UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

Unified Patents, LLC,
Petitioner,

v.

DigiMedia Tech, LLC,
Patent Owner.

Case IPR2022-01182
Patent 6,473,532

PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT 6,473,532

TABLE OF CONTENTS

LIST OF EXHIBITS

Exhibit	Description
Ex. 1001	U.S. Patent No. 6,473,532 to Sheraizin et al. ("the '532 patent")
Ex. 1002	Declaration of Dr. Lina Karam ("*Karam*")
Ex. 1003	Curriculum Vitae of Dr. Lina Karam
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	U.S. Provisional Application No. 60/136,633, filed on May 27, 1999 ("*Gu*'s provisional")
Ex. 1008	Arun N. Netravali, Barry G. Haskell, DIGITAL PICTURES: REPRESENTATION, COMPRESSION, AND STANDARDS, (2d ed., 1994) ("*Netravali*")
Ex. 1009	Reserved
Ex. 1010	Declaration of Kevin Jakel
Ex. 1011	Docket, *DigiMedia Tech, LLC v. Panasonic Corp.*, Case No. 1:20-cv-04995 (N.D.G.A.) (December 9, 2020)
Ex. 1012	Docket, *DigiMedia Tech, LLC v. ViacomCBS, Inc. et al.*, Case No. 1:21-cv-01831 (S.D.N.Y.) (March 2, 2021)
Ex. 1013	Docket, *DigiMedia Tech, LLC v. Lenovo US Inc. et al.*, Case No. 1:21-cv-00227 (D. Del.) (February 18, 2021)
Ex. 1014	Scheduling Order, *DigiMedia Tech, LLC v. Lenovo US Inc. et al.*, Case No. 1:21-cv-00227 (D. Del.) (issued September 28, 2021)
Ex. 1015	Excerpt of LegalMetric District Report: Delaware District Court in Patent Cases, January 2018-April 2022

Exhibit	Description
Ex. 1016	Scheduling Order, *DigiMedia Tech, LLC v. Lenovo US Inc. et al.*, Case No. 1:21-cv-00227 (D. Del.) (issued April 25, 2022)
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*")

I. Introduction

Unified Patents, LLC ("Petitioner") requests *inter partes* review of claims 6 and 16 of DigiMedia Tech, LLC's U.S. Patent No. 6,473,532 (Ex. 1001 ("the '532 patent)). Petitioner requests review as follows:

Ground No.	Obviousness
1	Claims 6 and 16 are obvious based on U.S. Patent No. 6,366,705 to Chiu et al. (Ex. 1005 - "*Chiu*") in view of U.S. Patent No. 7,006,568 by Gu et al. (Ex. 1006 - "*Gu*") and the knowledge of a person of ordinary skill in the art ("POSA").

II. State of The Art

A. Video Coding Overview: Motion Compensation and Transformation

A video is a series of pictures (i.e., frames) capturing visual information in time. To be stored and reproduced on a screen, pictures are divided into subparts called picture elements or pixels. *Netravali*, 5-7; *Karam*, ¶¶ 17-18. Pixels are made up of bits, which are spaces that represent binary data. *Netravali*, 5; *Karam*, ¶ 19. Whether a pixel is on or off, bright or dimmed, or blue or green, depends on how many bits are used. *Netravali*, 5; *Karam*, ¶¶ 17-19. Videos thus contain a lot of information. Because video files are large, they must be compressed to be stored efficiently or transmitted between computers. *See, e.g.*, *Gu*, 1:25-50; *Netravali*, 614; *Karam*, ¶ 20. The compression process is usually referred to as video encoding.

Karam, ¶¶ 20-22. Motion compensation and transformation are often used in video compression techniques. *Karam*, ¶¶ 20-22.

Motion compensation is the process of estimating or predicting motion (i.e., temporal change) between successive frames. *See Chiu*, 1:20-2:57; *Netravali*, 339-40; *Karam*, ¶ 22. 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*, 615-16; *Karam*, ¶ 22. Thus, only the motion information, rather than the later frame itself, needs to be stored or transmitted. *See Chiu*, 3:29-35; *Netravali*, 615-16; *Karam*, ¶ 22. 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*, 615-16; *Karam*, ¶ 22. This reduces the amount of information to be coded and stored. *See Chiu*, 1:20-2:57; *Netravali*, 339-40; *Karam*, ¶ 22.

Transforming signals also is a part of video coding that aims to reduce the amount of information that is coded. *See Chiu*, 1:45-62; *Netravali*, 175-76; *Karam*, ¶ 21. Transform coding specifically attempts to reduce redundancy within a frame (e.g., between pixels) rather than between frames. *See Chiu*, 1:45-62; *Netravali*, 175-76; *Karam*, ¶ 21. Transforming changes how the information within the frame is represented—usually converting the information from the spatial domain to the frequency domain. *See Netravali*, 18-21; *Karam*, ¶ 21.

B. Perceptual Coding

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; *Karam*, ¶ 23. For example, one pixel may be masked (i.e., rendered 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; *Karam*, ¶ 23. Also, a portion of a frame may be more or less noticeable because of motion (i.e., change in time) between frames. *Gu*, 5:44-65; *Karam*, ¶ 23.

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; *Karam*, ¶ 24. Perceptual coding is largely a method of lossless compression—a compression method that recodes videos using fewer bits based on what the human visual system perceives. *See Chiu*, 6:1-8:40; *Gu*, 6:35-67; *Karam*, ¶ 24. Perceptual coding 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*, 5:44-65; *Karam*, ¶ 24. Where the 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*, 7:38-50; *Gu*, 11:61-64; *Karam*, ¶¶ 24.

III. U.S. Patent 6,473,532

A. Overview

The '532 patent was filed on March 14, 2000, and issued on October 29, 2002. '532 patent, p. 1. Its earliest priority claim is to Israel Application No. 134182, which was filed on January 23, 2000.[1] *Id.* Titled, "METHOD AND APPARATUS FOR VISUAL LOSSLESS IMAGE SYNTACTIC ENCODING," the '532 patent generally relates to the "processing of video images and, in particular, to syntactic[2] encoding of images for later compression by standard compression techniques." *Id.*, 1:1-9.

The '532 patent describes and claims visual lossless encoders and methods of preprocessing frames as a form of visual lossless encoding. *Id.*, 2:25-28. The resulting pre-processed video sequence can be subsequently compressed by a video compression encoder to produce a reduced bit-rate for a fixed desired quality or to produce an improved quality at the same bit-rate as compared to using only a video compression encoder alone. *See, e.g.*, *id.*, 5:34-44. For example, "[a]n object of the

[1] Although not contested here, Petitioner does not waive its ability to contest the Patent Owner's priority claims.

[2] Syntactic is not explicitly defined in the '532 patent, but a POSA would have understood syntactic to mean syntax or rule-based coding. *Karam*, ¶ 25 (note 1).

present invention is to provide a method and apparatus for video compression which is generally lossless vis-a-vis what the human eye perceives." *Id.*, 2:22-24. Figure 2 generally describes "a video compression system having a visual lossless syntactic [VLS] encoder." *Id.* 4:31-32. "VLS syntactic encoder 20 transforms an incoming video signal" into a further compressed video signal. *Id.* 5:39-44.

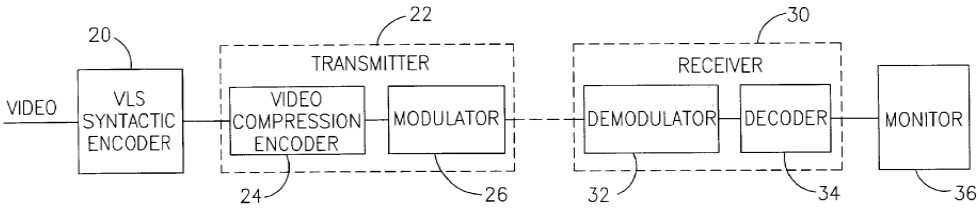


FIG.2

Figure 3 "is a block diagram illustration of the details of the visual lossless syntactic encoder of FIG. 2." *Id.*, 4:35-36; Fig. 3 (annotated below).

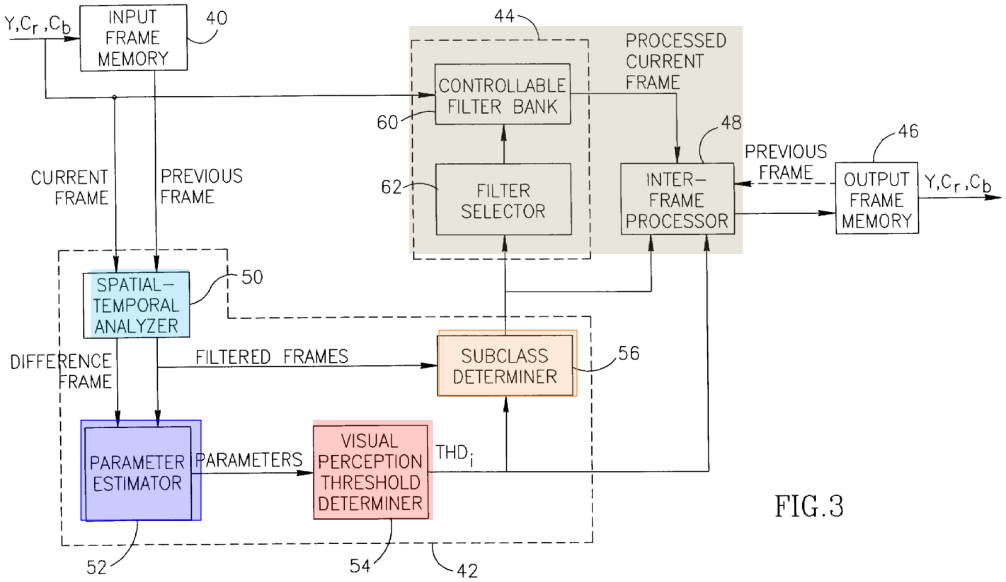


FIG.3

The '532 patent describes that "frames are composed of pixels, each having luminance Y and two chrominance C_r and C_b components, each of which is typically defined by eight bits. VLS encoder 20 generally separately processes the three components." *Id.*, 6:13-17. The "spatial-temporal analyzer 50 generates a plurality of filtered frames from the current frame, each filtered through a different high pass filter (HPF), where each high pass filter retains a different range of frequencies therein." *Id.*, 6:30-34. It also "generates difference frames between the current frame and another, earlier frame." *Id.*, 7:5-6. Then, the "[p]arameter estimator 52 takes the current frame and the filtered and difference frames and generates a set of parameters that describe the information content of the current frame. The parameters are determined on a pixel-by-pixel basis or on a per frame basis, as relevant." *Id.*, 7:10-17. "Visual perception threshold determiner 54 determines the visual perception threshold THD_i per pixel" using the estimated parameters. *Id.*, 8:26-27. The "[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 . . . . [it] then defines the subclass to which the pixel belongs." *Id.*, 8:35-42. Intra-frame processor 44 and inter-frame processor 48 then transforms the pixels based on their subclasses. *See id.*, 9:16-11:45.

The method of the encoding process is described in independent claim 6:

> A method of visual lossless encoding of frames of a video signal, the method comprising steps of:
>
> spatially and temporally separating and analyzing details of said frames;
>
> estimating parameters of said details;
>
> defining a visual perception threshold for each of said details in accordance with said estimated detail parameters;
>
> classifying said frame picture details into subclasses in accordance with said visual perception thresholds and said detail parameters; and
>
> transforming each said frame detail in accordance with its associate subclass.

The specification describes details as pixels or groups of pixels related based on human visual perception. *Id.*, 5:28-33 "The present invention separates the details of a frame into two different types, those that can only be detected (for which only one bit will suffice to describe each of their pixels) and those which can be distinguished (for which at least three bits are needed to describe the intensity of each of their pixels)." *Id.* The '532 patent further notes that "[t]he filter unit [also called the spatial-temporal analyzer] divides the video frame into portions having different detail dimensions." *Id.*, 2:35-36. For example, "filters HPF-R1 and HPF-C1 retain only very small details in the frame (of 1-4 pixels in size) while filter HPF-R3 retains much larger details (of up to 11 pixels)." *Id.*, 6:66-7:2. The specification 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."). It is not until the classifying step that the codex 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"). Thus, while "detail" is not explicitly defined, the term at least encompasses a pixel or a group of pixels that is detectable or distinguishable by the human visual system. *Karam*, ¶¶ 25-35; 73.

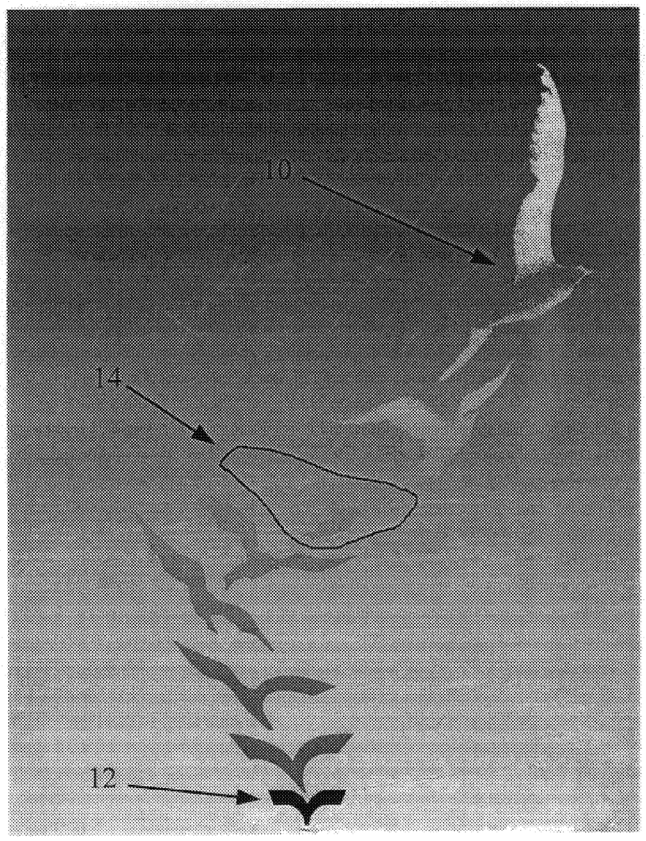


Fig. 1

B. Prosecution History

The '532 patent has 17 total claims, of which claims 1-5 are directed to a visual lossless encoder and claims 6-17 are directed to a method of visual lossless encoding. '532 patent, 12:10-14:34. The claims received a first-action allowance during prosecution. Ex. 1004, 88-95. The Notice of Allowance included a statement of the reasons for allowance that generally asserts the limitations recited in the issued independent claim 1 are not taught by the prior art on the record. *Id.*, 89. The

statement also asserted "[t]he prior art [] fails to teach the corresponding visual lossless encoding method [in claims 6-17]." *Id*.

IV. Level of Ordinary Skill in the Art

A person of ordinary skill in the art ("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. *Karam*, ¶¶ 36-37.[3] 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. *Id.*

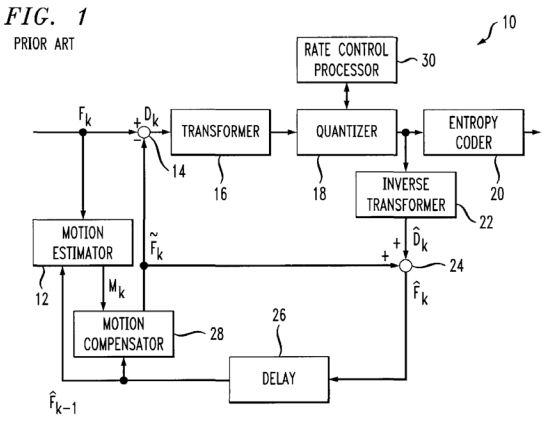
[3] Petitioner submits the declaration of Dr. Karam (Ex. 1002), an expert in the field of the '532 patent. Dr. Karam was a POSA of the '532 patent as of its earliest claimed priority date. *See Karam*, ¶¶ 2-6; Curriculum Vitae of Dr. Lina Karam (Ex. 1003).

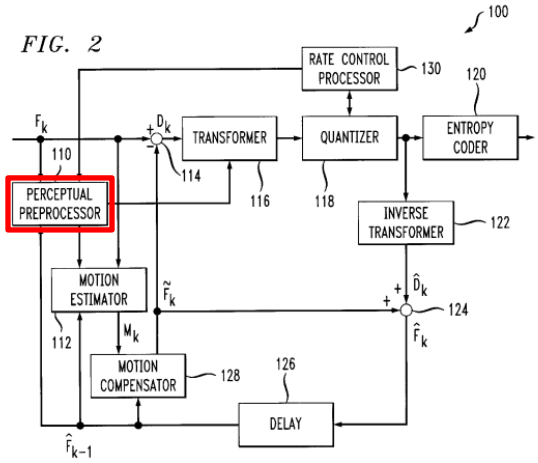
V. Ground 1: Claims 6 and 16 are obvious over *Chiu* in view of *Gu and the Knowledge of a POSA*

A. *Chiu*

Chiu was filed on January 28, 1999, and issued on April 2, 2002. The reference is prior art to the '532 patent under at least 35 U.S.C. § 102(e).

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.





This perceptual preprocessor "provides such advantages by making use of 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. *Chiu* describes methods of "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.*, 9:27-45.

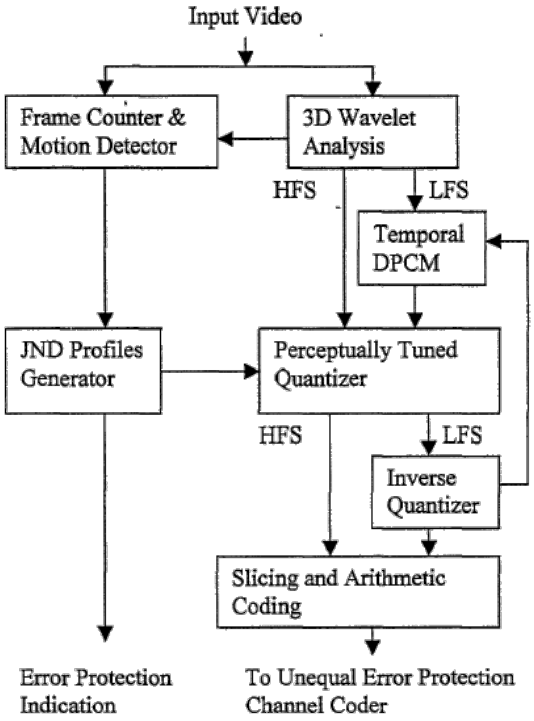
Chiu is analogous art to the '532 patent. *Karam*, ¶¶ 58-64; 87-93. The reference is from the same field of endeavor—visual lossless encoding—and at least reasonably pertinent to a problem addressed by the '532 patent—methods of video compression that are "lossless vis-a-vis what the human eye perceives." '532 patent, 2:23-24; *Karam*, ¶¶ 39-40; 58-64; 87-93.

B. Gu

Gu was filed as a PCT application on May 26, 2000. *Gu*, p. 1. It entered the US national stage on November 27, 2001. *Id.* *Gu* claims priority to U.S. Provisional Application No. 60/136,633 ("*Gu*'s Provisional"), which was filed on May 27, 1999. *Id.* As shown in the claim chart below, the *Gu*'s Provisional application provides written description and enablement support for at least claim 1 of *Gu*. Petitioner's expert also provides detailed mapping of *Gu*'s claim 1 to the *Gu*'s Provisional application. *Karam*, ¶¶ 41-57. Thus, *Gu* is prior art to the '532 patent under at least

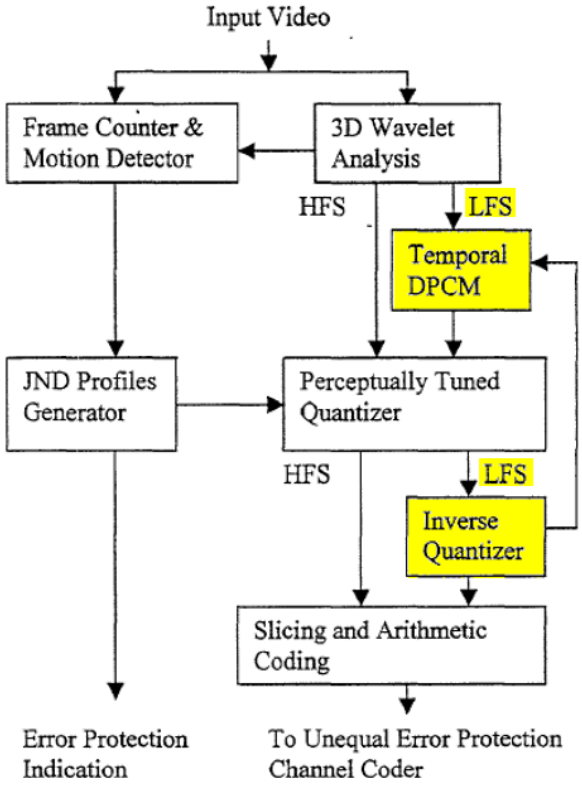
35 U.S.C. § 102(e). *See Dynamic Drinkware, LLC v. Nat'l Graphics, Inc.*, 800 F.3d 1375, 1378 (Fed. Cir. 2015).

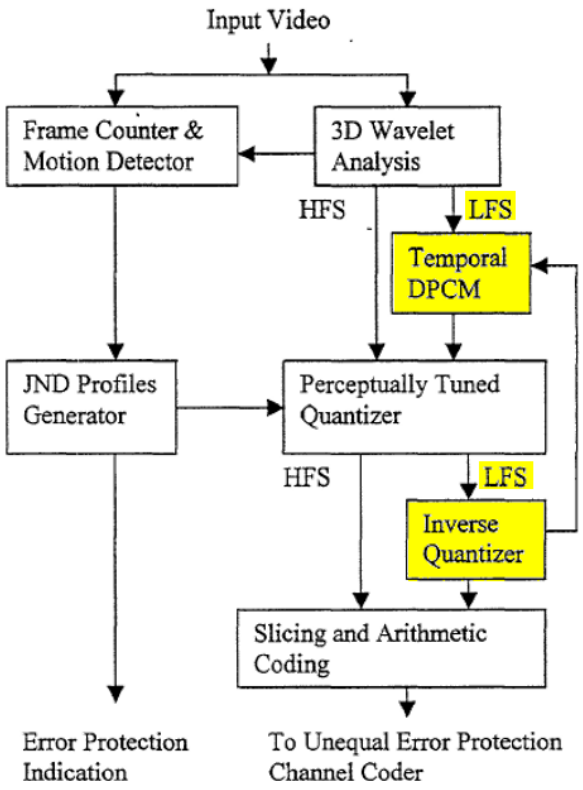
***Gu* Claim 1**	**Support from *Gu*'s Provisional**
1. A method, comprising:	Preamble
(a) arranging two successively received video frames into a set;	"The JND profile of a **video sequence** is a function of local signal properties, such as brightness, background texture, luminance changes **between two frames,** and frequency distribution." Ex. 1007, 39. "When **each pair of video frames is decomposed** into 11 spatio-temporal subbands as in Figure 2 (the details of such a 3D wavelet decomposition is described in Chapter 3), the relationship between the full-band JND profile and the component JND profiles can then be obtained by the following equations . . ." *Id.* at 52. **"Figure 3** and **Figure 4** show the JND model based video encoder and decoder respectively. In the video encoder, the input **video sequence** is decomposed into eleven spatio-temporal frequency subbands in the 3-D wavelet analysis module." *Id.* at 56.

***Gu* Claim 1**	**Support from *Gu*'s Provisional**
	 **Figure 3 JND Based Video Encoder** *Id.* at 57 (Fig. 3). "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. The details are as described in Chapter 2." *Id.* at 64; *Karam*, ¶ 47.
(b) decomposing said set into a plurality of high frequency and low frequency subbands;	"The signal is decomposed into frequency subbands and these subbands are encoded independently or dependently." Ex. 1007, 39. "We have implemented a video encoding/decoding scheme based on 3-

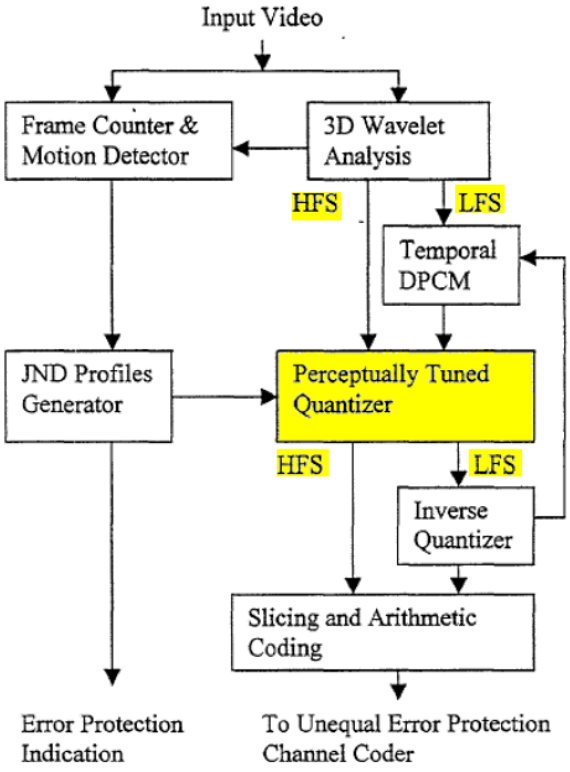
***Gu* Claim 1**	**Support from *Gu*'s Provisional**
	D wavelet decomposition and a human perceptual model." *Id.* at 42. "When each pair of video frames is decomposed into 11 spatio-temporal subbands as in Figure 2 (the details of such a 3D wavelet decomposition is described in Chapter 3), the relationship between the full-band JND profile and the component JND profiles can then be obtained by the following equations . . ." *Id.* at 52. "We use 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; *Karam*, ¶¶ 48-49.
(c) generating a human perceptual model for each of the plurality of subbands;	"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." Ex. 1007, 52. "The JND Model Generators estimate the spatio-temporal JND profile from analyzing local video signals and the distortion allocation algorithm that determines the JND profile for each subband." *Id.* at 56; *Karam*, ¶ 50.

***Gu* Claim 1**	**Support from *Gu*'s Provisional**
(d) encoding said low frequency subbands to produce encoded low frequency subbands;	"The signal is decomposed into frequency subbands and these subbands are encoded independently or dependently." Ex. 1007, 39. "Figure 3 and Figure 4 show the JND model based video encoder and decoder respectively. . . . **The spatial LLLL temporal L subband will be encoded by DPCM**." *Id.* at 56 (emphasis added). "We use 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, . . ." *Id.* at 62.

Gu Claim 1	Support from *Gu*'s Provisional
	 **Figure 3 JND Based Video Encoder** *Id.* at 57 (Fig. 3, annotated). "The reduction of irrelevancy is the result of amplitude quantization. In a signal compression algorithm, the inputs of the quantizing system are typically sequences of prediction errors or transform coefficients." *Id.* at 47. "The wavelet coefficients in the lowest frequency subband (LFS) are encoded with DPCM before quantization." *Id.* at 93; *Karam*, ¶ 51.

***Gu* Claim 1**	**Support from *Gu*'s Provisional**
(dd) inverse quantizing said encoded low frequency subbands:	"Figure 3 and Figure 4 show the JND model based video encoder and decoder respectively. In the video encoder, the input video sequence is decomposed into eleven spatio-temporal frequency subbands in the 3-D wavelet analysis module. . . . The Perceptually Tuned Quantizer implements quantization for the wavelet coefficients in each subbands according to their JND profiles. The spatial LLLL temporal L subband will be encoded by DPCM." Ex. 1007, 56.  **Figure 3 JND Based Video Encoder**

***Gu* Claim 1**	**Support from *Gu*'s Provisional**
	Id. at 57 (Fig. 3, annotated); *Karam*, ¶ 52.
(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;	"In video encoder, the input video sequence is decomposed into eleven spatio-temporal frequency subbands in 3-D wavelet analysis module. . . . The JND model generators estimate the spatio-temporal JND profile from analyzing local video signals, the distortion allocation algorithm determines the JND profile for each subband. The perceptually tuned quantizer quantizes the wavelet coefficients in each subband according to their JND profiles. . . . The data stream from each subband goes through the slicer which break the data into small packets to prevent the error propagation of arithmetic decoding before entropy coding." Ex. 1007,13. "In the video encoder, the input video sequence is decomposed into eleven spatio-temporal frequency subbands in the 3-D wavelet analysis module. . . . The JND Model Generators estimate the spatio-temporal JND profile from analyzing local video signals and the distortion allocation algorithm that determines the JND profile for each subband. The 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

***Gu* Claim 1**	**Support from *Gu*'s Provisional**
	slicing and entropy coding. Afterward we get compressed video signals in 11 bit streams." *Id.* at 56.  **Figure 3 JND Based Video Encoder** *Id.* at 57 (Fig. 3, annotated). "The 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 (emphasis added). "Therefore, one quantization table that leads to the uniform error energy

***Gu* Claim 1**	**Support from *Gu*'s Provisional**
	distribution over all subbands is set up for each subband. It is transmitted in the header of the bit stream for this subband. If after quantization the proportion of zero signals in a block is larger than 7/8 or 15/16, this block is assumed to be unimportant. Its stepsize is recorded as 0, and in the decoder, all values in this block are recovered as 0's. So the coefficients in such an unimportant block need not be transmitted." *Id.* at 66. "The index of levels for the optimum quantizer is transmitted in the header of the bit stream of this subband." *Id.* at 69; *Karam*, ¶ 53.
(f) redefining said generated bitstreams for each of said high frequency subbands to produce a plurality of redefined high frequency bitstreams;	"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. These 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." Ex. 1007, 56. "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

***Gu* Claim 1**	**Support from *Gu*'s Provisional**
	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 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 $\{\mathbf{b_1}, \mathbf{b_2}, \ldots \mathbf{b_M}\}$. " *Id.* at 92 (emphasis added) ; *Karam*, ¶ 54.
(g) redefining said generated bitstreams for each of said low frequency subbands to produce a plurality of redefined low frequency bitstreams;	"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. These 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." Ex. 1007, 56. "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

***Gu* Claim 1**	**Support from *Gu*'s Provisional**
	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 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 $\{\mathbf{b_1}, \mathbf{b_2}, \ldots \mathbf{b_M}\}$. " *Id.* at 92 (emphasis added) ; *Karam*, ¶¶ 54-55.
(h) channel coding said plurality of redefined high frequency bitstreams and said plurality of redefined low frequency bitstreams; and	"These 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." Ex. 1007, 56. "3.6 Perceptual Channel Coding

***Gu* Claim 1**	**Support from *Gu*'s Provisional**
	In 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; *Karam*, ¶ 56.
(i) repeating steps (a)–(h) for a next set until each of said received video frames have been compressed.	"For 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." Ex. 1007, 48. "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." *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

***Gu* Claim 1**	**Support from *Gu*'s Provisional**
	model is refreshed with the updated signal." *Id.* at 63; *Karam*, ¶ 57.

Gu "relates generally to the compression and encryption of data, and more particularly to compressing data with a human perceptual model." *Gu*, 1:22-24. Like the '532 patent, *Gu* identifies that "[c]ompeting with the need to compress data is the need to maintain the subjective visual quality of an image, arguably, the ultimate objective of video transmission systems. To this end, performance metrics that take the psycho visual properties of the human visual system (HVS) into account are needed." *Id.*, 1:37-42. It 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.

To analyze details of a frame before coding, *Gu* teaches how "[a] 3-D wavelet analyzer module is used to generate low and high frequency subbands" of frames and how "[a] just-noticeable distortion model generator is used to calculate a profile

of the video image based upon local signal properties." *Id.*, 2:25-28. 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. "For 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

Gu is analogous art to the '532 patent. *Karam*, ¶¶ 58-64; 87-93. The reference is from the same field of endeavor—visual lossless encoding—and at least reasonably pertinent to a problem addressed by the '532 patent—methods of video compression that are "lossless vis-à-vis what the human eye perceives." '532 patent, 2:23-24; *Karam*, ¶¶ 41-64; 87-93.

C. Independent claim 6

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

To the extent the preamble is limiting, it would have been obvious over *Chiu* in view of the knowledge of a POSA. *See Pacing Techs., LLC v. Garmin Intern., Inc.*, 778 F.3d 1021, 1023-1024 (Fed. Cir. 2015) (finding that preambles may be limiting), in certain circumstances); *Karam*, ¶¶ 65-70.

The '532 patent states that its object "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 particular, even if the video compression 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. Therefore, a POSA would have understood the claimed "method of visual lossless encoding of frames of a video signal" is lossless only with respect to eye distinguishable information, regardless of whether eye indistinguishable information is lost or not. *Karam*, ¶¶ 65-70.

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. *Chiu's* 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"). Therefore, a POSA would have understood that *Chiu*'s perceptual preprocessor is programmed to perform a video encoding method that does not cause any loss of information perceived, i.e., distinguishable, by a human eye. *Karam*, ¶¶ 65-70; 87-93. *Chiu* also discloses that the visual lossless video encoding method is performed on frames of a video signal. *Chiu*, 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."). 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. *Karam*, ¶¶ 58; 65-70; 87-93.

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

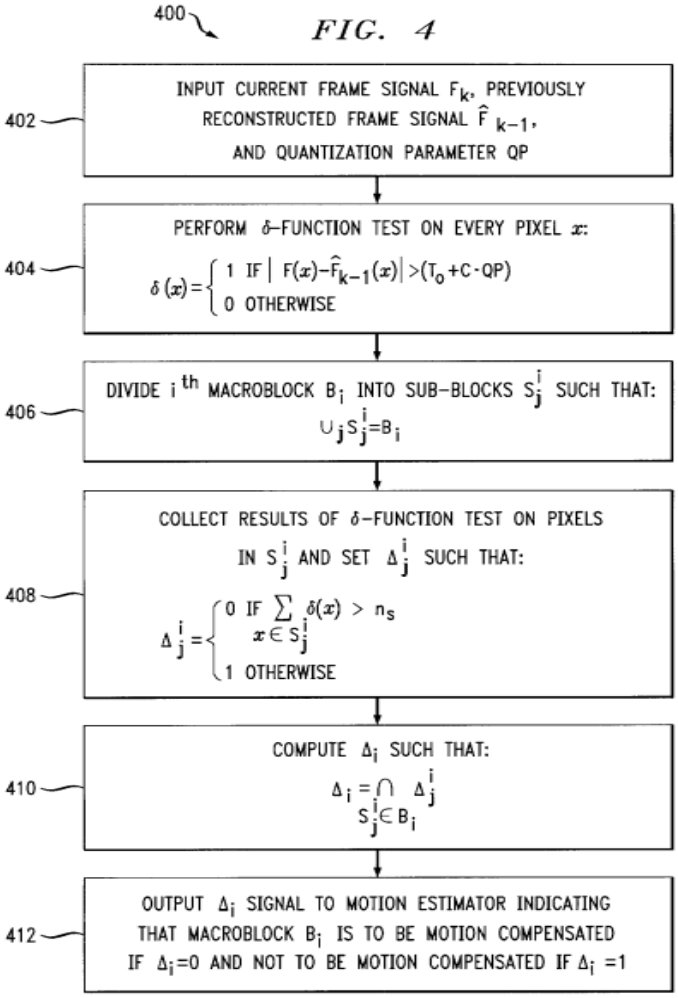
Chiu in view of *Gu* and the knowledge of a POSA teaches this limitation. *Karam*, ¶¶ 71-93. *Chiu* discloses a perceptual preprocessing method that uses a

visually perceptible threshold value to analyze the degree of visual perception of the details, where the threshold is modeled based on a just-noticeable-distortion (JND) value. *See Chiu*, 12:25-38; *Karam*, ¶¶ 73-75. *Gu* discloses that the JND value is generated by spatially and temporally separating and analyzing details. *Gu*, 9:50-11:55; *Karam*, ¶¶ 76-86. 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. *Karam*, ¶¶ 58-64; 87-93.

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

As described in Section III.A, "details" in the '532 patent claims at least encompasses a pixel or a group of pixels that is detectable or distinguishable by the human visual system

Chiu discloses a video encoding method used to process pixels or groups of pixels that are detectable or distinguishable by a human eye. *Chiu*, 10:17-11:50; *Karam*, ¶¶ 39-40; 74-75. For example, Figure 4 of *Chiu* illustrates a "perceptual preprocessing method 400:"


400
FIG. 4
402
INPUT CURRENT FRAME SIGNAL F_k, PREVIOUSLY RECONSTRUCTED FRAME SIGNAL $\hat{F}_{k-1}$, AND QUANTIZATION PARAMETER QP
404
PERFORM δ–FUNCTION TEST ON EVERY PIXEL x:
$\delta(x)=\begin{cases}1 \text{ IF } \left|F(x)-\hat{F}_{k-1}(x)\right|>(T_0+C\cdot QP)\\0 \text{ OTHERWISE}\end{cases}$
406
DIVIDE i^{th} MACROBLOCK B_i INTO SUB–BLOCKS S_j^i SUCH THAT:
$\cup_j S_j^i=B_i$
408
COLLECT RESULTS OF δ–FUNCTION TEST ON PIXELS IN S_j^i AND SET Δ_j^i SUCH THAT:
$\Delta_j^i=\begin{cases}0 \text{ IF } \sum_{x\in S_j^i}\delta(x) > n_s\\1 \text{ OTHERWISE}\end{cases}$
410
COMPUTE Δ_i SUCH THAT:
$\Delta_i=\bigcap_{S_j^i\in B_i}\Delta_j^i$
412
OUTPUT Δ_i SIGNAL TO MOTION ESTIMATOR INDICATING THAT MACROBLOCK B_i IS TO BE MOTION COMPENSATED IF Δ_i=0 AND NOT TO BE MOTION COMPENSATED IF Δ_i =1

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."[4] *Id*., 10:23-36.

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." *Id*., 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 perceive any impairment in the video if the discrepancies are below the JND value." *Id*., 7:19-22. In view of *Chiu*'s disclosure, a POSA would have understood that the JND is used

[4] Equation (9) in *Chiu* appears to have a typo by missing the constant "C." *Karam*, ¶ 74. The equation is correctly written in Figure 4, reproduced above. The typo does not impact the analysis in the present petition.

to measure whether a pixel is perceived, i.e., detectable or distinguishable, by a human eye. *Karam*, ¶ 75. 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 encode perceivable pixels, i.e., "details." *Karam*, ¶ 73-75.

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

Like *Chiu*, *Gu* 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. *Id.*, 9:50-11:55; Karam, ¶¶ 41-64; 76-93.

Gu discloses "[t]he JND model . . . provides each pixel with a threshold of error visibility, below which reconstruction errors are rendered imperceptible." *Gu*, 1:47-50. "The generation of the JND model consists of several steps." *Id*., 9:65-66.

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.

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

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

Id., 11:15-28.

Further, as the second step for generating the JND model, *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.

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." '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. Likewise, in the above steps for generating the JND model, *Gu* teaches the claimed "spatially and temporally separating . . . details of

said frames" by disclosing transforming the pixel information in each video frame n 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.

Gu, 9:50-11:55; *Karam*, ¶¶ 79-82. Thus, *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. *Karam*, ¶¶ 80-84.

Specifically, *Gu* teaches that mg(x,y) quantifies the "spatial nonuniformity of the background luminance," also known as local texture characteristics. *Gu*, 5:61-63; 10:13-15. *Gu* further teaches that mg(x,y) is computed using four directional high-pass filters "$G_k(I,j)$, 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."). "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. 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. *Karam*, ¶ 81. 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. *Karam*, ¶ 81. Therefore, *Gu* teaches that $G_k(I,j)$ filters pixel information associated with large spatial nonuniformity. *Karam*, ¶ 81.

Gu teaches that bg(x,y) quantifies the local "average value of background luminance" in the local region (i.e., neighborhood) associated with pixel (x,y). *Id*., 5:49; 10:15; 11:11-28. bg(x,y) "is calculated by a weighted lowpass [filter] operator B(I, j), I[sic], j=1 . . . 5." *Id*., 11:11-12. Therefore, bg(x,y) filters pixel information associated with small change in background luminance. *Karam*, ¶ 82.

Gu teaches that 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. 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. *Karam*, ¶ 83. Therefore, ild(x,y,n) is a highpass filer and filters pixel information associated with large temporal change. *Id*.

Accordingly, *Gu* teaches that 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 visual characteristics "due to motion," i.e., changes between frames or temporal change. *Gu*, 10:13-15; 10:34-11:28; 11:34-43; *Karam*, ¶¶ 81-84. 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. *Karam*, ¶¶ 76-84.

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. *Id*., ¶¶ 80-86. *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 $G_k(I, j)$ and weighted lowpass operator B(I, j), all in the form of matrices, are spatial filters

that analyze how one pixel relates to another. *Karam*, ¶¶ 81-86. Similar to the disclosures in *Gu*, the specification of the '532 patent discloses the use of filters, or operators, as an embodiment of spatial analysis. *See, e.g.*, '532 patent, 6:30-34 ("spatial-temporal analyzer 50 generates a plurality of filtered frames from the current frame"); *Karam*, ¶¶ 30, 81-86. Moreover, in generating the temporal component, *Gu* performs motion analysis by calculating "the average interframe luminance difference between the nth and (n−1)th frame." *Gu,* 11:34-43; *Karam*, ¶¶ 81-86; *compare to* '532 patent, 7:5-6 ("[spatial-temporal] analyzer 50 also generates difference frames between the current frame and another, earlier frame"). 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. *Karam*, ¶¶ 76-79; 81-86.

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. *Gu*, 9:50-11:55; *Karam*, ¶¶ 76-84. 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. *Gu*, 9:50-11:55; *Karam*, ¶¶ 76-86.

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

As discussed above, *Chiu* discloses a "visually perceptible threshold value T" for each pixel as $T = \mathrm{To} + \mathrm{C} * \mathrm{QP}$, where T_O is the JND, QP is the quantization parameter, and C is a constant. *Chiu*, 10:30-36. Although *Chiu* discloses that "T_O is the JND," it does not expressly provide details about to calculate the JND. *Id*. However, *Gu* discloses a JND model used for the same purpose as the JND in *Chiu*, namely "[to] provide[] each pixel with a threshold of error visibility, below which reconstruction errors are rendered imperceptible." *Gu*, 1:47-50. 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. *Karam*, ¶¶ 58-54, 71-93.

A POSA would have been motivated to make such combination because *Chiu* and *Gu* are directed to the same field of endeavor, namely video encoding. In particular, *Chiu* and *Gu* aim to solve the same problem, improving methods of video compression using perceptually lossless coding video encoding based human visual perception. *Karam*, ¶¶ 58-54, 87-93. For example, *Chiu* notes that "[a] main objective of a video codec is to represent the original signal with minimal bit rate

while maintaining acceptable picture quality, delay, and computational complexity." *Chiu*, 3:54-57. *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." *Gu*, 1:54-56. Another goal of *Chiu* is to reduce computational complexity and improve the compression ratio (i.e., reduce the bit rate) 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*, 4:22-29; 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. 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). 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. *Karam*, ¶¶ 58-54, 71-93.

Chiu and *Gu* also propose the same solutions to solve the above problems. 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.

Moreover, *Chiu* and *Gu* both disclose using the JND model to study the visual perceptibility and visual redundancy of the details. *Karam*, ¶¶ 90-91. *Chiu*'s method 400 uses the JND to determine the visually perceptible threshold value. *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."). 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* use the same methods (i.e., the JND model) to solve the same problem (i.e., using visually perceptible threshold to locate or identify perceptually important signals). *Karam*, ¶¶ 90-91. 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 the teachings of *Chiu*'s use of JND in per detail threshold determination with *Gu*'s teachings of how to calculate JND values for each detail. *Karam*, ¶¶ 90-91. Such a combination would have required minimal modifications to *Chiu*'s method because *Chiu*'s perceptual preprocessor is already the part of the video codec where *Chiu*'s detail thresholds are determined and calculated based on similar 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. *Karam*, ¶¶ 90-91.

A POSA would have also appreciated that combining the teachings of *Chiu* and *Gu* would yield a predictable result. *Karam*, ¶ 92. This is because *Chiu* and *Gu* both disclose similar JNDs. *Id*., ¶¶ 58-64; 87-92. For example, *Chiu* discloses "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. "The information for updating such pixels can be considered to be perceptual redundancy." *Id*., 7:22-24. The JND can be modeled "[i]n accordance with Weber-Fechner theory." *Id*., 7:14-16. And *Gu* discloses its JND model is used "[t]o remove . . . the irrelevancy of perceptually insignificant components from video signals." *Gu*, 7:2-6. "JND provides each signal to be coded with a visibility threshold of

distortion, below which reconstruction errors are rendered imperceptible." *Id*., 7:6-9. *Gu*'s JND model is also modeled after the Weber-Fechner theory. *Id*., 5:49-53. Indeed, both *Chiu* and *Gu* discuss the JND in terms of the spatial domain rather than the frequency domain. *See Chiu*, 6:64-7:37; *Gu*, 7:9-11. Given these teachings, a POSA would have appreciated that the JNDs described in *Chiu* and *Gu* are the same kind of component (i.e., a threshold level below which reconstruction error is imperceptible by the human eye) used for quantifying the same physical phenomena (i.e., visual masking). *Karam*, ¶¶ 64, 73-86. Therefore, the JND value calculated using *Gu*'s method is similar to the JND value used by *Chiu* for determining its visually perceptible threshold value. *Id*., ¶¶ 64, 73-93. Accordingly, a POSA would have had a reasonable expectation of success for using the JND generation method taught by *Gu* to compute the JND required by *Chiu*. *Id*., ¶¶ 64, 93.

In other words, this combination would have involved "the mere application of a known technique"—*Gu*'s JND generation—"to a piece of prior art"—*Chiu*'s method of encoding details using JNDs—"ready for the improvement." *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 417 (2007). And the result would have been a "predictable variation," as *Gu*'s method would merely have "improve[d] similar devices"—*Chiu*'s perceptual preprocessor—"in the same way." *Id.* There further would have been a reasonable expectation of success for this combination because *Gu*'s JND is similar to the JND used by *Chiu* for determining its visual perception

threshold, and adding JND models into video codecs would have been a routine combination for a POSA before the earliest priority date of the '532 patent. *Karam*, ¶ 58-64; 71-93.

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

Chiu in view of *Gu* and the knowledge of a POSA renders obvious this limitation because *Gu*'s JND generation includes the estimation of parameters. *Karam*, ¶¶ 94-99.

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. Examples of estimating parameters include: "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." *Id.*, 7:31-34. "Y_i is the luminance level of a pixel in the current frame." *Id.,* 7:56-57. In essence, while $N\Delta_i$ is a parameter measuring temporal change of the details, NY_i is a parameter describing the intensity of a pixel in a frame. "It is noted that the parameters do not have to be calculated to great accuracy as they are used in combination to determine a per pixel, visual perception threshold THD." *Id.*, 7:14-16. While these are examples of parameters, the otherwise broad

characterization indicates that the claim is not limited to only the parameters disclosed in the specification. *See id.*, 7:10-8:25. Indeed, the specification notes that "[a]t least some of the following parameters are determined," implying that other parameters may be used in the '532 patent's methods or that not all of the listed parameters must be estimated. *Id.*; *Karam*, ¶¶ 94-99; 116-117. The calculations involved in *Gu*'s JND generation demonstrate spatial and temporal separation and analysis, the resulting values teach the estimation of parameters. *Gu*, 10:3-16; 11:29-36. The final values of mg(x,y), bg(x,y), ild(x,y,n) components, which are calculated on a per pixel basis, each describe information about the current frame, as the '532 patent teaches. *Id.*, 10:3-16; 11:29-36. For example, "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. And "ild(x, y, n) denotes the average interframe luminance difference between the n th and (n−1) th frame." *Id*., 11:29-36.

Gu teaches that these components are then used to compute the spatial and temporal masking visibility threshold components of the JND, as described below:

- 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; *Karam*, ¶¶ 94-99; 116-117. "Here, masking refers to the ability of one signal to hinder the perception of another within a certain time or frequency range." *Gu*, 5:25-28.

Gu teaches that f_1, f_2, and f_3 are all components that "describe the information content of the current frame" as described in the '532 patent. *Gu*, 10:3-11:55; *see* '532 patent, 7:10-17. 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., changes between frames or temporal change. *Id.*, 9:65-11:44; *Karam*, ¶ 97. Therefore, a POSA would have considered that *Gu's* teachings regarding f_1, f_2, and f_3 teach the claimed "estimating parameters

of said details" because they describe the content of the current frame. *Karam*, ¶¶ 96-97.

A POSA would have also considered *Gu* to disclose other parameters as well. *Karam*, ¶¶ 96-98. For example, *Gu*'s $(p(x,y,n) - p(x,y,n-1))$ would have been considered a parameter. $(p(x,y,n) - p(x,y,n-1))$ 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, $(p(x,y,n) - p(x,y,n-1))$ 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 also is describing the information content of the frames much like $N\Delta_i$. *Karam*, ¶¶ 97-98. Thus, a POSA would have considered $(p(x,y,n) - p(x,y,n-1))$ a parameter as well. *Karam*, ¶ 98. Because *Gu*'s JND generation method includes the estimation of parameters, a POSA would have found it obvious to use *Gu*'s parameters to calculate the JND when used in *Chiu*'s method in accordance with the proposed combination. *Karam*, ¶¶ 94-99. This combination is obvious for the same reasons discussed in Section V.C.2. *Karam*, ¶¶ 58-64, 71-93.

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

Chiu in view of *Gu* and the knowledge of a POSA renders obvious this limitation. *Karam*, ¶¶ 100-104. *Chiu* discloses defining a visual perception

threshold for each of said details in accordance with a JND value. *Chiu*, 7:38-52; 10:17-38. *Gu* discloses that the JND value is generated using f_1, f_2, and f_3 ("estimated detail parameters"). *Gu*, 10:3-11:55. A POSA would have been motivated to combine the teachings of *Chiu* and *Gu*, to use f_1, f_2, and f_3 to generate the JND as taught in *Gu*, which a POSA would have found obvious, in accordance with the proposed combination, to further use in *Chiu's* method to generate *Chiu*'s visual perception threshold. *Karam*, ¶¶ 58-64, 71-93, 100-104.

As discussed in Section V.C.2 regarding element [6.b], a POSA would have found it obvious to use *Gu*'s JND generation method to calculate the JND in *Chiu*. *Karam*, ¶¶ 58-64, 73-93. *Gu* discloses a spatio-temporal JND determined by:

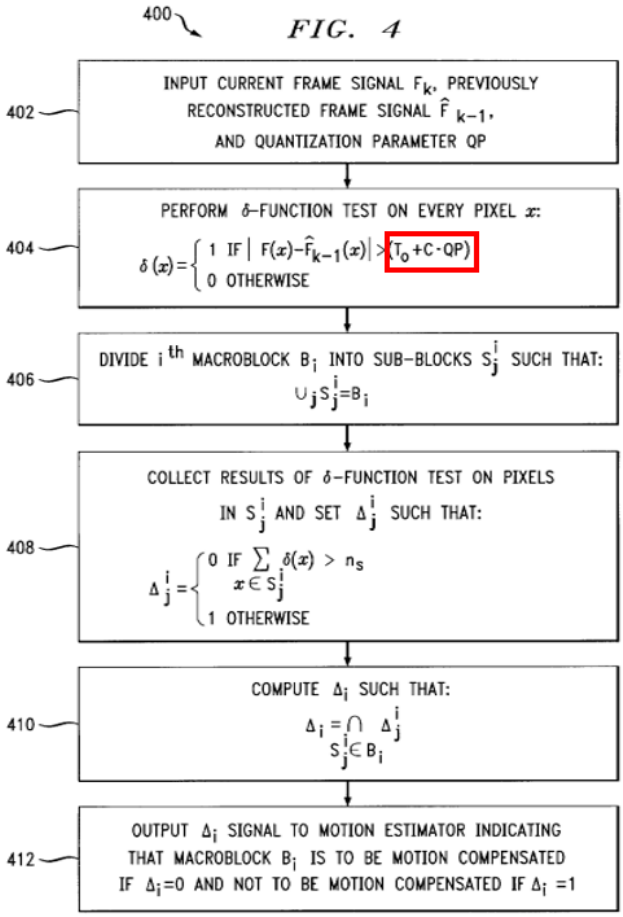
$$JND_{S\text{-}T}(x,y,n)=f_3(ild(x,y,n))\cdot JND_S(x,y,n) \quad (6)$$

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

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

where f_1, f_2, and f_3 are estimated detail parameters as described in Section V.C.3. *Gu*, 9:65-11:55. These equations demonstrate that *Gu*'s JND is based on estimated detail parameters, f_1, f_2, and f_3.

Chiu's perceptual preprocessing method (400 in Fig. 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.



Chiu, Fig. 4 (annotated).

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. *Karam*; ¶¶ 100-104. *Chiu*'s thresholding equation, as modified by *Gu*, would read:

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

See Chiu, 10:23-34; *Gu*, 10:3-11:55; *Karam*, ¶¶ 109-10.

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; *see, e.g.*, *Karam*, ¶ 104. 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. *Karam*, ¶¶ 100-104. This combination is obvious for the same reasons discussed in Section V.C.2. *Karam*, ¶¶ 58-64, 71-93.

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

Chiu in view of *Gu* and the knowledge of a POSA renders this limitation obvious. *Karam*, ¶¶ 105-19. Specifically, *Chiu*'s visual lossless encoding method uses the visually perceptible threshold value T, as improved by the proposed combination of *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. *See Chiu,* 10:17-11:50; *Gu*, 9:50-11:55; *Karam*, ¶¶ 105-19. In the proposed combination, the improved process classifies the details in the macroblocks into small, random details (subclass 1) and large, connected details (subclass 2). *Karam*, ¶¶ 109-19.

a. *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 claim 6 requirement of classifying "details into subclasses in accordance with said visual perception thresholds and said detail parameters" is described in the preferred embodiment of the '532 patent specification. *See, e.g.*, '532 patent, 3:1-3:11; 3:65-4:4; 4:31-60 (Figs. 2, 3, and 11); 8:35-9:10. The '532 patent's specification describes how its subclass determiner completes this classifying step in its preferred embodiment. *See id.*, 4:31-60 (Figs. 2, 3, and 11); 8:35-9:10. The '532 patent's 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." '532 patent, 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. *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 the filtered frames, then it belongs to a very small single detail. *Id.*

The language of the "preferred embodiment" is very similar to the recitation in the claims of 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 narrows the classifying step).[5] *Id.*

The '532 patent does not otherwise describe what the classification step entails when it is "in accordance with said visual perception thresholds ***and*** said detail parameters." Indeed, throughout the specification and in claims 1-5 and 11, the classifying step is based only on the threshold and detail dimensions. *See, e.g.*, *id.*, 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"). And the claims use detail dimensions and parameters differently. *See id.*, claims 1, 5, and 6, 12:9-63.

[5] 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.

Because the '532 patent describes how to classify details into subclasses using thresholds and detail dimensions, a POSA would have understood classifying "in accordance with said visual perception thresholds and said detail parameters" to mean at least classifying in accordance with "said visual perception thresholds as previously defined in accordance with the detail parameters." *See Karam*, ¶¶ 106-14.

Chiu's perceptual preprocessing method in Figure 4 describes classifying details. *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). In the proposed combination, *Chiu*'s method as improved by *Gu's* teachings provides:

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

See Chiu, 10:23-34; *Gu*, 10:3-11:55; *Karam*, ¶¶ 109-10. 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. *See Gu*, 10:3-11:55; *Karam*, ¶¶ 96-98; 100-04. 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 difference frames. *Karam*, ¶¶ 98-99. For each pixel, the δ-function test compares the value of the difference frame (which is calculated using the pixel values of the current frame F_k and the previously reconstructed picture $\hat{F}_{k-1}$) to the visually perceptible threshold value T. *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 Δ^i_j associated with each sub-block is assigned a value as follows:

$$\Delta^i_j = \begin{cases} 0 & \text{if } \sum_{x \in S^i_j} \delta(x) > n_s \\ 1 & \text{otherwise} \end{cases} \quad (10)$$

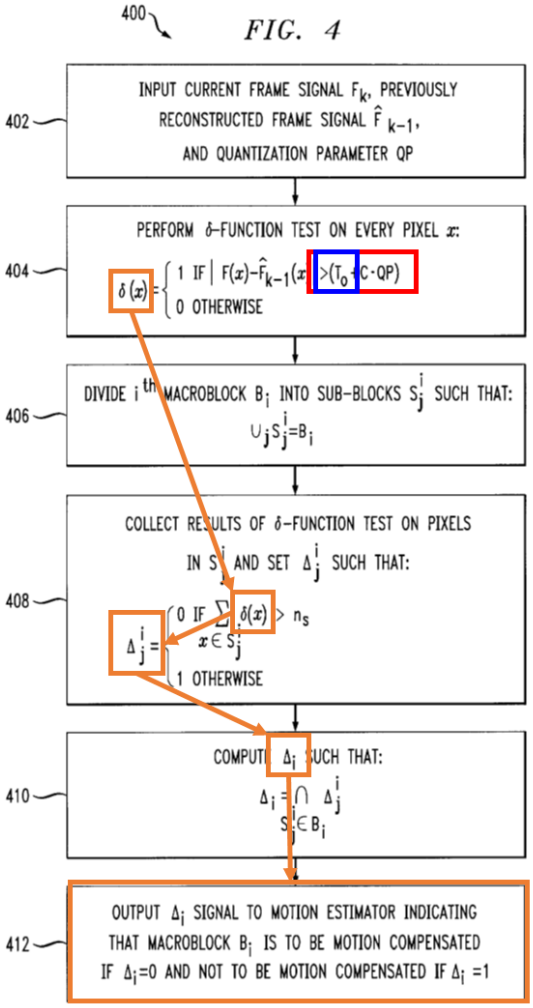
where n_s is a sub-block decision value." *Id.*, 11:15-25. "In step 410, the skip-information-bit Δ_i for the macroblock is determined as follows:

$$\Delta_i = \bigcap_{x \in Bi} \Delta^i_j \quad (11)$$

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.

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



Id., Fig. 4 (annotated); *Karam*, ¶¶ 106-14. As shown, the pixel values and the visually perceptible threshold value T (annotated in the red box), defined in accordance with the detail parameters (annotated in the blue box), are used in steps 408 through 412 to determine whether to skip motion compensation. *Karam*,

¶¶ 106-14. Among them, step 408 uses the thresholds, defined in accordance with the detail parameters, 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. *Karam*, ¶¶ 112-14. Step 410 determines whether all the sub-blocks in a particular macroblock likely have connected surviving pixels (i.e., details). *Karam*, ¶¶ 112-14. As understood by a POSA, if all the sub-blocks in a particular macroblock likely have connected surviving pixels (i.e., details), the surviving pixels (i.e., details) in the macroblock likely form large, connected details (annotated in the orange boxes); otherwise, the surviving pixels (i.e., details) in the macroblock are small, random details (annotated in the orange boxes). *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."); *Karam*, ¶¶ 105-14.

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:

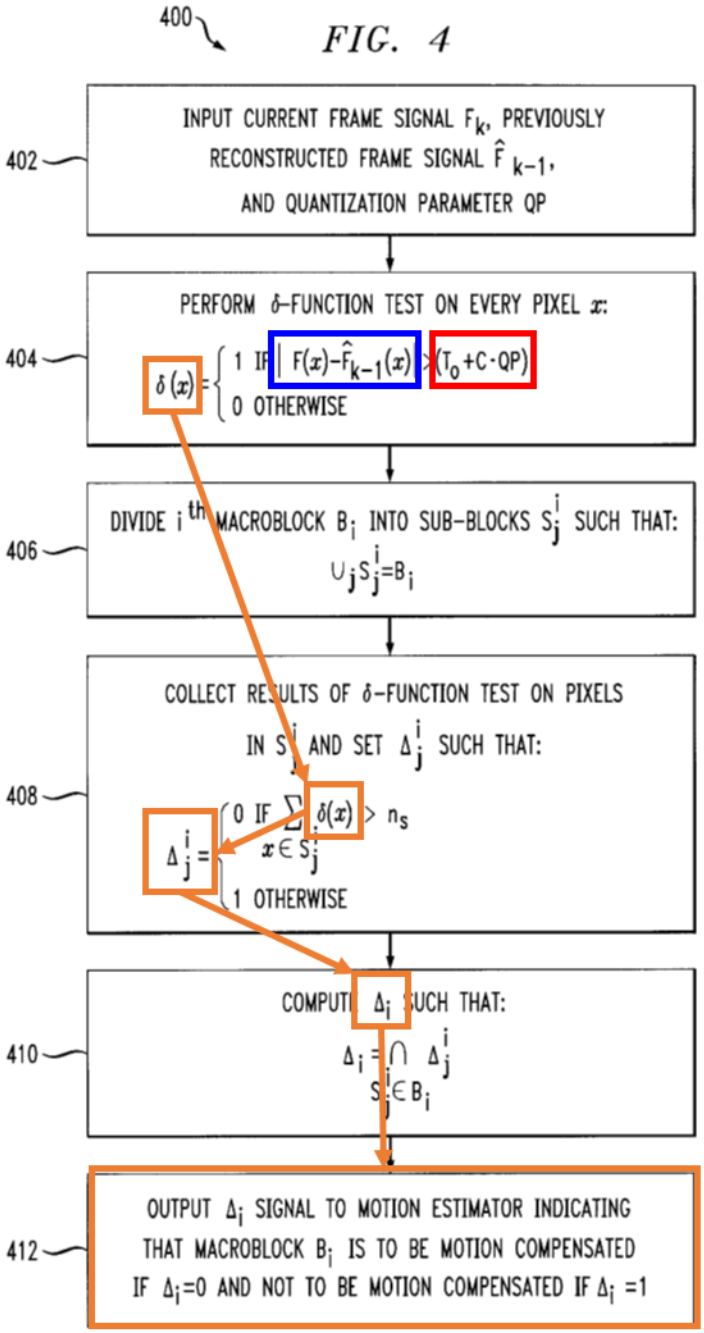
- Small, random details;
- Large, connected details.

Karam, ¶¶105-14. *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 but are applicable to any related groups of pixels (i.e., details) in a frame. *See Karam*, ¶ 114. 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. *Karam*, ¶¶ 105-14.

b. 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.

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; *Karam*, ¶ 115. And as discussed in Section V.C.3, $p(x,y,\mathrm{n}) - p(x,y,n-1)$ would have been considered a parameter. 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. *See Chiu*, 10:17-11:50; *Gu*, 9:65-11:55; *Karam*, ¶¶ 115-119. Based on the structure of the '532 patent's claims, a POSA would have understood parameters in claim 6 to include "at least one parameter." *Karam*, ¶ 116. A POSA would have understood *Chiu*'s classification step, as discussed in depth in Section V.C.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; *Karam*, ¶¶ 105-19. Similar to the model shown above, this can be modeled by *Chiu*'s annotated Figure 4.



See Chiu, Fig. 4 (annotated).

Despite being recited in the plural form, the term "parameters" does not require more than one parameter. *Karam*, ¶¶ 115-19. Dependent claim 8 narrows the estimation step to "at least [one][6] of the following steps" and then lists NΔi and NYi as examples. Because claim 8 can estimate only one parameter, claim 1 must be interpreted in a consistent manner. *See, e.g.*, *Wright Med. Tech., Inc. v. Osteonics Corp.*, 122 F.3d 1440, 1445 (Fed. Cir. 1997) (holding that an independent claim must not be interpreted "in a way that is inconsistent with a claim which depends from it"). Although plural terms are usually presumed to mean to two or more, this presumption "can be overcome when the broader context shows a different meaning applies." *See Apple Inc. v. MPH Techs. Oy*, 28 F.4th 254, 262 (Fed. Cir. 2022) (citing *Versa Corp. v. Ag-Bag Int'l Ltd.*, 392 F.3d 1325, 1330 (Fed. Cir. 2004)). In accordance with claim 8 and the corresponding preferred embodiment within the specification, only "at least one" parameter needs to be estimated and only "at least one" parameter needs to be used in classifying. *See* '532 patent, claim 8; 3:17-49.

[6] Claim 8 recites "The method according to claim 6 and wherein the step of estimating comprises at least **cane** of the following steps." Based on the specification of the '532 patent, Petitioner assumes that "cane" is a typo and is supposed to say "one."

Therefore, for "parameters" to be interpreted consistently between claims 6 and 8, the term must mean "at least one parameter." *Karam*, ¶¶ 115-19.

The use of "parameters" in claim 6 is also distinguishable from "plurality of picture parameters" in claim 5. Claim 5 recites "means for generating a plurality of picture parameters describing at least one of the following parameters." *See id.*, claim 5. The ordinary definition of "plurality of" is "two or more." *See, e.g.*, *Dayco Prod., Inc. v. Total Containment, Inc.*, 258 F.3d 1317, 1328 (Fed. Cir. 2001) (holding a plural term to include one when the patentees used "plurality of" elsewhere in the claims). Here, as in *Dayco*, Patentee used both a "plurality of" and the plural form of parameters. The different modifications of "parameters" suggests different meanings of the term. If Patentee wanted "parameters" to exclude one, then it would have used "plurality of" as it did in claim 5.

Grammatical consistency also supports the interpretation that "parameters" means "at least one parameter." *See In re Omeprazole Pat. Litig.*, 84 F. App'x 76, 80 (Fed. Cir. 2003) ("the plural term does not necessarily require multiple compositions, but instead reflects an attempt to achieve grammatical consistency."); *Karam*, ¶ 117. 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. As drafted, "parameters of said details" may include "multiple parameters for each detail," but it does not

exclude "a parameter for each detail." Therefore, "parameters" means "at least one parameter."

Accordingly, 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. *See Chiu*, 10:17-11:50; *Gu*, 9:65-11:55; *Karam*, ¶¶ 115-19.

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

Chiu in view of the knowledge of a POSA renders this limitation obvious. *Karam*, ¶¶ 120-24.

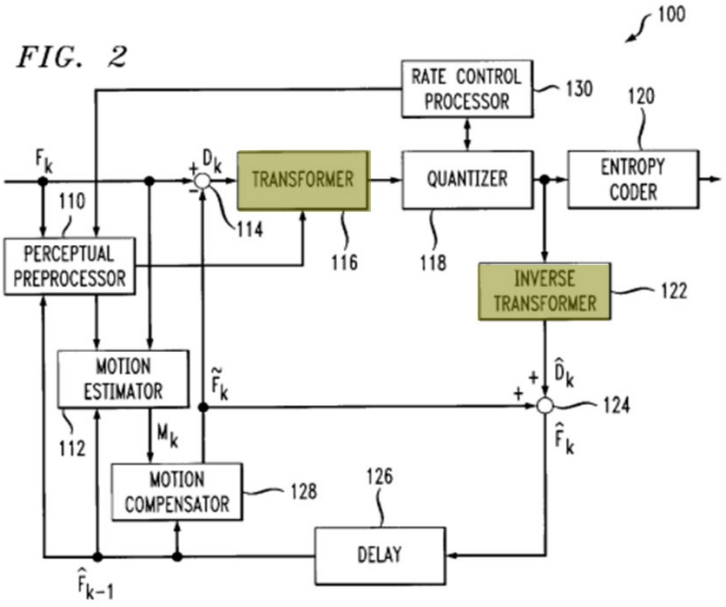
With regard to step 410 in the perceptual preprocessing method 400, *Chiu* discloses that the macroblocks with large, connected details are motion compensated, and macroblocks with small, random details are skipped for motion compensation. *Chiu*, 11:33-36. If a macroblock is skipped for motion compensation, 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; Fig. 4 (annotated); *Karam*, ¶ 121.

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 1. *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. *Karam*, ¶ 122.

Moreover, as shown in the following annotated Figure 2, *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.



Id., Fig. 2 (annotated); *Karam*, ¶¶ 121-24.

In sum, *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, 3:7-8; 11:15-12:6. *Karam*, ¶ 124. 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. *Karam*, ¶¶ 120-24.

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

Claim 16 is obvious over *Chiu* in view of *Gu* and the knowledge of a POSA. *Karam*, ¶¶ 125-28. As an initial matter, "intensity" recited in this claim lacks an

antecedent basis. The term does not appear previously in claim 6.[7] However, to the extent "intensity" refers to the intensity of "details" recited in claim 6, the claim element is rendered obvious in view of the proposed combination. *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"); *Fitbit, Inc. v. Valencell, Inc.*, 964 F.3d 1112, 1119-20 (Fed. Cir. 2020) (holding that the Board should consider the patentability of challenged claims lacking an antecedent when a correction is not subject to a reasonable debate based on claim language, the specification, and prosecution history).

The recitation in claims 16 of "wherein said intensity is a luminance value" would have been obvious to a POSA for similar reasons to those recited above for claim 6. *See supra*, Section V.C.2.c. *Chiu* provides that its thresholds are based on intensity values. *See Chiu*, 7:34-37. *Chiu* further teaches that for achromatic images,

[7] A reference to "intensity" can be found in claim 8, which also depends from claim 6; however, claims 8 and 16 are not otherwise related such that there is an antecedent basis for the term.

"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." *Id.*, 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. Based on these disclosures, and the general knowledge of a POSA, a POSA would have known that luminance is a measure of light intensity. *Karam*, ¶¶ 127-28.

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. *Karam*, ¶¶ 126-28.

VI. Non-Institution Under 35 U.S.C. §§ 314(a) or 325(d) Would Be Improper

A. 35 U.S.C. § 314(a) Does Not Favor Denial

The '532 patent was asserted in three district court cases listed in Section VII.B (Related Matters). Under the factors articulated in *Apple Inc. v. Fintiv, Inc.*, IPR2020-00019, Paper 11 (PTAB Mar. 20, 2020) (precedential), non-institution in

light of the cases would be improper, because at least five of the *Fintiv* factors weigh against discretionary denial and only one is neutral.

Factor 1 favors institution or is at least neutral, as Unified is not a participant to any proceedings involving the '532 patent. This indicates that no stays would be necessary. It is speculative whether a district court would grant a stay, and it is likely a stay would be granted in any future proceedings involving the '532 Patent that are filed after the Institution Decision. *Apple Inc. v. Fintiv, Inc.*, IPR2020-00019, Paper 15, at 12 (PTAB May 13, 2020); *Western Digital Corp. v. Martin Kuster*, IPR2020-01391, Paper 10, at 8-9 (PTAB Feb. 16, 2021); *Dish Network v. Broadband iTV, Inc.*, IPR2020-01280, Paper 17, at 12-14 (PTAB Jan. 21, 2021). Furthermore, while none of the litigation cases have been stayed, no defendant has moved to stay. *Int'l Bus. Machines Corp. v. Trusted Knight Corp.*, IPR2020-00323, Paper 15 at 9 (PTAB Jul. 10, 2020). Indeed, two of the cases have been terminated. *See* Ex. 1011 (*DigiMedia Tech, LLC v. Panasonic Corp.*, Case No. 1:20-cv-04995 (N.D.G.A.)); Ex. 1012 (*DigiMedia Tech, LLC v. ViacomCBS, Inc.*, Case No. 1:21-cv-01831 (S.D.N.Y.)). Only the *DigiMedia Tech, LLC v. Lenovo US Inc. et al.*, Case No. 1:21-cv-00227 (D. Del.) case is still pending. Ex. 1013.

Factor 2 weighs against discretionary denial because Unified is not a participant in any cases involving the '532 patent. A jury trial date has been set in the only case that remains pending for October 30, 2023, in *DigiMedia Tech, LLC*

v. Lenovo US Inc. et al., Case No. 1:21-cv-00227 (D. Del). Ex. 1014. The USPTO has provided guidance however, that **district court "scheduled trial dates are unreliable and often change."** USPTO *Interim Procedure For Discretionary Denials In AIA Postgrant Proceedings With Parallel District Court Litigation*, June 22, 2022, at 8. Thus, the USPTO suggests considering evidence of "median time-to-trial for civil actions in the district court in which the parallel litigation resides." *Id*. at 8-9. The average time-to-trial for a patent case in the District of Delaware is around three years (see LegalMetric Report, Ex. 1015, 2), which would be after the PTAB's final written decision in this matter. Thus, factor 2 favors institution.

Factor 3 weighs against discretionary denial because Petitioner is not a party to any litigation involving the '532 patent and, therefore, has not expended resources in that litigation. Further, Unified does not have any knowledge of resources expended by the parties in the parallel district court litigation other than public information. However, based on publicly available information, in the only remaining *Lenovo US* case, the parties have invested only minimal resources on validity issues regarding the '532 patent and the court has made minimal investment. Furthermore, the claim construction hearing is not scheduled to occur until January 6, 2023. Ex. 1016. Where, as here, "the district court[s] ha[ve] not issued orders related to the patent at issue in the petition, this fact weighs against exercising discretion to deny institution." *Fintiv* at 10.

Factor 4 weighs against discretionary denial. Unified is unaware of any overlap between this proceeding and the district court litigation other than the citation to the '532 patent in the complaint generally. Petitioner has filed this IPR due to its unique business model of deterring non-practicing entities from asserting patents of poor quality against strategic technologies and industries. *See*, *e.g.*, Jakel Decl. (Ex. 1010), ¶¶2-3. As Petitioner's perspective, goals and business model is unique, it has independently pursued invalidity grounds, prior art and claim selection without discussion, interaction, involvement, or input of any kind with any defendant. *Id*., ¶¶4-5, 13.

Factor 5 weighs against discretionary denial because Unified is not a defendant in the three litigation cases and none of those defendants is a real party-in-interest here. *Fintiv* at 13-14.

Factor 6 weighs against discretionary denial. As demonstrated by the Petition, the challenged claims are unpatentable over *Chiu* and *Gu*—which were not considered during prosecution. Further, this is the only petition challenging the '532 patent, even though DigiMedia Tech has asserted it in three different litigation cases. Thus, a determination of its validity by the Board would conserve resources.

B. 35 U.S.C. § 325(d) Does Not Favor Denial

Non-institution under § 325(d) would also be improper in view of the *Advanced Bionics* framework and *Becton Dickinson* factors. *See Advanced Bionics*

LLC v. Med-EL Elektromedizinische Geräte GMBH, IPR2019-01469, Paper 6, (PTAB Feb. 13, 2020) (precedential); *Becton Dickinson & Co. v. B. Braun Melsungen AG*, IPR2017-01586, Paper 8, 6-7 (PTAB Dec. 15, 2017).

The evidence identified in this Petition was not before the Examiner during prosecution. *Chiu* and *Gu* were not cited by the Examiner in any Office Action or provided in the Applicant's Information Disclosure Statement. '532 patent, pages 1-2; *see also* Ex. 1004. The Examiner issued a first action allowance for the '532 patent on the basis that the claims are not taught by the prior art. Ex. 1004, 88-95. As shown by the prior art relied upon in this petition, the Examiner's determination was in error. Furthermore, the Examiner also did not have the benefit of Dr. Karam's declaration, which explains what a POSA would have understood from the prior art as of the filing date of the '532 patent. *Karam*, ¶¶ 39-128. Accordingly, any argument for a discretionary denial of institution under 35 U.S.C. §325(d) would be misplaced.

VII. Mandatory Notices Under 37 C.F.R. § 42.8

A. Real Party-in-Interest

Pursuant to 37 C.F.R. §42.8(b)(1), Petitioner certifies that Unified Patents, LLC, is the real party-in-interest, and further certifies that no other party exercised control or could have exercised control over Unified's participation in this proceeding, the filing of this petition, or the conduct of any ensuing trial. In view of

Worlds Inc. v. Bungie, Inc., 903 F.3d 1237, 1242-44 (Fed. Cir. 2018), Unified has submitted a declaration to support its certification. *See* Declaration of Kevin Jakel (Ex. 1010).

B. Related Matters

Petitioner is aware of three cases involving the '532 patent: *DigiMedia Tech, LLC v. ViacomCBS, Inc. et al.*, Case No. 1:21-cv-01831 (S.D.N.Y.), filed March 2, 2021; *DigiMedia Tech, LLC v. Lenovo US Inc. et al.*, Case No. 1:21-cv-00227 (D. Del.), filed February 18, 2021; and *DigiMedia Tech, LLC v. Panasonic Corporation*, Case No. 1:20-cv-04995 (N.D.G.A.), filed December 9, 2020.

C. Lead and Backup Counsel

Petitioner identifies the following lead and backup counsel:

Lead Counsel	Backup Counsel
Trenton Ward (Reg. No. 59,157) Finnegan, Henderson, Farabow, Garrett & Dunner, LLP 271 17th Street, N.W. Suite 1400 Atlanta, GA 30363-6209 Telephone: 404-653-6441 Facsimile: 202-408-4400 Email: trenton.ward@finnegan.com	Roshan S. Mansinghani (Reg. No. 62,429) Unified Patents, LLC 4445 Willard Ave., Suite 600 Chevy Chase, MD 20815 Telephone: 214-945-0200 Email: roshan@unifiedpatents.com David C. Seastrunk (Reg. No. 73,723) Unified Patents, LLC 4445 Willard Ave., Suite 600 Chevy Chase, MD 20815 Telephone: 214-945-0200 Email: david@unifiedpatents.com William C. Neer (Reg. No. 78,874) Finnegan, Henderson, Farabow, Garrett & Dunner, LLP 901 New York Avenue, NW Washington, DC 20001-4413 Telephone: 202-408-4000 Email: william.neer@finnegan.com

VIII. Grounds for Standing

Pursuant to 37 C.F.R. § 42.104(a), Petitioner certifies that the '532 patent is available for IPR and that Petitioner is not barred or estopped from requesting this review.

Date: June 27, 2022

Respectfully Submitted,

/Trenton A. Ward/
Trenton A. Ward (Reg. No. 59,157)
Lead Counsel

CERTIFICATION UNDER 37 C.F.R. § 42.24(d)

This Petition complies with the requirements of 37 C.F.R. § 42.24. As calculated by the word count feature of Microsoft Word, it contains 13,629 words, excluding the words contained in the following: Table of Contents, Table of Authorities, List of Exhibits, Mandatory Notices, Certification Under § 42.24(d), and Certificate of Service.

/Trenton A. Ward/
Trenton A. Ward (Reg. No. 59,157)
Lead Counsel
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, LLP

CERTIFICATE OF SERVICE

Pursuant to 37 C.F.R. §§ 42.6(e) and 42.105(a), the undersigned certifies that on June 27, 2022, a copy of the foregoing **Petition for *Inter Partes* Review of U.S. Patent 6,473,532**, the associated P**ower of Attorney**, and **Exhibits 1001-1008 and 1010-1019** were served by FedEx Priority Overnight on the correspondence address of record indicated in the Patent Office's public PAIR system for U.S. Patent No. 6,473,532:

Michael R. Houston
FOLEY & LARDNER LLP
3000 K STREET N.W.
SUITE 600
WASHINGTON, DC 20007-5109

Date: June 27, 2022

By: /William Esper/
William Esper
Case Manager and PTAB Coordinator
FINNEGAN, HENDERSON, FARABOW,
GARRETT & DUNNER, LLP