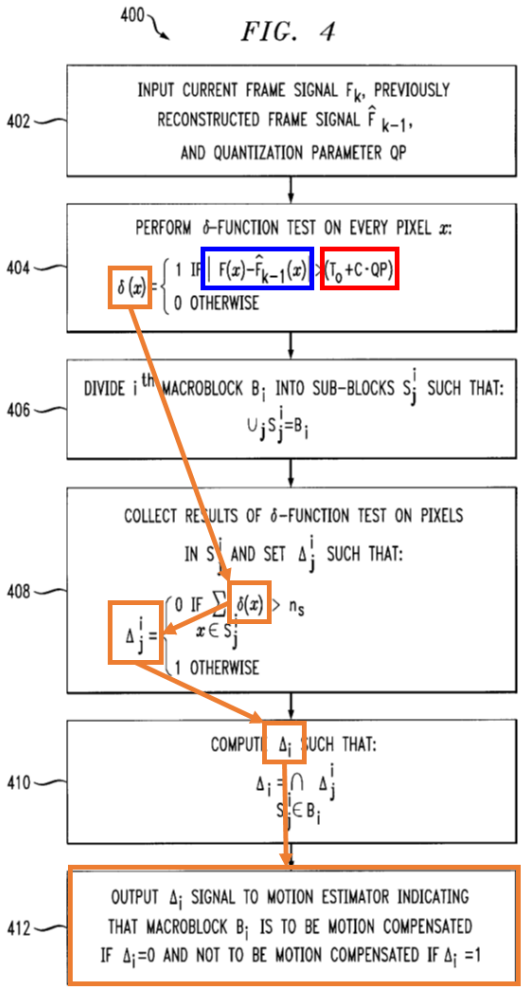
116. Although claim 6 uses "parameters" in the plural form, a POSA would have understood the term to only require "at least one parameter." Claim 6 mentions "parameters" three times within the claim—only in the plural form. Dependent claim 8, however, narrows the estimation step to "at least [one][4] of the following steps" and then lists NΔi and NYi as examples. The corresponding preferred embodiment within the specification also provides that only "at least one" parameter needs to be estimated. *See* '532 patent, claim 8; 3:17-49. Because claim 8 can estimate only one parameter, a POSA would have understood claim 1 to include this limitation as well. And because the estimation step informs the classifying step, a

[4] Claim 8 recites: "The method according to claim 6 and wherein the step of estimating comprises at least **cane** of the following steps." When this claim is recited in the specification as a preferred embodiment, it says "one" instead of "cane." *See* '532 patent, 3:17-49. Therefore, I understand claim 8 to have a typo.

POSA would have believed that only "at least one" parameter needs to be used in classifying too.

117. Grammatical consistency also supports the interpretation that "parameters" means "at least one parameter." For example, "estimating parameters" is being done for "details," plural. Parameters must be plural for the clause to be grammatically correct. Saying "estimating parameter of said details" would not be correct. Therefore, "parameters" means "at least one parameter."

118. To the extent that claim 6 requires "parameters" to be separately used to classify details in combination with the thresholds, *Chiu* in view of *Gu* teaches this limitation. A POSA would have understood *Chiu*'s classification step, as discussed in depth in Section VIII.D.5.a, to involve at least one parameter ($p(x, y, \mathrm{n}) - p(x, y, n-1)$) and a threshold, defined by JND. *See Chiu*, 10:17-11:50; *Gu*, 9:65-11:55. This can be modeled by *Chiu*'s annotated Figure 4.



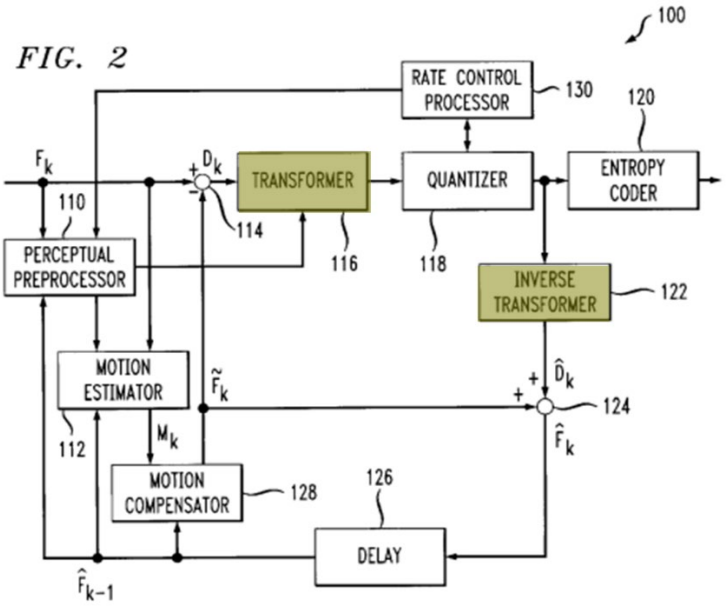
See Chiu, Fig. 4 (annotated).

119. Therefore, *Chiu*'s method as modified by *Gu* would have included a classifying step done "in accordance with said visual perception thresholds and said detail parameters." *See Chiu*, 10:17-11:50; *Gu*, 9:65-11:55.

6. [6.f]: transforming each said frame detail in accordance with its associate subclass.

120. *Chiu* in view of the knowledge of a POSA renders this limitation obvious.

121. *Chiu* discloses that the macroblocks with large, connected details are motion compensated, and macroblocks with small, random details are skipped for motion compensation. *Chiu*, 11:33-36. If a macroblock is skipped for motion compensation, a skip-information-bit is provided to the motion estimator 112, the transformer 116, and other affected components of the encoder 100 (i.e., quantizer 118, entropy encoder 120, etc.), to inform them that the particular macroblock is not to be processed thereby. *Id*., 11:36-47.



Id., Fig. 2 (annotated).

122. *Chiu* teaches it was well known in the art that the motion compensation involves transforming a macroblock. *Chiu* discloses that "[t]emporal correlation usually can be reduced significantly via forward, backward, or interpolative prediction based on motion compensation." *Chiu*, 1:53-56. "The remaining spatial

correlation in the temporal prediction error images can be reduced via transform coding." *Id.*, 1:56-57. One way this is accomplished is through the use of "the transformer 16 and the quantizer 18" of Figure 2. *Id.*, 3:7-8.

> The transformer may, for example, be a discrete cosine transform (DCT) generator which generates DCT coefficients for macroblocks of frames. The quantizer then quantizes these coefficients. . . . [which] generally has a substantial 15 effect on the resultant bit rate: a large QP performs coarse quantization, reducing the bit rate and the resulting video quality, which leads to a higher bit rate and higher resulting image quality.

Id., 3:8-19. *See also Jayant* at 1397-1417 (discussing different forms of transformations).

123. Moreover, as shown in annotated Figure 2 above, *Chiu* discloses that the motion compensated macroblocks are sent to the transformer 116 for transformation, while the skipped macroblocks are made equal to the corresponding macroblocks on the previously reconstructed picture $\hat{F}_{k-1}$, which are inverse transformed by inverse transformer 122. A POSA would understand that this means that every frame is transformed even if it is part of a subclass that is not motion compensated.

124. *Chiu* teaches that the subclasses of details in [6.e] are transformed as follows:

- Small, random details: skipped, use previously reconstructed picture $\hat{F}_{k-1}$ obtained by inverse transform;
- Large, connected details: motion compensated, transform current macroblock.

Chiu, 11:15-12:6. Therefore, *Chiu* in view of the knowledge of a POSA renders obvious "transforming each said frame detail in accordance with its associate subclass," as claimed.

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

125. Claim 16 is obvious over *Chiu* in view of *Gu* and the knowledge of a POSA.

126. Although "intensity" recited in this claim lacks an antecedent basis from claim 6, a POSA would have understood "intensity" to refer to "details." *See* '532 patent, 5:28-33 ("The present invention separates the details of a frame into . . . those which can be distinguished (for which at least three bits are needed to describe the intensity of each of their pixels"); claim 8, 13:1-31 (describing how detail parameter estimation takes place on a per pixel basis and can include, for example, "normalized intensity value per-pixel"). A POSA would also understand this to be true based on common knowledge at the time. For example, pixels, which make up details, were said to have a given intensity based on their brightness or color. *See, e.g.*, *Netravali* at 3, 5-10.

127. The obviousness of claim 6 over *Chiu* in view of *Gu* discussed above

applies here as well. And accordingly, it would have been obvious to a POSA to make the proposed combination. *See supra*, Sections VIII.C and VIII.D.2.c. *Chiu* provides that its thresholds are based on intensity values. *See Chiu*, 7:34-37. *Chiu* further teaches that for achromatic images, "the correct choice of the bush-hog threshold T for a particular pixel" can depend on "(i) average spatial and temporal background luminance level against which the pixel is presented; (ii) supra-threshold luminance changes adjacent in space and time to the test pixel; and (iii) spatio-temporal edge effects." *Chiu*, 7:38-44. Thus, *Chiu* discloses how its methods work when "intensity is a luminance value." *Gu* also discusses luminance in terms of intensity. For example, "[i]f movement is drastic, such as scene change, the perceived spatial and **intensity** resolution is significantly reduced immediately after the scene change. It was found that the eye is noticeably more sensitive to flicker at high **luminance** than at low luminance." *Gu*, 6:12-17. Thus, a POSA would have known luminance is a measure of light intensity.

128. Consequently, to the extent that intensity refers to the intensity of the details in claim 6, *Chiu* in view of *Gu* and the knowledge of a POSA renders claim 16 obvious.

IX. RIGHT TO SUPPLEMENT

129. I reserve the right to supplement my opinions in the future to respond to any arguments that the Patent Owner raises and to take into account new information as it becomes available to me.

X. JURAT

130. I 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.

Dated: 24 June 2022

By: 

Lina J. Karam, Ph.D.