IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

AMAZON.COM, INC., and
AMAZON WEB SERVICES, INC.,

Petitioners,

v.

DIVX, LLC,

Patent Owner.

Case: IPR2025-01062

U.S. Patent No. 10,715,806

PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 10,715,806

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Page

TABLE OF CONTENTS
(Continued)

Page

TABLE OF CONTENTS
(Continued)

Page

TABLE OF CONTENTS
(Continued)

Page

TABLE OF CONTENTS
(Continued)

Page

TABLE OF CONTENTS
(Continued)

Page

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Page

TABLE OF AUTHORITIES

Page(s)

CASES

STATUTES AND RULES

OTHER AUTHORITIES

EXHIBIT LIST (37 C.F.R. § 42.63(e))

Ex. No.	Description
1001	U.S. Patent No. 10,715,806 ("the 806 Patent")
1002	Excerpts of Prosecution File History for the 806 patent
1003	Declaration of Hari Kalva, Ph.D.
1004	Sambe, Yasuo et al., High Speed Distributed Video Transcoding for Multiple Rates and Formats, IEICE Transactions on Information and Systems, August 2005 ("Sambe")
1005	U.S. Patent No. 6,490,320 ("Vetro")
1006	U.S. Patent Application Publication No. 2010/0189183 ("Gu")
1007	Xin, Jun et al., Digital Video Transcoding, Proceedings of the IEEE, Vol. 93, No. 1, January 2005
1008	Ahmad, Ishfaq et al., Video Transcoding: An Overview of Various Techniques and Research Issues, IEEE Transactions on Multimedia, Vol. 7, No. 5, October 2005
1009	Vetro, Anthony et al., Object-Based Transcoding for Adaptable Video Content Delivery, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 3, March 2001
1010	Kuhn, Peter et al., MPEG-7 Transcoding Hints for Reduced Complexity and Improved Quality, 2001
1011	Lao, Feng et al., Parallelizing Video Transcoding Using Map-Reduce-Based Cloud Computing, 2012
1012	Dogan, S. et al., MPEG-4 Video Transcoder for Mobile Multimedia Traffic Planning, Centre for Communication Systems Research (CCSR), University of Surrey, UK, 2001
1013	Richardson, Iain E., The H.264 Advanced Video Compression Standard, 2nd ed., 2010

Ex. No.	Description
1014	Sun, Huifang et al., Digital Video Transcoding for Transmission and Storage, 2005
1015	Vetro, Anthony et al., Video Transcoding Architectures and Techniques: An Overview, IEEE Signla Processing Magazine, March 2003
1016	Springer - Handbook of Data Compression, 5th ed., 2010
1017	Van Beek, Peter et al., Metadata-Driven Multimedia Access, IEEE Signal Processing Magazine, March 2003
1018	Ho, Chia-Chiang et al., Building MPEG-7 Transcoding Hints from Intrinsic Characteristics of MPEG Videos, IEEE Transactions on Consumer Electronics, Vol. 50, No. 1, February 2004
1019	Ely, S. R., MPEG Video Coding A Simple Introduction, EBU Technical Review, 1995
1020	Yu, Lu and Wang, Jian-peng, Review of the current and future technologies for video compression, Journal of Zhejiang University-Science C (Computers & Electronics), 2010
1021	Shi, Changui et al., A Fast MPEG Video Encryption Algorithm, ACM Multimedia, 1998
1022	Xie, Bo and Zeng, Wenjun, A Sequence-Based Rate Control Framework for Consistent Quality Real-Time Video, IEEE Transactions and Systems for Video Technology, Vol. 16, No. 1, January 2006
1023	Bozoki, S. et al., Parallel Algorithms for MPEG Video Compression with PVM, Dept. of Telecommunications, 1996
1024	He, Yong et al., A Software-Based MPEG-4 Video Encoder Using Parallel Processing, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 8, No. 7, November 1998

Ex. No.	**Description**
1025	A true and correct copy of the MARC record for the journal *IEICE Transactions on Information and Systems* from the Linda Hall Library of Science, Engineering & Technology online catalog
1026	Library of Congress Authorities, MARC Display, Subject Heading for Information Technology Periodicals
1027	A true and correct copy of the MARC record for the journal IEICE Transactions on Information and Systems obtained from the OCLC bibliographic database
1028	Declaration of Sylvia D. Hall-Ellis, Ph.D.
1029	G. J. Sullivan, J. -R. Ohm, W. -J. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649-1668, Dec. 2012, doi: 10.1109/TCSVT.2012.2221191. PDF available here: https://ieeexplore.ieee.org/abstract/document/6316136
1030	Curriculum Vitae of Hari Kalva, Ph.D.

Pursuant to 35 U.S.C. § 311 and 37 C.F.R. § 42.100, Amazon.com, Inc. and Amazon Web Services, Inc. ("Petitioner") requests *Inter Partes* Review ("IPR") of claims 1-21 (the "Challenged Claims") of U.S. Patent No. 10,715,806 (the "806 patent"), assigned to DivX, LLC. The Challenged Claims are obvious over the prior art. Accordingly, the Patent Trial and Appeal Board should institute review of the Challenged Claims.

I. REQUIREMENTS FOR IPR UNDER 37 C.F.R. § 42.104

A. Standing (37 C.F.R. § 42.104(a))

Petitioner certifies that the 806 patent is available for IPR and that Petitioner is not barred or estopped from requesting IPR of the Challenged Claims.

B. Identification of Challenge (37 C.F.R. § 42.104(b))

In view of the prior art and evidence presented, the Challenged Claims are unpatentable and should be cancelled. 37 C.F.R. § 42.104(b)(1). Based on the prior art identified below, IPR of the Challenged Claims should be instituted. 37 C.F.R. § 42.104(b)(2).

Proposed Grounds of Unpatentability	
Ground 1	Claims 1-4, 7-14, and 17-21 are obvious under pre-AIA § 103 over Sambe, Vetro, and General Knowledge of a POSITA
Ground 2	Claims 5-6 and 15-16 are obvious under pre-AIA § 103 over Sambe, Vetro, Gu, and General Knowledge of a POSITA

Sections IV.A-B identify where each element of the Challenged Claims is found in the prior art. 37 C.F.R. § 42.104(b)(4). The exhibit numbers of the evidence relied on to support the challenges are provided above and the relevance of the evidence to the challenges raised are provided in Sections II-III. 37 C.F.R. § 42.104(b)(5). Exhibits 1001-1030 are attached.

C. Claim Construction (37 C.F.R. § 42.104(b)(3))

For the purposes of this Petition, all claims are interpreted with their ordinary and customary meaning as understood from the perspective of a person of ordinary skill in the art ("POSITA").[1]

II. OVERVIEW OF THE 806 PATENT

A. Technical Background

1. Introduction to Video Coding

Video coding is the process of compressing (encoding) and decompressing (decoding) digital video. EX1013, 7; EX1003, ¶30. Digital video is composed of a series of successive images, known as frames—commonly 24 per second for film, and more for television and other applications. *See* EX1014, 207-208. Each frame includes a matrix of hundreds of thousands of pixels that combine to form an image. *See* EX1013, 7-16; *see also* EX1014, 12-16. This large amount of information

[1] Petitioner reserves the right to offer different claim constructions in other forums.

contained within digital videos makes them unwieldy to transmit and store in their original uncompressed form. *See* EX1013, 26; *see also* EX1014, 33-36.

Video coding addresses this problem by reducing video size, allowing for efficient transmission and storage. EX1014, 36; *see also* EX1019, 13; EX1003, ¶31. The coding process first breaks each video frame into processing units known as macroblocks or coding units. EX1016, 883; *see also* EX1019, 17, 19; EX1014, 37-45; EX1003, ¶32. Macroblocks or coding units are each individually processed—the coding steps described below are applied to them individually. EX1019, 14, 19; *see also* EX1014, 37-38, 44-45; EX1013, 31-35, 42-48l; EX1003, ¶33.

Then, operating at the macroblock or coding unit level, the coding process eliminates pixel redundancies within a given frame and between different frames in the video. EX1013, 25-26; *see also* EX1014, 36-47; EX1019, 13; EX1003, ¶33. This is possible because "pixel values are not independent but are correlated with their [sic] neighbours, both within the same frame and across frames." EX1013, 25-26; *see also* EX1014, 36-37; EX1019, 13; EX1003, ¶33. Redundancy within the same frame is known as spatial redundancy (*e.g.*, many black pixels representing a night sky in one frame). EX1013, 26; EX1003, ¶33. Redundancy across different frames is known as temporal redundancy (*e.g.*, pixels representing the same portion of the night sky in successive frames). EX1013, 25-26; *see also* EX1014, 44; EX1003, ¶33. Accordingly, the value of a pixel can be predicted based on values of neighboring pixels within the same frame (spatial) or between different frames (temporal). EX1019, 13; *see also* EX1016, 869-873; EX1013, 25-26; *see also* EX1014, 36-47; EX1003, ¶¶34-37.

These spatial and temporal redundancies are removed by constructing a prediction of a current macroblock based on neighboring frame and/or pixel values. EX1013, 27; EX1014, 44-45; *see also* EX1019, 16-19; EX1003, ¶¶35-37. The output, called a "prediction residual," represents the prediction macroblock subtracted from the original uncompressed macroblocks. EX1013, 27; *see also* EX1019, 19; EX1003, ¶35.

A further mathematical transformation—typically involving applying a discrete cosine transform to the prediction residual and quantizing the resulting transform coefficients—then compresses the prediction residual, resulting in a set of values, known as quantized transform coefficients. EX1013, 27; *see also* EX1014, 37-41; EX1019, 14-15; EX1003, ¶¶38-40. The quantized transform coefficients are then passed to an entropy encoder for further processing. EX1013, 27; EX1019, 15, 19, Fig. 4; EX1003, ¶¶40-45. The entropy encoder removes "statistical redundancy in the data, [like] for example representing commonly occurring vectors and coefficients by short binary codes." EX1013, 27; EX1019, 15, 19, Fig. 4. Entropy encoding assigns fewer bits to frequently recurring transform coefficients, thereby minimizing the number of bits needed to encode each frame or block. EX1020, 9; EX1021, 82.

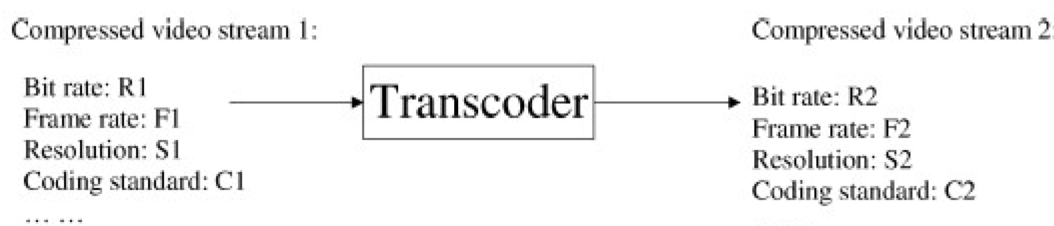
This coding process results in an encoded video that is smaller than the original video file. EX1013, 27; EX1003, ¶45. To view the encoded video requires using a decoder to recreate the original video by reversing the encoding process. EX1013, 27-28. The resulting decoded video will not be identical to the original video due to loss of data during the encoding process, but the difference will be unnoticeable to a viewer. *Id.*, 25; EX1003, ¶46.

2. Introduction to Video Transcoding

The 806 patent is directed to a form of video coding known as transcoding. *See generally* EX1001; EX1003, ¶47. Video transcoding refers to converting a video from one encoded format to another. EX1007, 84; *see also* EX1008, 793; EX1015, 18. Transcoding emerged in the early 2000s to deal with rapid advancements in multimedia systems. EX1015, 18. In particular, emerging multimedia systems allowed users to access video across a diverse array of different devices (computers, cell phones, and televisions), platforms (video streaming), and networks (Wireless Local Area Networks and cellular). *Id.* To account for this diversity, video providers needed to be able to distribute and play content across all devices, platforms, and networks. *Id.* But there was no one-size-fits-all approach to video coding for use across all devices, platforms, and networks. *See* EX1004, 1923; *see also* EX1009,

387. Transcoding addressed this problem by converting videos from one encoded format to another. *See id.*; EX1003, ¶¶47-51.

Transcoding could change a video file from one encoding standard, such as MPEG-2, to another encoding standard, such as MPEG-4. *Id*. To address different device or network requirements, transcoding could also change coding parameters for a video, such as bit rate, frame rate, or resolution. EX1007, 84-85, Fig. 1. For example, a video could be changed from a constant bitrate format to a variable bitrate format, allowing the transcoder to allocate bit rates based on video content complexity (assigning higher bit rate to complex scenes), which ensures a more consistent level of video quality regardless of scene complexity. EX1014, 176; EX1010, 1, 3; EX1003, ¶ 48. An example of transcoding is shown in the below flow diagram. EX1007, Fig. 1. A compressed video stream 1 is received and converted by a transcoder to a compressed video stream 2 having a different encoding standard and different coding parameters. *Id.*



EX1007, Fig. 1.

Transcoders are composed of both a decoder and an encoder. *See, e.g.*, EX 1005, 2:14-31, Fig. 1; *see also, e.g.*, EX1007, 85, Fig, 4; EX1014, 105; EX1003,

¶52. This is shown in the below image. A transcoder receives an encoded video with given bitrate (Rin). EX1005, 2:14-31, Fig. 1. The transcoder decodes the encoded video (step 110) and re-encodes it (step 120) at a new bitrate (Rout) based on the specific requirements 102 of a user's device. *Id.*

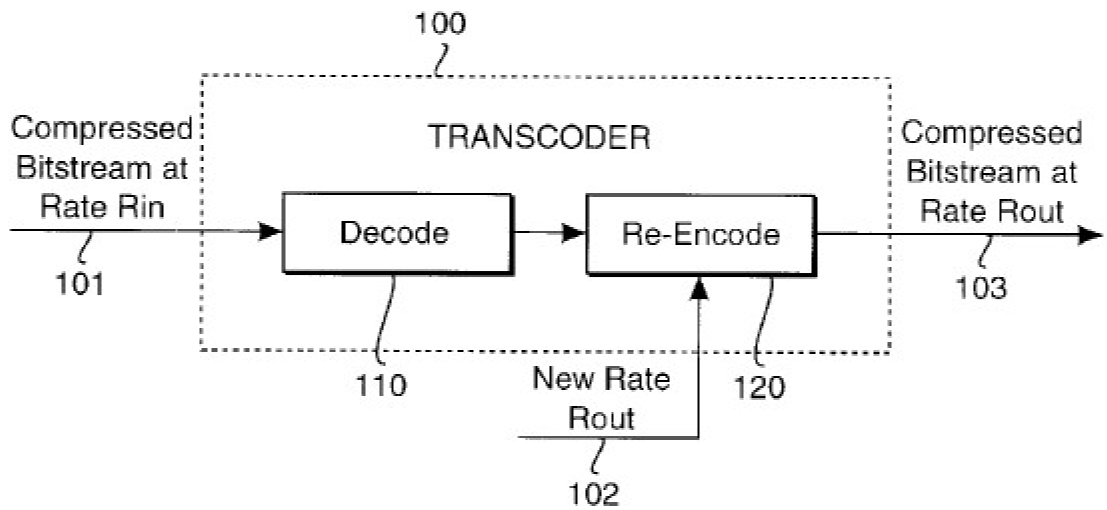


FIG. 1
PRIOR ART

EX1005, Fig. 1.

3. Well-Known Transcoding Implementations

An early area of focus for POSITAs was transcoding speed. Using a single transcoder to decode and then reencode an entire video was limited by the computing resources and processing capacity available to that single transcoder. EX1003, ¶53; *see also cf.* EX1011, 2905 (describing parallel transcoding as a solution for fast transcoding); EX1012, 3-4 (describing parallel transcoding as "more efficient and less delay prone" than sequential transcoding); *see also* EX1023, 1, 12; EX1024,

910. So, POSITAs began using multiple transcoders to transcode different segments of the same video in parallel. *See* EX1011, 2905; *see also* EX1004, 1923; EX1003, ¶53. Those segments would then be recombined to form the reencoded video. *See* EX1011, 2905; *see also* EX1004, 1924.

But using multiple transcoders to transcode different segments of a video had its own set of challenges—primarily, ensuring a consistent video quality regardless of which transcoder handled which segment. EX1004, 1923; *see also* EX1011, 2906. Ensuring consistent video quality requires proper budgeting and allocation of bitrates to each video segment. EX1007, 84-85, 91; *see also* EX1009, 391; EX1010, 1; EX1003, ¶54. This is because a video segment with significant motion (e.g., showing a football game with active play) requires more sophisticated motion estimation and therefore more bitrate than a video segment with minimal motion (e.g., showing a news anchor sitting at a desk). EX1017, 48; *see also* EX1005, 21:7-56; EX1003, ¶54. In other words, more complex video segments require more bits to compress and maintain a consistent quality than less complex video segments. *Id.* To properly budget and allocate bitrates for each segment, a POSITA would have to understand the complexity of each segment individually. EX1017, 48; *see also* EX1005, 21:7-56; EX1003, ¶54.

Since at least the introduction of MPEG-7 in 2001, well before the 2013 priority date claimed by the 806 patent, one well-known way to understand the complexity of a video segment was by extracting metadata from the video. EX1017, 47-48; *see also* EX1009, 391-92; EX1010, 1; EX1005, 4:15-5:11. Metadata is "any suitable information about the media content" of an encoded video. EX1001, 4:25-26. This can include general information about a video, such as the encoding standard with which a given video is encoded, or specific information about the video, such as complexity information about a given scene within the video. EX1017, 47-48; *see also* EX1009, 391-92; EX1010, 1; EX1005, 21:8-22:61. EX1003, ¶55. Accordingly, to understand the complexity of each video segment, a POSITA would have looked to extracted metadata, such as, for example, the "difficulty hint" defined by the MPEG-7 standard. EX1017, 47-48; *see also* EX1005, 18:24-20:14, 21:7-57; EX1009, 391-92; EX1010, 1; EX1003, ¶¶55-56. "The difficulty hint describes the encoding complexity of segments within a video sequence (or regions within an image)." EX1017, 48; *see also* EX1018, 306; EX1010, 3-4. Using the difficulty hint, finite bitrate resources would be budgeted between video segments to ensure that each segment, regardless of its complexity, is transcoded with a consistent quality. EX1017, 48; *see also* EX1010, 3-4; EX1005, 21:7-57; *see also* EX1009, 391; EX1010, 1. But for this to work, the transcoders must know the encoding difficulty of video segments prior to transcoding. EX1017, 48; EX1018, 306; EX1010, 3. Otherwise, the bitrate allocation for each of the video segments would be imperfect due to a lack of information regarding the frames coming immediately before and after the video segment. *Cf.* EX1004, 1923 (describing other techniques); EX1003, ¶¶57-59. This would result in inconsistent video quality and improper budgeting of bitrates not only for the earlier transcoded video segments but would have a cascading effect on the bitrate allocation for the remaining segments in a video. EX1004, 1923, 1924-1927 (describing issues relating to poor quality information); EX1003, ¶59.

B. Patent Description

The 806 patent is directed to a well-known process of extracting metadata and using a plurality of transcoders to transcode segments of a video using that metadata. EX1003, ¶¶60-86. As explained in the Background Section, transcoders can decode a video encoded in a first format into a raw video file, and then reencode that file into a second format using a set of encoding parameters. EX1001, 1:33-40. The 806 patent claims require extracting metadata during transcoding but prior to decoding and then using information based on the extracted metadata for parallel transcoding using a plurality of transcoding devices. *Id.*, 13:49-59; EX1003, ¶86. The Applicant considered this to be novel; during prosecution, the Applicant responded to a prior art rejection by arguing that, unlike the prior art, the 806 patent teaches "generating … media metadata related to the source video file *prior* to decoding at least a portion of the source video file." EX1002, Response to OA dated July 9, 2019, at 9-10 (emphasis added). The metadata described in the 806 patent includes "complexity information" about video scenes. EX1001, 4:31-50. The "complexity information" can include "any suitable information about the spatial and/or temporal complexity of an image." *Id.*, 4:45-50. The claimed limitations, including the purportedly novel limitation to extract metadata during transcoding but prior to decoding, were all well-known to a POSITA at the time of the claimed invention. *See supra* § II.A.3.

C. Priority Date

The 806 patent was filed as a continuation of application No. 15/905,695 filed on February 26, 2018, which is a continuation of application No. 13/841,943, filed on March 15, 2013. EX1001, Cover. For purposes of this proceeding only, Petitioner does not dispute that the priority date is March 15, 2013. EX1003, ¶87.

D. Level of Ordinary Skill in the Art

A POSITA would have a bachelor's or master's degree in electrical engineering, computer engineering, or a related field, with specialized training or coursework in digital video coding and processing and related technologies. The individual would have at least four years of practical experience in digital video coding and processing, including software and/or hardware architectures used to perform digital video coding and processing. A person could also have qualified, as a POSITA, with more formal education and less technical experience, or vice versa. EX1003, ¶¶88-91.

III. PRIOR ART OVERVIEW[2]

A. Sambe

"High-Speed Distributed Video Transcoding for Multiple Rates and Formats" is a printed publication written by Yasuo Sambe *et al.* and published in August 2005 ("Sambe," EX1004). EX1004, 1923. Sambe is prior art to the 806 patent under at least 35 U.S.C. §§ 102(a) and (b). EX1028. Sambe was not disclosed to the Patent Office during prosecution of the 806 patent. EX1003, ¶¶92-93.

Sambe describes a distributed video transcoding system that divides source video files into segments and transcodes those segments in parallel. EX1004, 1923-24; EX1003, ¶¶93-99. This process is illustrated in Figure 1 below. The Source PC, the red box in Figure 1 below, receives a source video file and divides it into multiple video segments. EX1004, 1924, Fig. 1 (annotated). These segments are transmitted to a plurality of Transcoding PCs, the yellow box in Figure 1 below, along with transcoding parameters. *Id.* These parameters include filter function specifications, spatial resolution, temporal resolution, and re-encoding formats. *Id.*; EX1003, ¶94. The Transcoding PCs each decode and re-encode the video segments based on these parameters. EX1004, 1923-24; EX1003, ¶95. Specifically, the video segments are divided into frames, and each frame is passed to the encoder. EX1004, 1923-24.

---

[2] Pre-AIA 35 U.S.C. § 102 applies to the 806 patent.

The frame-by-frame based transcoding architecture provides for greater flexibility in transcoding, by, for example, allowing "new encoding modules or new filter operations like digital water marking [to] be easily added." *Id.*

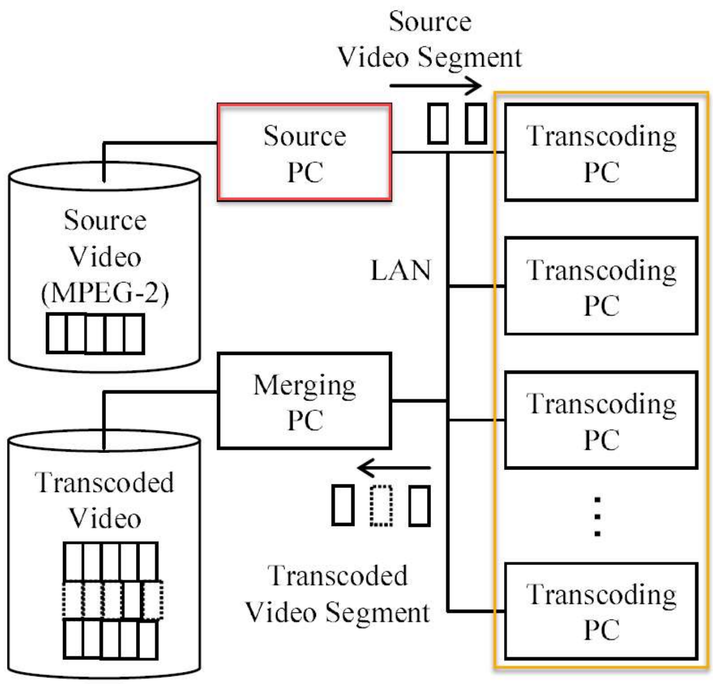


Fig. 1 Distributed video transcoder

Id., 2, Fig. 1.

Figure 17 of Sambe, shown below, demonstrates the plurality of Transcoding PCs processing the segments of the video file in parallel. *Id.*, 7; EX1003, ¶97. The Source PC of Sambe divides a video stream into segments, like, for example, segment 1 and segment 2 (colored red and green for illustrative purposes, respectively) shown below. EX1004, Fig. 17 (annotated). Segments 1 (red) and 2 (green) are then transmitted to Transcoding PC (1) and (2), respectively, which then transcode their respective segments "in parallel." *Id.*, 1923-24, 1929, Figs. 2, 17.

After transcoding, encoded segments 1 (red) and 2 (green) are sent to the Merging PC where they are concatenated (stitched together) with the remaining transcoded segments to form the desired video file. *Id.*, 1923-24, 1929, Figs. 2, 17. Using a plurality of Transcoding PCs to process different segments of a video in parallel increased transcoding speeds by a factor of seven. *Id.*, 1923, 1931.

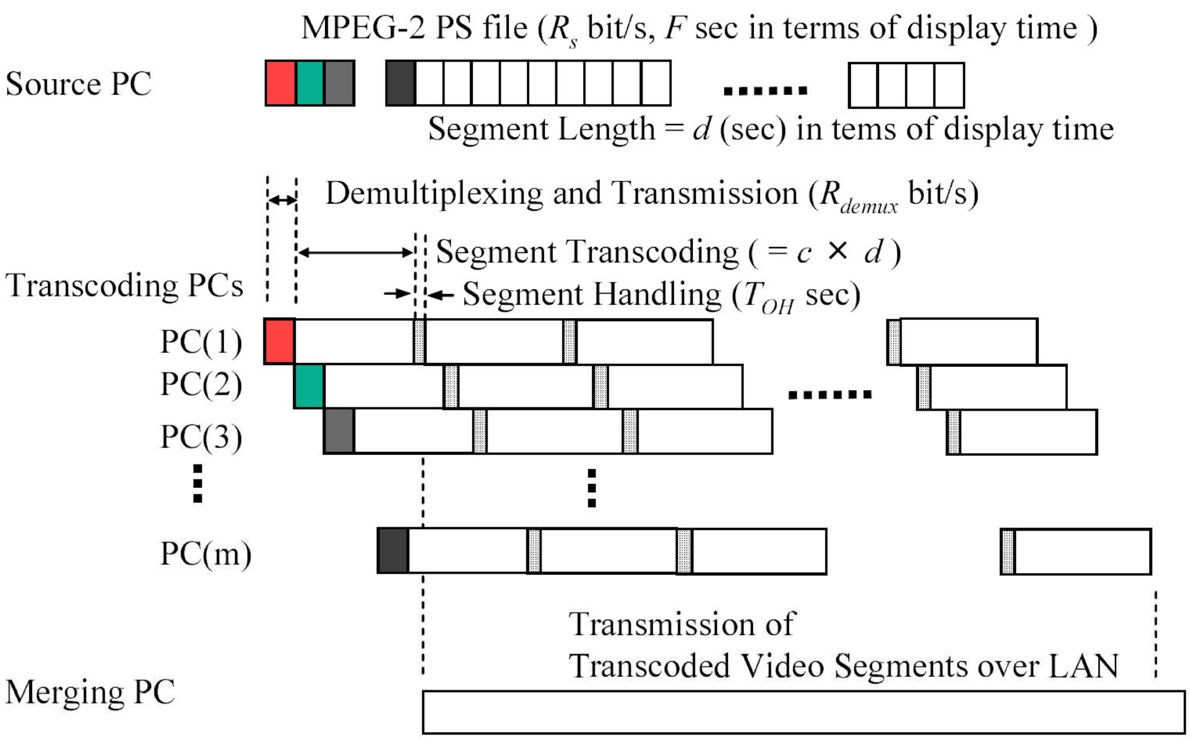


Fig. 17 Process flow of the distributed transcoding

Id., Fig. 17 (annotated).

B. Vetro

U.S. Patent No. 6,490,320 B1 to Anthony Vetro et al., titled "Adaptable Bitstream Video Delivery System," was filed for on April 11, 2000, and issued on December 3, 2002 ("Vetro," EX1005). EX1005, Cover. Vetro is prior art to the 806 patent under 35 U.S.C. §§102(a), (b), and (e). Vetro was not disclosed to the Patent Office during prosecution of the 806 patent. EX1003, ¶¶100-101.

In Figure 3, shown below, Vetro describes a transcoding system that ensures efficient delivery of content while maintaining its quality. EX1005, 5:61-6:23; EX1003, ¶102-106. The content classifier 310 (green box) extracts information describing the bitstream (e.g., motion activity, video change information and texture) from the bitstream 301. EX1005, 7:62-8:1, Fig. 3 (annotated). This information is then mapped to "a finite set of semantic classes or high-level meta-data." *Id.*, 7:62-8:5, 8:21-36, 8:60-9:1. To assist with this mapping of the extracted information to "meta-data," content classifier 310 receives "metadata information 303." *Id.*, 8:1-5, Fig. 3 (annotated). This metadata is also sent directly to the CND Manager and the Switchable Transcoder to assist with transcoding, and "include[s] descriptors and description schemes" specified by the MPEG-7 standard. *Id.*, 8:1-5, Fig. 3 (annotated); EX1003, ¶105. These MPEG-7 standard descriptors and description schemes describe many features of the video content, including the relative complexity of "shots," which are segments of the video. EX1005, 17:31-19:22. A POSITA would have understood the "shots" to be akin to scenes. EX1003, ¶105. For example, the 806 patent describes "a shot change" as a type of scene change. EX1001, 4:31-44; *see also, e.g.*, EX1005, 17:41-45 ("A 'shot' can be a group of frames (GOF's)" and represent "smaller segments of video that begin when a camera is turned [on] and last until the camera is turned off."). The extracted information, which are directly mapped into metadata, allow the CND manager to allocate bitrates among the "shots" to meet the bitrate and quality objectives. *Id.*, 7:62-8:1, 21:8-49; EX1003, ¶105.

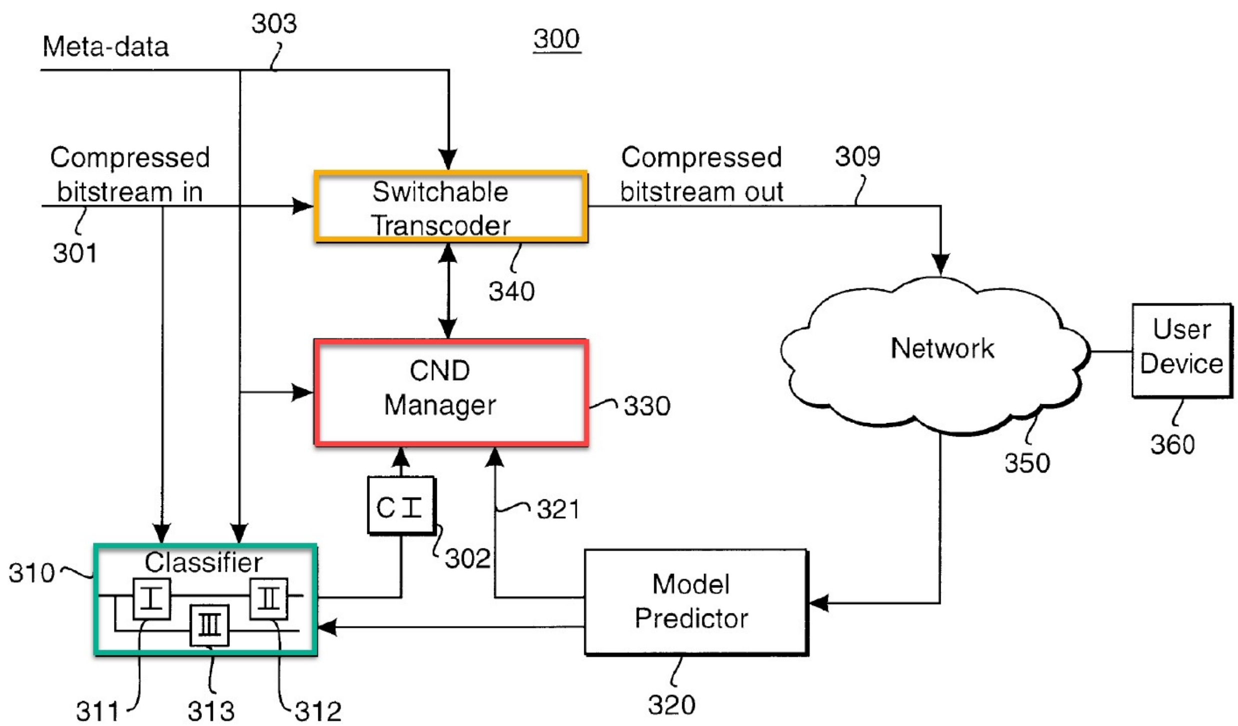


FIG. 3

EX1005, 7:40-43, Fig. 3 (annotated).

Figure 4 of Vetro, shown below, depicts an expanded view of the Vetro transcoder, illustrating the inner workings of the switchable transcoder. Using the metadata provided by the content classifier, the CND manager determines the optimal rate-quality function for the video content. *Id.*, 10:10-18; EX1003, ¶¶105-

112. This function models the optimal quality that can be achieved for a given bit rate and user device, and factors into determining which of the transcoders in switchable transcoder 340 is best suited to transcode the video content. *Id.*, 9:48-63. Based on the rate-quality function, the CND manager determines whether the discrete summary transcoder 441, the continuous conversion transcoder 442, or some other transcoder 443 is best suited to transcode the video content. *Id.*, 9:31-39, 10:7-24, Fig. 4.

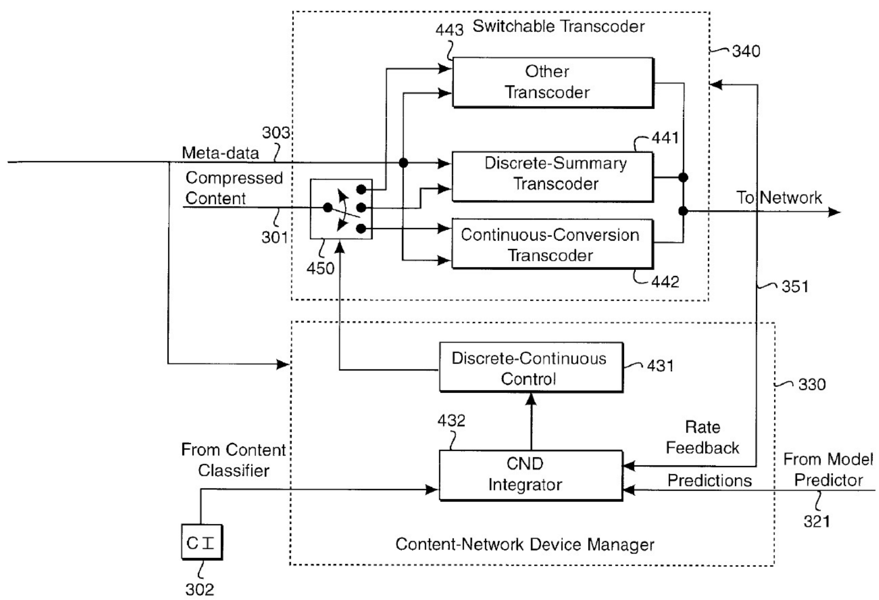


FIG. 4

Id., Fig. 4.

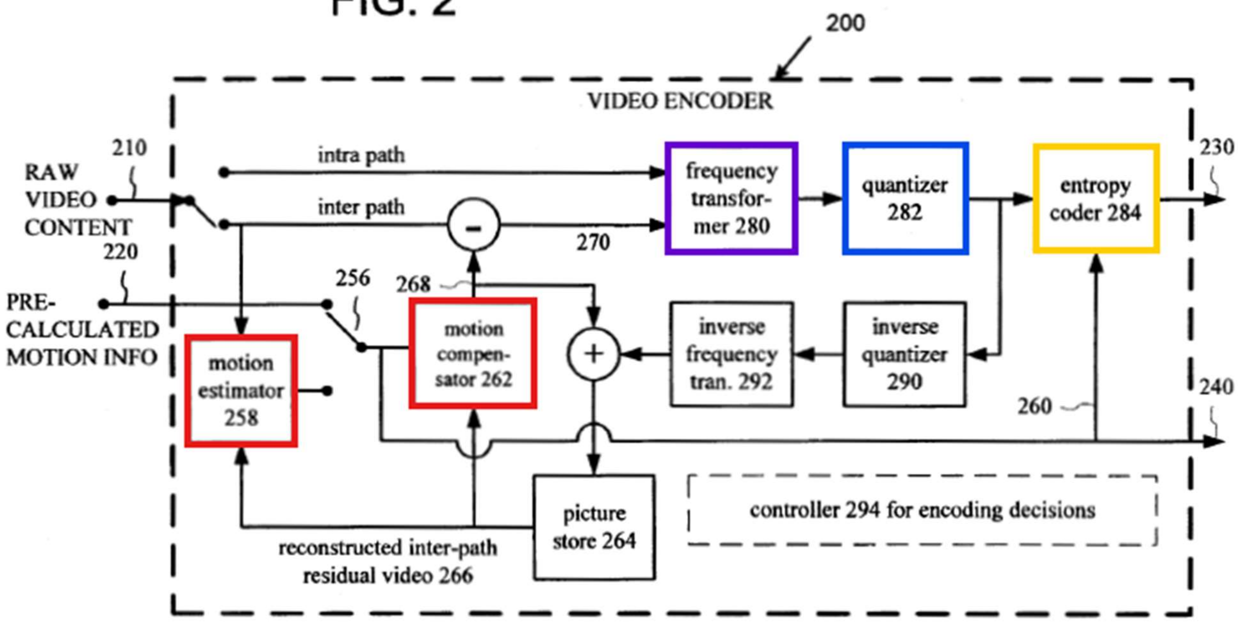
C. Gu

U.S. Patent Application Publication No. 2010/0189183 A1 to Chuang Gu et al., titled "Multiple Bit Rate Video Encoding Using Variable Bit Rate and Dynamic Resolution for Adaptive Video Streaming," was filed on January 29, 2009, and published on July 29, 2010 ("Gu," EX1006). Gu is prior art to the 806 patent under 35 U.S.C. §§ 102(a), (b), and (e). Gu was disclosed to the PTO during prosecution of the 806 patent but was not individually addressed by the Examiner during prosecution. EX1003, ¶¶113-114.

The video encoder of Gu uses well-known video coding steps to process the video content. *Supra* § II.A.1; EX1003, ¶¶115-123. Specifically, in Figure 2, as shown below, the motion estimator 258 and motion compensator 262 (red boxes below), together, prepare prediction frames and/or macroblocks, referred to as motion-compensated current pictures and/or macroblocks, and by calculating the difference between the prediction frames and/or macroblocks and the original frame and/or macroblocks, prepare prediction residuals for each frame and/or macroblock. EX1006, [0031]-[0033], Fig. 2 (annotated); EX1003, ¶¶118-121. The frequency transformer 280 (purple box below) receives the prediction residuals, and "applies a DCT, variant of DCT, or other forward block transform to blocks of … prediction residual data, producing blocks of frequency transform coefficients." EX1006, [0034], Fig. 2 (annotated); EX1003, ¶121. The quantizer 282 (blue box below) quantizes the generated transform coefficients on a frame-by-frame or macroblock-by-macroblock basis. EX1006, [0035]–[0036], Fig. 2 (annotated); EX1003, ¶122. Finally, entropy coder 284 (yellow box below) encodes "the output of the quantizer

282" resulting in a compressed bitstream as output. EX1006, [0037]–[0038], Fig. 2 (annotated); EX1003, ¶123.

FIG. 2



EX1006, Fig. 2 (annotated).

IV. INVALIDITY

A. Ground 1: Claims 1-4, 7-14, and 17-21 are obvious over Sambe, Vetro, and the General Knowledge of a POSITA

1. A POSITA would have combined Sambe and Vetro

A POSITA would have been motivated to combine Sambe and Vetro. Specifically, a POSITA would have sought to combine the distributed video transcoding architecture of Sambe, including a source PC and a plurality of transcoding PCs operating in parallel, with the teachings of the transcoding system of Vetro. EX1004, 1924, Fig. 1; EX1005, 7:40-43, Fig. 3; EX1003, ¶124. This combination would take advantage of both the increased transcoding speed offered by Sambe's distributed transcoding architecture, and improved understanding of coding complexity for transcoding disclosed in Vetro. EX1003, ¶124.

Sambe describes a high-speed distributed transcoding system intended to fulfill a growing demand for transcoding video between different encoding formats used by internet video applications. EX1004, 1923. Sambe explains that traditional parallel transcoding methods can generate videos with quality discontinuity and degradation at video segment boundaries "because of a lack of information such as the coding complexity of the previous video segment." *Id.* To deal with this, the transcoding PCs of Sambe ***estimate*** the coding complexity for the first frames of each video segment for use in transcoding. *Id.*, 1924-27; EX1003, ¶¶125-130. These estimates help with the quality issue identified by Sambe, but a POSITA would have understood that actual complexities would be even more useful than an estimate. EX1003, ¶130; *see also* EX1007, 92; EX1022, 56. Vetro describes extracting and using metadata describing complexity of "shots," or video segments, to understand the ***actual*** complexity of the "shots." EX1005, 18:24-19:21; EX1003, ¶130-136. By using metadata about the actual complexity of the "shots," Vetro optimizes bitrate allocation per shot and ensures a consistent video quality. EX1005, 21:8-21; EX1003, ¶¶134-136.

A POSITA combining these teachings of Sambe and Vetro would have had a reasonable expectation of success in doing so because it would simply require incorporating the content classifier and CND manager components of Vetro into the source PC of Sambe and improving the transcoding PCs of Sambe with the teachings of the switchable transcoder of Vetro. EX1003, ¶¶136-138. The content classifier generates metadata that describes, among other things, the complexity of the "shots" making up the video content, and the CND manager uses the metadata to determine how to optimally transcode each of the shots— ensuring each shot is encoded at a consistent quality with an optimal number of bits. EX1005, 5:61-6:23, 8:60-9:9, 9:25-10:24, 18:24-19:58; EX1003, ¶136.

A POSITA would have known to improve each of the transcoding PCs of Sambe with the teachings of the switchable transcoder of Vetro. EX1003, ¶138. The additional metadata would provide even more guidance on which of the multiple transcoding approaches available in the switchable transcoder is best suited for each video segment. EX1003, ¶136; *see also* EX1005, 21:8-57. The improvement would also have been straightforward, because Sambe's transcoding PCs are both flexible and modular. As Sambe explains, its "transcoding architecture gives greater flexibility in transcoding" allowing "new encoding modules or new filter operations" to "be easily added." EX1004, 1924, Figs. 1-2. Thus, a POSITA would have understood that Sambe anticipated customization of the specific transcoder employed in its distributed transcoding architecture. EX1003, ¶138. A POSITA would have further understood that the content classifier and CND manager of Vetro should be incorporated into the source PC of Sambe. EX1005, 5:31-32, Fig. 4 (describing the CND manager as the "transcoder manager"); EX1003, ¶137. This would ensure that any pre-processing of the source video, such as segmenting, determining transcoding parameters, media metadata generation, and determining how the compressed bitstream gets decoded, occurs once at the Sambe source PC for efficiency. EX1003, ¶137. Thus, a POSITA would know how to achieve this combination with minimal design impact. EX1003, ¶138.

2. Claim 1

(1) [1.pre] "A method for transcoding a source video file into a set of multiple alternate video streams, the method comprising:"

To the extent the preamble is considered limiting, Sambe and Vetro individually render it obvious. EX1003, ¶¶143-147.

A POSITA would have understood that Sambe discloses a method for transcoding a source video file into a set of multiple alternate video streams. EX1003, ¶¶144-145. Sambe "describes a distributed video transcoding system that can simultaneously transcode an MPEG-2 video file into various video coding formats with different rates." EX1004, 1923. Specifically, Sambe's "distributed video transcoder" includes a source PC that transmits video segments and their associated transcoding parameters to a plurality of transcoding PCs to be transcoded. EX1004, 1924, Fig. 1. Sambe further discloses that each transcoding PC transcodes its respective video segments, and then sends them to the merging PC "to form the desired video format files." EX1004, 1923-24, Figs. 2, 17; EX1003, ¶146. A POSITA would have understood the desired video format files of "various video coding formats with different rates" to be the set of multiple alternative video streams. EX1003, ¶146.

Similarly, Vetro discloses a transcoding system that "generate[s] variation and/or summary bitstreams 1308 from an original compressed bitstream (Video In) 1301." EX1005, 23:26-30. Vetro explains that these variations will be stored, "so that in the future, a bit stream for some real-time operating condition will be readily available to the downstream transcoders." *Id.*, 23:50-54. A POSITA would have understood the generated bitstreams from an original compressed bitstream to be a set of multiple alternate video streams. EX1003, ¶147.

(2) [1.a] "generating, at a computer system configured as a media metadata generation device, media metadata related to the source video file prior to decoding, during a transcoding of, at least a portion of the source video file, where the media metadata comprises scene complexity information;"

The combination of Sambe in view of Vetro renders limitation 1[a] obvious. EX1003, ¶¶148-157.

Sambe's "distributed video transcoder" includes a source PC. EX1004, 1923. The source PC "divides in-coming MPEG-2 data with minimum duplication and the data are used to determine re-encoding parameters." *Id.*, 1923-24. The source PC then transmits video segments and their associated transcoding parameters (the re-encoding parameters) to a plurality of transcoding PCs to be transcoded. *Id.*, 1924. The transcoding parameters the source PC sends to a transcoding PC may include "filter function specifications, spatial resolution, temporal resolution, and re-encoding formats desired," which "specify the operation of the filter and the encoder of the transcoding PC." *Id.*; EX1003, ¶149.

To the extent that Sambe does not explicitly disclose that the source PC itself generates the "media metadata [that] comprises scene complexity information" during transcoding but prior to decoding, Vetro does. EX1003, ¶150.

Vetro discloses a computer system that generates media metadata from a source video file before it is decoded in the transcoding process. EX1005, 7:40-43, Fig. 3 (310); EX1003, ¶¶150-155. In Figure 3, Vetro's "[Content] Classifier" 310, shown in the green box below, generates media metadata that categorizes "shots" by their complexity during transcoding but prior to decoding. EX1005, 7:62-8:26, 17:32-19:58, Fig. 3 (annotated); EX1003, ¶151. The "[Content] Classifier" 310 receives content information from the input bitstream 301 and metadata information 303 that "include[s] descriptors and description schemes" as "specified by the emerging MPEG-7 standard." EX1005, 7:62-8:5, Fig. 3 (annotated). It then "map[s] semantic features of content characteristics, such as motion activity, video change information and texture, into a set of parameters that are used to make rate-quality trade-offs in the content network manager," and within these parameters, also known as semantic classes, the content is differentiated based on the coding complexity. *Id.,* 7:62-8:5*,* 8:60-9:5; EX1003, ¶¶150-153.

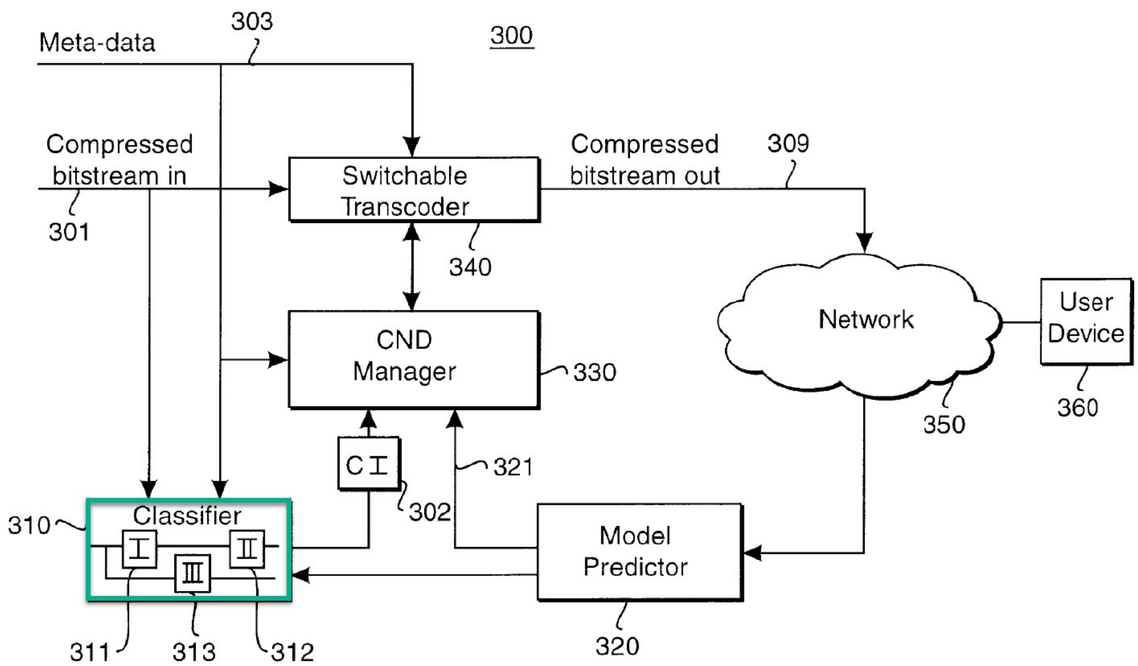


FIG. 3

Id., 7:40-43; Fig.3 (annotated).

A POSITA would have understood the parameters generated by the classifier constitute "media metadata" because they describe the associated media bitstream data. EX1003, ¶154 (citing EX1005, 7:62-8:5, 8:60-9:5). A POSITA would further understand that the generated parameters comprise scene complexity information. EX1003, ¶154. The generated parameters include the classification of the "shots" of Vetro—each of which constitutes a scene—into categories of similar motion activity and spatial distribution, which a POSITA would understand to be describing the complexity of the "shots." EX1005, 8:60-9:5, 18:24-19:58; EX1003, ¶154; *see also supra* § III.B.

A POSITA would also have understood that the content classifier generates the media metadata during transcoding but prior to decoding. EX1003, ¶155. Vetro discloses that the media metadata generated by Vetro's content classifier "in part, characterizes the *potential performance*" of Vetro's switchable transcoder. EX1005, 9:6-9 (emphasis added). Thus, a POSITA would have understood that the content classifier generates the media metadata as part of the transcoding process. EX1003, ¶155. The CND manager then processes and uses that media metadata to "pre-select the best strategy for the transcoder," such as determining whether to transcode the source video file using the discrete-summary transcoder or the continuous-conversion transcoder within Vetro's switchable transcoder. *Supra* § III.B; EX1005, 9:25-10:24, 19:53-58, 20:63-64, 21:8-15. Because the CND manager processes and uses the media metadata to determine which transcoder to select and the source video file is not decoded until it reaches the selected transcoder, a POSITA would understand that the media metadata is necessarily generated prior to decoding.

EX1003, ¶155; *see also* EX1005, Fig. 1 (depicting the decoder as part of the transcoder).

A POSITA would have been motivated to incorporate the content classifier portions of the Vetro transcoding system into Sambe's distributed transcoding architecture. Specifically, a POSITA would have known to incorporate Vetro's content classifier that generates parameters (i.e., media metadata), into Sambe's source PC, to determine how to allocate bits and ensure consistent video quality. EX1005, 21:8-57; EX1003, ¶156. In this combination, the source PC would use the parameters to determine a strategy for how to optimally transcode the compressed bitstream and send information describing that strategy to the transcoding PCs along with the video segments to be transcoded. EX1004, 1924; EX1003, ¶156. This process would ensure that any pre-processing of the source video, such as segmenting, determining transcoding parameters, media metadata generation, and determining transcoding strategy, occurs only once at the Sambe source PC, which reduces cost and improves efficiency. EX1003, ¶156.

This is a combination of prior art elements according to known methods to yield predictable results, as well as use of a known technique to improve a similar method in the same way. EX1003, ¶157; *see also supra* § II.A.3. As noted above, because the source PC in Sambe is already responsible for determining re-encoding parameters (transcoding parameters) from the source video file that "specify the operation of the filter and the encoder of the transcoding PC[s]" and transmitting them to the transcoding PCs, this would be a reasonable modification that provides the expected results—that of the content classifier of the source PC generating parameters (media metadata) used to determine how the source file will be transcoded. EX1004, 1923-24; EX1003, ¶157.

(3) [1.b] "providing information based on the media metadata from the computer system to a plurality of transcoding devices; and"

The combination of Sambe in view of Vetro renders limitation 1[b] obvious. EX1003, ¶¶158-164.

As noted above, the source PC of Sambe determines re-encoding parameters (transcoding parameters) from the source video file that "specify the operation of the filter and the encoder of the transcoding PC[s]" and transmits them to a plurality of transcoding PCs. *Supra* § IV.A.2(2). Alongside the transcoding parameters, the source PC also sends the video segments associated with the transcoding parameters to each of the plurality of transcoding PCs. EX1004, 1924.

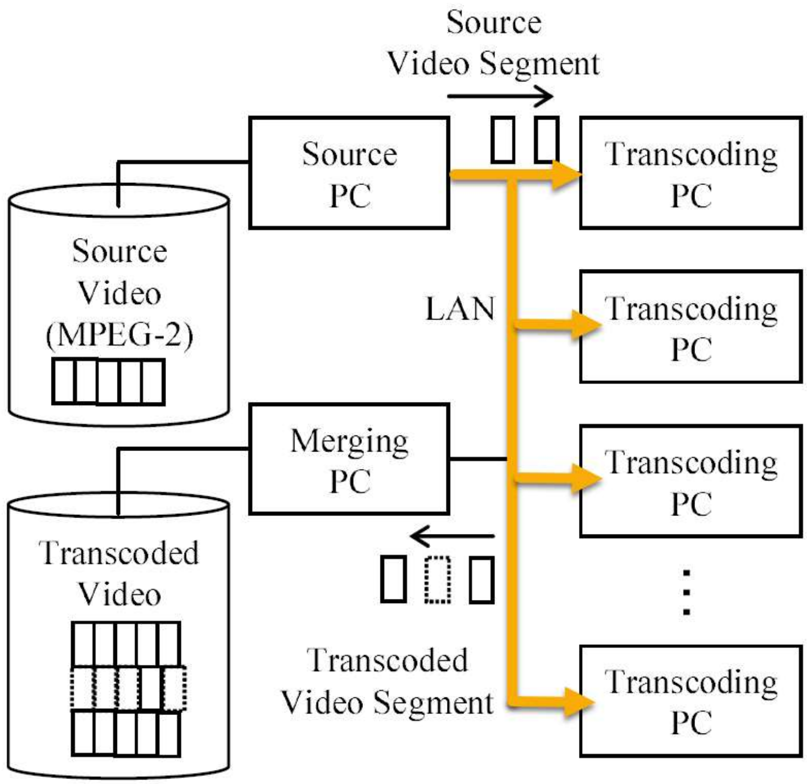


Fig. 1 Distributed video transcoder

EX1004, 1924, Fig. 1 (annotated).

To the extent that Sambe does not explicitly disclose providing information based on media metadata to the plurality of transcoding devices, Vetro does. EX1003, ¶¶160-162. Vetro discloses that information based on the generated metadata described in limitation [1.a] is provided to the transcoding devices. As noted above, the content classifier of Vetro generates media metadata (parameters). *See supra* § IV.A.2(2). The content classifier then, as shown below in Figure 4, transmits these parameters (green arrow) to the Content-Network Device Manager 330. EX1005, Fig. 4 (annotated). The Content-Network Device Manager 330 (yellow box) then processes the parameters to derive the transcoder's strategy and make real-time selection decisions regarding bitrate allocation and coding parameters. EX1005, 8:21-8:26, 21:8-22, Fig. 4 (annotated). This processed strategy information (purple arrow) is then passed to the switchable transcoder (red box) and guides the transcoding process for Vetro's "shots." EX1005, 8:21-8:26, Figs. 3, 4 (annotated); EX1003, ¶161.

For example, if the content classifier identifies a shot as simple to encode, such as a relatively static scene with an anchorman sitting behind a desk, it passes that classification to the CND manager. EX1005, 19:40-20:52. The CND manager then uses that information to "pre-select" the best transcoding strategy. *Id.* In the low-motion, simple example of the anchorperson, the CND will determine that the Discrete Summary Transcoder 441, which reduces the entire shot to a single frame, is the optimal option. *Id.*, 18:49-19:59. This strategy is passed to the transcoding switch, which directs the transcoding strategy. *Id.*; *see also id.*, 9:31-35. In this way, the media metadata is first processed and then used to determine how to allocate bits among the "shots" based on their relative complexities. EX1003, ¶162.

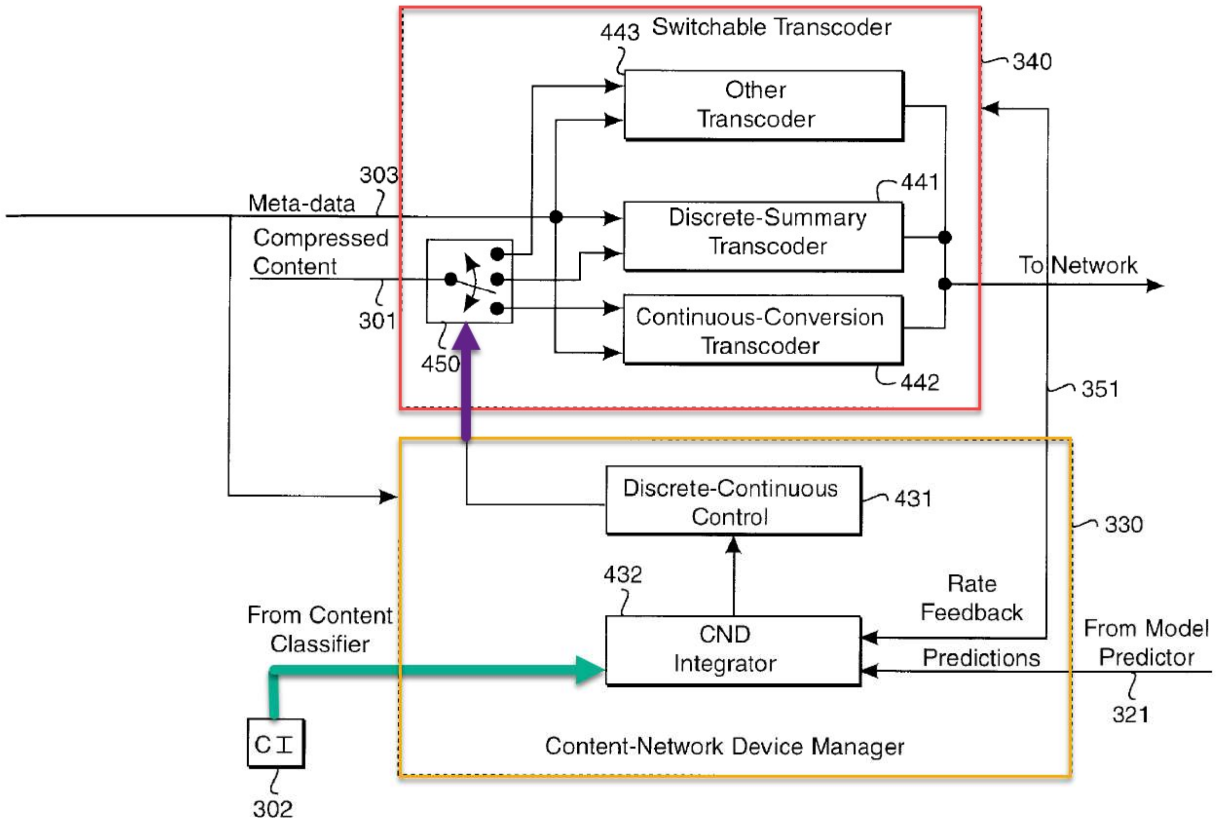


FIG. 4

EX1005, Fig. 4 (annotated); EX1003, ¶¶161-162.

A POSITA would have been motivated to incorporate the Vetro transcoding system into Sambe's distributed transcoding architecture, including by incorporating Vetro's CND manager into Sambe's source PC. EX1003, ¶163. In this combined configuration, the source PC with its incorporated content classifier and content-network device manager would process and send the information based on the media metadata (the transcoding strategy information) to the plurality of transcoding PCs along with the video segments to be transcoded. EX1004, 1924; EX1003, ¶163. A POSITA would understand, as explained in Vetro, that this information would permit the transcoders to determine how to allocate bits and ensure consistent video quality. *Id.* Further, this configuration would ensure that any pre-processing of the source video, such as segmenting, determining transcoding parameters, media metadata generation, and determining transcoding strategy, occurs once at the Sambe source PC. EX1003, ¶163.

This is a combination of prior art elements according to known methods to yield predictable results, as well as use of a known technique to improve a similar method in the same way. EX1003, ¶164; *see also supra* § II.A.3. As noted above, because the source PC in Sambe is already responsible for determining re-encoding parameters (transcoding parameters) from the source video file that "specify the operation of the filter and the encoder of the transcoding PC[s]" and transmitting them to the transcoding PCs, this would be a reasonable modification that provides the expected results—that of taking the strategy information derived from the content classifier and CND manager in Vetro to the transcoder into account when transcoding the input video. EX1003, ¶164.

(4) [1.c] "performing the following at each of the plurality of transcoding devices in parallel:"

The combination of Sambe in view of Vetro renders limitation 1[c] obvious. EX1003, ¶165. Sambe expressly discloses transcoding using multiple transcoding PCs in parallel. *Supra* § III.A; EX1004, 1924. Based on transcoding parameters sent from the source PC, each transcoding PC may perform a variety of different steps in parallel. *Id.*; EX1003, ¶165. The following subsections explain why the combination of Sambe and Vetro further discloses each of the claimed processes.

(5) [1.c(i)] "receiving the at least a portion of the source video file, including a first plurality of encoded images encoded according to a source format, from a media content source;"

Sambe discloses limitation [1.c(i)]. EX1003, ¶¶166-168.

In Sambe, the parallel transcoding PCs receive video segments from the source file to be transcoded, each of which is a portion of the source video file. EX1004, 1923-24, Figs. 1, 2. As Sambe explains, a "video segment consists of one or more consecutive Groups of Pictures (GOP)." *Id.*, 1924. And Sambe discloses a "***frame by frame*** based transcoding architecture" that decodes and encodes each frame of the GOP(s) in each video segment. *Id.*, 1924 (emphasis added). A POSITA would have recognized that the frames making up the GOPs comprise a plurality of encoded images. EX1003, ¶166.

A POSITA would also have understood that the encoded images in the received video segments are encoded according to a source format. *Id.*, ¶167. Sambe discloses transcoding from a source video format to a different target video format. EX1004, 1923. For example, Sambe describes converting "pre-encoded MPEG-2 digital video in archives to other compressed formats such as MPEG-1, MPEG-4, [and] H.263," and altering video content "in terms of bit-rate and resolution to meet network bandwidth and terminal capacity." *Id.*, 1923; *see also,* *id.,* Fig. 2, Table 1. A POSITA would have understood both the coding standard (e.g., MPEG-2) and the coding parameters (e.g., resolution and bitrate) of a source video to reflect the coding format. EX1003, ¶168; *see also* EX1001, 1:26-32 (defining format both in terms of coding standards and coding parameters such as bitrate and resolution).

(6) [1.c(ii)] "decoding the at least a portion of the source video file based on the source format to generate a decoded portion of video including a plurality of decoded images;"

Sambe discloses limitation [1.c(ii)]. EX1003, ¶¶169-171.

Sambe discloses that the parallel transcoders decode a portion of the source video file to generate a decoded portion of video including a plurality of decoded images. Sambe states, "[e]ach transcoding PC decodes and re-encodes the video segments into the different video formats specified. ***Decoded frames*** are filtered, resized, and passed to the encoder frame by frame." EX1004, 1924 (emphasis added). A POSITA would thus understand that the parallel transcoders decode the source video file into decoded frames and further that each decoded frame is a decoded image for the encoder to process. EX1003, ¶170. A POSITA would further understand that the decoded frames of the source video segments comprise a plurality of decoded images. EX1003, ¶170.

Sambe also discloses that the parallel transcoding PCs perform decoding based on the source format of the source video file. Sambe explains that the "video segments" are each a portion of the source video file, which has a source format. *See supra* § IV.A.2(6) (citing EX1004, 1925-26, Table 1); EX1003, ¶171. Because decoders must understand the specific rules and algorithms of the corresponding video segment encoding method to accurately decode the video segment, a POSITA would have understood that decoding of the video segments would be based on their source format. EX1003, ¶171. For example, Figure 2 of Sambe, a block diagram of a transcoding PC, includes an "MPEG-2 Decoder" because the distributed video transcoder of Sambe transcodes from the "MPEG-2" coding format to other coding formats, including but not limited to MPEG-1, MPEG-2, and MPEG-4. EX1004, 1923-24, Fig. 2.

(7) [1.c(iii)] "receiving the information based on the media metadata from the computer system; and"

The combination of Sambe in view of Vetro renders limitation [1.c(iii)] obvious. EX1003, ¶¶172-174.

As noted above for claim limitation [1.a.], the distributed video transcoder of Sambe includes a source PC that transmits video segments and their associated transcoding parameters to a plurality of transcoding PCs for parallel transcoding. *See supra* § IV.A.2(2). Further, as noted above for claim limitations [1.a.] and [1.b], in the combined system, this source PC is modified to include Vetro's content classifier and CND manager. *See supra* § IV.A.2(2)-(3). And within the source PC, the content classifier generates parameters (the claimed media metadata) comprising scene complexity information and the CND manager subsequently generates the transcoding strategy (the claimed information based on the media metadata) using the media metadata. *Id.* Then, as demonstrated below in Figure 4 of Vetro, once the CND manager determines the transcoding strategy (the claimed information based on the media metadata), it transmits it (purple arrow) to the Switchable Transcoder (red box) where it controls how the video segments (portion of the source video file) are transcoded—e.g., whether to transcode the video segment using the Discrete-Summary Transcoder or Continuous-Conversion Transcoder. EX1005, 21:8-14, Fig. 4; EX1003, ¶173.

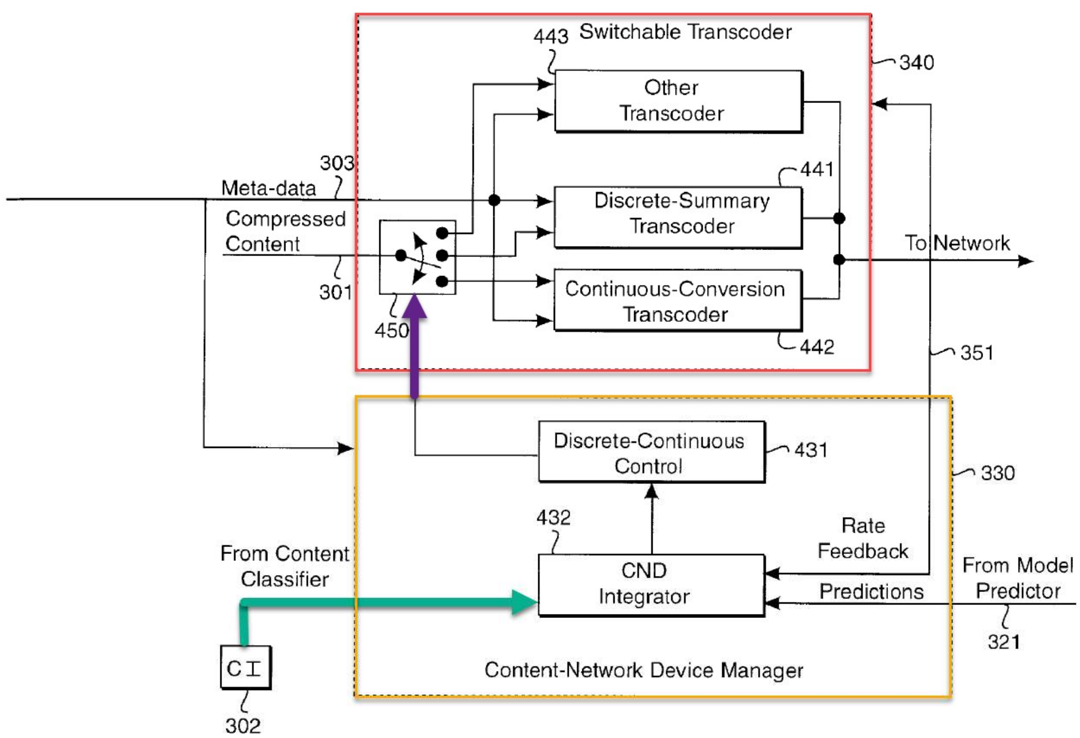


FIG. 4

EX1005, Fig. 4 (annotated).

In the combined system, as noted above the CND manager is in the source PC, and, thus, a POSITA would have understood that the source PC transmits the transcoding strategies (the claimed information based on the media metadata) to the plurality of transcoding PCs. EX1003, ¶174. Therefore, a POSITA would also have understood that in this combined system, each of the parallel transcoding PCs receive the transcoding strategy (information based on the media metadata) as noted above for limitation [1.b], from the source PC. EX1003, ¶174. Further, as noted above, a POSITA would have been motivated to make this combination to improve efficient allocation of bits and to ensure consistent video quality. A POSITA would have a reasonable expectation of success in making this combination, for the same reasons stated in limitation 1.b. *See supra* § IV.A.2(3).

(8) [1.c(iv)] "encoding the plurality of decoded images of the decoded portion of video into an alternate video stream including a second plurality of encoded images based on a target format and the information based on the media metadata, the alternate video stream being one of the set of multiple alternate video streams."

The combination of Sambe in view of Vetro renders limitation [1.c(iv)] obvious. EX1003, ¶¶175-177. Sambe discloses that each transcoding PC decodes the source video file based on the source format into a plurality of decoded images, referred to as frames, to prepare for encoding. *See supra* § IV.A.2(3).

Sambe further discloses that each transcoding PC encodes the decoded images of the decoded portion of video into an alternate video stream including a second plurality of encoded images based on a target format. Sambe discloses that after decoding, the transcoding PCs filter, resize, and pass the images to the encoder. EX1004, 1924. The plurality of transcoding PCs then "re-encodes the video segments into the different video formats specified . . . frame-by-frame." *Id.*; *see also* EX1003, ¶176. The encoded portion of the video is then sent to the merging PC where it (with the other encoded portions of the video) forms the desired video format file (the alternate video stream). *See* EX1004, 1923-24, Figs. 2, 17; EX1003, ¶176. A POSITA would, thus, have understood that the combined transcoding system encodes "the plurality of decoded images of the decoded portion of video into an alternate video stream including a second plurality of encoded images based on a target format." EX1003, ¶176. Additionally, as noted above, Sambe discloses preparing a set of alternate video streams of the source video file. *See supra* § IV.A.2(3). A POSITA would, thus, have also understood that each alternate video stream Sambe encodes is one of the set of multiple alternate video streams. EX1003, ¶176.

Additionally, as noted above, in the combined system, each of the transcoding PCs receive transcoding strategy (information based on the media metadata) from the source PC of Sambe through Vetro's content classifier and CND manager. *See supra* § IV.A.2(7). The transcoding strategy informs the transcoding of the received video segments. EX1003, ¶177; *see also, e.g.,* EX1004, 1924; EX1005, 7:31-38, 7:40-43, 8:21-26, 9:26-39, Fig. 4 (340). Further, as noted above, a POSITA would have been motivated to make this combination to improve efficient allocation of bits and to ensure consistent video quality and would have a reasonable expectation of success in making this combination, for the same reasons stated in limitation 1.b. *See supra* § IV.A.2(3).

3. Claim 11

(1) [11.pre]: A system for transcoding video data, the system comprising:

To the extent the preamble is considered limiting, both Sambe and Vetro individually render it obvious for the same reasons discussed regarding limitation [1.pre]. *Supra* § IV.A.2(1); EX1003, ¶201.

(2) [11.a]: a computer system configured as a media metadata generation device, wherein the computer system is configured to:

The combination of Sambe in view of Vetro renders limitation [11.a] obvious for the same reasons as discussed regarding limitation [1.a]. *Supra* § IV.A.2(2); EX1003, ¶202.

(3) [11.a(i)]: generate media metadata related to the source video file prior to decoding, during a transcoding of, at least a portion of the source video file, where the media metadata comprises scene complexity information; and

The combination of Sambe in view of Vetro renders limitation [11.a] obvious for the same reasons as discussed regarding limitation [1.a]. *Supra* § IV.A.2(2); EX1003, ¶203.

(4) [11.a(ii)] provide information based on the media metadata to a plurality of transcoding devices; and

The combination of Sambe in view of Vetro renders limitation [11.a(ii)] obvious for the same reasons as discussed regarding limitation [1.b]. *Supra* § IV.A.2(3); EX1003, ¶204.

(5) [11.b]: the plurality of transcoding devices, configured to perform the following at each of the plurality of transcoding devices in parallel:

The combination of Sambe in view of Vetro renders limitation [11.b] obvious for the same reasons as discussed regarding limitation [1.c]. *Supra* § IV.A.2(4); EX1003, ¶205.

(6) [11.b(i)]: receive the at least a portion of the source video file, including a first plurality of encoded images encoded according to a source format, from a media content source;

Sambe discloses limitation [11.b(i)] for the same reasons as discussed regarding limitation [1.c(i)]. *Supra* § IV.A.2(5); EX1003, ¶206.

(7) [11.b(ii)]: decode the at least a portion of the source video file based on the source format to generate a decoded portion of video including a plurality of decoded images;

Sambe discloses limitation [11.b(ii)] for the same reasons as discussed regarding limitation [1.c(ii)]. *Supra* § IV.A.2(6); EX1003, ¶207.

(8) [11.b(iii)]: receive the information based on the media metadata from the computer system; and

The combination of Sambe in view of Vetro renders limitation [11.b(iii)] obvious for the same reasons as discussed regarding limitation [1.c(iii)]. *Supra* § IV.A.2(7); EX1003, ¶208.

(9) [11.b(iv)]: encode the plurality of decoded images of the decoded portion of video into an alternate video stream including a second plurality of encoded images based on a target format and the information based on the media metadata, the alternate video stream being one of the set of multiple alternate video streams.

The combination of Sambe in view of Vetro renders limitation [11.b(iv)] obvious for the same reasons as discussed regarding limitation [1.c(iv)]. *Supra* § IV.A.2(8); EX1003, ¶209.

4. Claim 21

(1) [21.pre]: "A method for transcoding a source video file into a set of multiple alternate video streams, the method comprising:"

To the extent the preamble is considered limiting, both Sambe and Vetro individually render it obvious for the same reasons discussed regarding limitation [1.pre]. *Supra* § IV.A.2(1); EX1003, ¶227.

(2) [21.a]: "generating, at a computer system configured as a media metadata generation device, media metadata related to the source video file prior to decoding, during a transcoding of, at least a portion of the source video file, where the media metadata comprises scene change information indicating the start and end of a scene, and scene complexity information;"

The combination of Sambe in view of Vetro renders obvious "generating, at a computer system configured as a media metadata generation device, media metadata related to the source video file prior to decoding, during a transcoding of, at least a portion of the source video file, where the media metadata comprises . . . scene complexity information" for the same reasons as discussed regarding limitation [1a]. *Supra* § IV.A.2(2); EX1003, ¶228. The combination of Sambe in view of Vetro also renders obvious "where the media metadata comprises scene change information indicating the start and end of a scene" for the same reasons as discussed below regarding claim 4. *Infra* § IV.A.7; EX1003, ¶228.

(3) [21.b]: "providing information based on the media metadata from the computer system to a plurality of transcoding devices; and"

For the reasons discussed above for limitation 1[b] and claim 4, the combination of Sambe in view of Vetro renders limitation [21.b] obvious. *Supra* § IV.A.2(3); *Infra* § IV.A.7; EX1003, ¶229.

(4) [21.c]: "performing the following at each of the plurality of transcoding devices in parallel:"

For the reasons discussed above for limitation 1[c] and claim 4, the combination of Sambe in view of Vetro renders limitation [21.c] obvious. *Supra* § IV.A.2(4); *Infra* § IV.A.7; EX1003, ¶230.

(5) [21.c(i)]: "receiving the at least a portion of the source video file, including a first plurality of encoded images encoded according to a source format, from a media content source;"

The combination of Sambe in view of Vetro renders limitation [21.c(i)] obvious for the same reasons as discussed regarding limitation [1.c(i)]. *Supra* § IV.A.2(5); EX1003, ¶231.

(6) [21.c(ii)]: "decoding the at least a portion of the source video file based on the source format to generate a decoded portion of video including a plurality of decoded images;"

The combination of Sambe in view of Vetro renders limitation [21.c(ii)] obvious for the same reasons as discussed regarding limitation [1.c(ii)]. *Supra* § IV.A.2(6); EX1003, ¶232.

(7) [21.c(iii)]: "receiving the information based on the media metadata from the computer system;"

For the reasons discussed above for limitation [1.c(iii)] and claim 4, the combination of Sambe in view of Vetro renders limitation [21.c(iii)] obvious. *Supra* § IV.A.2(7); *Infra* § IV.A.7; EX1003, ¶233.

(8) [21.c(iv)]: "dividing an image in the plurality of decoded images into a plurality of coding units based on a target format, wherein the source format and the target format have different resolutions;"

The combination of Sambe in view of Vetro renders "dividing an image in the plurality of decoded images into a plurality of coding units based on a target format" obvious for the same reasons as discussed regarding limitation [10.a]. *Infra* § IV.A.11(2); EX1003, ¶234. The combination of Sambe in view of Vetro renders "wherein the source format and the target format have different resolutions" obvious for the same reasons as discussed regarding claim 8. *Infra* § IV.A.9; EX1003, ¶234. Thus, the combination of Sambe in view of Vetro renders limitation [21.c(iv)] obvious for the same reasons as discussed regarding limitations [10.a] and claim 8. EX1003, ¶234.

(9) [21.c(v)]: "determining a number of bits to encode a group of pictures (GOP) based at least in part on a number of frames between the start and end of a scene as indicated by the information based on the media metadata; and"

For the reasons discussed for limitation [7.a] and claim 4, the combination of Sambe in view of Vetro renders limitation [21.c(v)] obvious. *Infra* §§ IV.A.8(2), IV.A.7; EX1003, ¶235.

(10) [21.c(vi)]: "encoding the plurality of decoded images of the decoded portion of video into an alternate video stream including a second plurality of encoded images based on the target format and the information based on the media metadata, the alternate video stream being one of the set of multiple alternate video streams."

For the reasons discussed for limitation [1.c(iv)] and claim 4, the combination of Sambe in view of Vetro renders limitation [21.c(vi)] obvious. *Supra* § IV.A.2(8); *Infra* § IV.A.7; EX1003, ¶236.

5. Claims 2 and 12: "The [method / system] of claim [1 / 11], wherein a group of the plurality of transcoding devices perform processes in parallel for the same alternate video stream from the set of multiple alternate video streams."

As discussed above, Sambe in combination with Vetro discloses every limitation of claims 1 and 11, from which claims 2 and 12 depend. *See supra* §§ IV.A.2, 3.

As noted above, parallel transcoding was well-known as one way to speed up transcoding. *See supra* § II.A.3; EX1003, ¶179. Sambe also discloses the limitations of claims 2 and 12. EX1003, ¶¶178-180, 210-211. For example, Sambe describes a test where its source PC "implement[ed] a simple round-robin scheduling method" in which the source PC divides input video into segments (e.g., segments 1, 2, and so on). EX1004, 1929; *see also id.*, Fig. 17. As demonstrated below in Figure 17, Sambe discloses that segments 1 (red) and 2 (green) are sent to Transcoding PC1 and Transcoding PC2 respectively, and Transcoding PC1 and Transcoding PC2 encode segments 1 and 2 in parallel. EX1004, 1923-24, 1929, Fig. 17 (modified); EX1003, ¶¶179-180. Segments 1 and 2, and the remaining segments of the video file, are sent to the Merging PC as they are transcoded, and "[a]fter the second segment for each desired file is received, the concatenation process begins and runs in parallel to the following segment transcoding." *See id.*, 1923-24, Figs. 2, 17; EX1003, ¶180. A POSITA would have understood that the Merging PC stitches the segments together into the same alternative stream (e.g., at a specific resolution and/or bit rate), referred to as the desired file, and this alternative stream is one of the set of multiple alternate video streams. EX1003, ¶180.

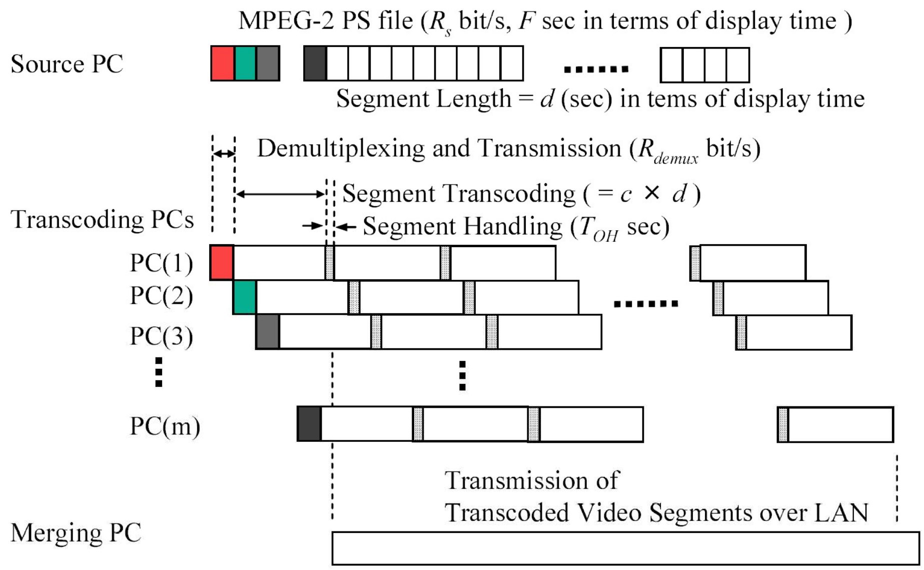


Fig. 17 Process flow of the distributed transcoding

EX1004, Fig. 17 (annotated).

6. Claims 3 and 13: "The [method / system] of claim [1 / 11], wherein a group of the plurality of transcoding devices perform processes in parallel for different alternate video streams in the set of multiple alternate video streams."

As discussed above, Sambe in combination with Vetro discloses every limitation of claims 1 and 11, from which claims 3 and 13 depend. *See supra* § IV.A.2, 3.

As noted above, parallel transcoding was well-known as one way to speed up transcoding. *See supra* § II.A.3; EX1003, ¶182. Sambe also discloses the limitations of claims 3 and 13. EX1003, ¶¶181-182, 212-213. For example, Sambe discloses that "[a]ll of the modules shown in [Figure 2], including transmission modules, are implemented by multi-thread programming." EX1004, 1924, Fig. 2. This allows both transmission from the source PC to the plurality of transcoding PCs, and subsequent processing by the transcoding PCs, to run in parallel. *Id.*, 1924. Also, as noted above for the preamble of claim 1, Sambe "describes a distributed video transcoding system that can *simultaneously* transcode an MPEG-2 video file into various video coding formats with different rates." *Id.*, 1923 (emphasis added). A POSITA would have understood that for Sambe to *simultaneously* transcode a video file into various coding formats with different rates, it must perform processes in parallel for different alternative video streams in the set of multiple alternate video streams. EX1003, ¶182.

7. Claims 4 and 14: "The [method / system] of claim [1 / 11], wherein the media metadata further comprises scene change information indicating the start and end of a scene."

As discussed above, Sambe in view of Vetro teaches every limitation of claims 1 and 11, from which claims 4 and 14 depend. *See supra* §§ IV.A.2, 3.

Vetro renders obvious media metadata comprising scene change information indicating the start and end of a scene. EX1003, ¶¶183-184, 214-215. Vetro discloses partitioning a video into "shots," or "smaller segments of video that begin when a camera is turned [on] and last until the camera is turned off." EX1005, 17:41-46. As was noted above, a POSITA would have understood the "shots" of Vetro to be scenes as claimed. *Supra* § III.B. Further, each "shot" of Vetro is categorized by its coding complexity, informing how it will be transcoded. *See, e.g.,* EX1005, 20:27-29. A POSITA would have understood that to use a "shot's" complexity information during transcoding, the Vetro system must maintain metadata describing where that "shot" starts and ends. EX1003, ¶184.

8. Claims 7 and 17

(1) Claims [7.pre] and [17.pre] "The [method / system] of claim [4 / 14], [by further performing / wherein the plurality of transcoding devices are configured to further perform] the following at each of the plurality of transcoding devices in parallel:"

As discussed above, Sambe in view of Vetro teaches every limitation of claims 4 and 14, from which claims 7 and 17 depend. *See supra* § IV.A.7. To the extent the preambles of claims 7 and 17 are limitations, the combination of Sambe in view of Vetro also discloses them. EX1003, ¶¶185, 216-217. As explained above, Sambe describes a system including transcoding devices, or PCs, that perform transcoding processes in parallel. *See supra* §§ III.A, IV.A.2.

(2) Claims [7.a] and [17.a] "determining a number of bits to encode a group of pictures (GOP) based at least in part on a number of frames between the start and end of a scene as indicated by the information based on the media metadata."

As discussed above, Sambe in view of Vetro teaches every limitation of claims 4 and 14, from which claims 7 and 17 depend. *See supra* § IV.A.7. The combination of Sambe in view of Vetro also discloses claim limitations [7.a] and [17.a]. EX1003, ¶¶186-187, 218.

As noted above, a POSITA would have understood that Vetro's teaching of generating "shot" level complexity information constitutes generating media metadata further comprising scene change information indicating the start and end of a scene. *Supra* § IV.A.7. Vetro also discloses this limitation. Vetro discloses that each "shot" can be a group of frames. EX1005, 21:35-36. As noted above, a POSITA would have understood each frame to constitute an "image" or "picture." *Supra* § IV.A.2(6). A POSITA would thus have understood that Vetro's "shots" can be a "group of pictures" as recited in this limitation. EX1003, ¶ 187. Vetro explains that "shot" level metadata, described above, "are used for classification, bit allocation and rate-quality considerations for that particular shot." EX1005, 17:65-18:1. For example, when a user requests multiple "shots" at the same time, the CND manager must determine how to allocate the bit rate amongst the "shots." *Id.*, 21:33-35. And for the discrete-summary transcoder, "this [bit] rate can correspond to *the number of frames* that are sent." *Id.*, 21:35-36 (emphasis added). A POSITA would have understood at the time of invention that determining the rate, or bit rate, of a "shot" based on its number of frames requires using the media metadata indicating the start and end of the scene to count the frames in between them (i.e, information based on the media metadata). EX1003, ¶187.

9. Claims 8 and 19: "The [method / system] of claim [1 / 11], wherein the source format and the target format have different resolutions."

As discussed above, Sambe in combination with Vetro discloses every limitation of claims 1 and 11, from which claims 8 and 19 depend. *See supra* §§ IV.A.2, 3.

Sambe discloses the limitations of claims 8 and 19. EX1003, ¶¶188-190, 223-224. For example, Sambe explains that due to modern terminals using displays with various sizes and resolutions, "it is often necessary for service providers delivering video content over the Internet to transcode the same content to yield different video formats, *spatial resolution*, and bit-rates simultaneously." EX1004, 1923 (emphasis added). As shown below in Table 1, Sambe also provides example transcoding conditions in which the source video resolution (720 x 480) differs from the output video resolution (360 x 240). *Id.*, 1926 (Table 1).

Table 1 Experimental transcoding conditions

Source Video Format	MPEG-2 MP@ML
Source Video	*football, flower garden mobile & calendar, sailboat*
Source Video Rate	8 Mbit/s
Source Frame Size	Horizontal 720 pixels, Vertical 480 lines
Output Video Coding	MPEG-2, MPEG-4 ASP
Output Video Rate	MPEG-2: 2 Mbps, 1 Mbps MPEG-4: 1 Mbps, 750 Kbps
Output Frame Size	Horizontal 360 pixels, Vertical 240 lines
GOP/GOV Structure	M=3, N=15

EX1004, Table 1 (annotated).

A POSITA would thus have understood that Sambe discloses a transcoding method where the source format and the target format may have different resolutions. EX1003, ¶¶190.

10. Claims 9 and 20: "The [method / system] of claim [1 / 11], wherein the source format and the target format correspond to different video encoding standards."

As discussed above, Sambe in combination with Vetro discloses every limitation of claims 1 and 11, from which claims 9 and 20 depend. *See supra* §§ IV.A.2, 3.

Sambe also teaches the limitations of claims 9 and 20. EX1003, ¶¶191-193, 225-226. For example, Sambe discusses the growing demand "to convert [] pre-encoded MPEG-2 digital video in archives to other compressed formats such as

MPEG-1, MPEG-4, H.263 and so on." EX1004, 1923. Also for example, in the distributed transcoding system of Sambe, "[e]ach transcoding PC decodes and re-encodes the video segments *into the different video formats specified*." *Id.*, 1924 (emphasis added). As shown below, Figure 2 depicts a such a transcoding PC consisting of encoder modules (MPEG-1 Encoder, MPEG-2 Encoder, and the MPEG-4 Encoder) that enable the transcoding PC to encode several different video formats. *Id.*; EX1003, ¶192.

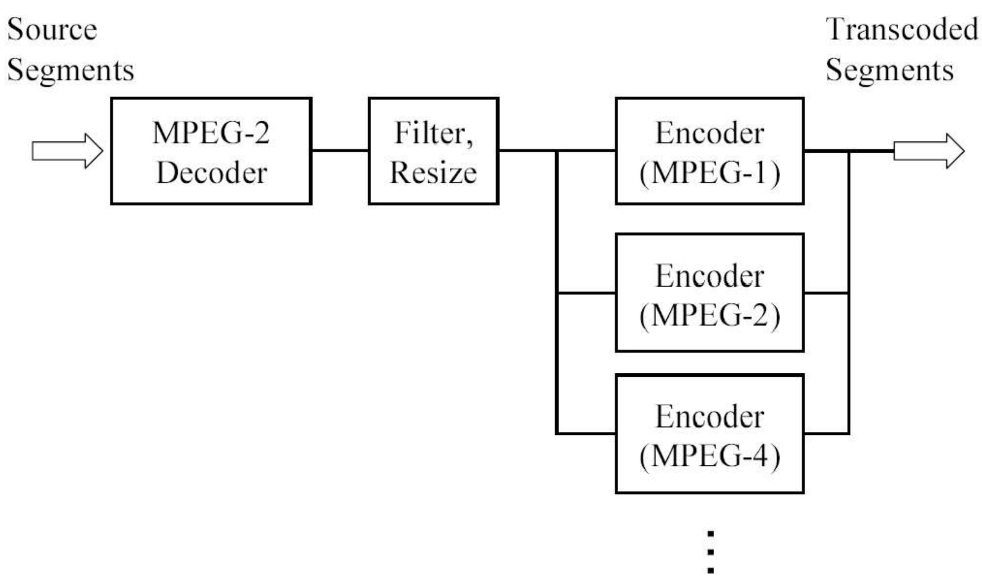


Fig. 2 Block diagram of a transcoding PC

EX1004, Fig. 2.

In Table 1 below, Sambe also provides example transcoding conditions, in which the source video format (MPEG-2 MP@ML) differs from at least one of the output video formats (e.g., MPEG-4). *Id.*, 1926 (Table 1) (modified).

Table 1 Experimental transcoding conditions

Source Video Format	MPEG-2 MP@ML
Source Video	*football, flower garden* *mobile & calendar, sailboat*
Source Video Rate	8 Mbit/s
Source Frame Size	Horizontal 720 pixels, Vertical 480 lines
Output Video Coding	MPEG-2, MPEG-4 ASP
Output Video Rate	MPEG-2: 2 Mbps, 1 Mbps MPEG-4: 1 Mbps, 750 Kbps
Output Frame Size	Horizontal 360 pixels, Vertical 240 lines
GOP/GOV Structure	M=3, N=15

EX1004, Table 1 (annotated).

A POSITA would have understood that the source and output video "format" and "coding" refer to different formats corresponding to different video encoding standards. EX1003, ¶193.

11. Claims 10 and 18

(1) Claims [10.pre] and [18.pre] "The [method / system] of claim [1 / 11], [by further performing / wherein the plurality of transcoding devices are configured to further perform] the following at each of the plurality of transcoding devices in parallel:"

As discussed above, Sambe in combination with Vetro discloses every limitation of claims 1 and 11, from which claims 10 and 18 depend. *See supra* §§ IV.A.2, 3.

To the extent the preambles of claims 10 and 18 are considered limiting, in addition to teaching every limitation of claims 1 and 11, the combination of Sambe and Vetro further teaches the preambles of claims 10 and 18. EX1003, ¶¶194-195, 219-220. As explained above, Sambe describes a system including transcoding devices, or PCs, that perform transcoding processes in parallel. *See supra* §§ III.A, IV.A.2.

(2) Claims [10.a] and [18.a] "dividing an image in the plurality of decoded images into a plurality of coding units based on the target format."

As discussed above, Sambe in combination with Vetro discloses every limitation of claims 1 and 11, from which claims 10 and 20 depend. *See supra* §§ IV.A.2, 3.

Sambe also teaches limitations [10.a] and [18.a]. EX1003, ¶¶196-200, 221-222. As noted above for limitation [1.c(ii)], the transcoding PCs of Sambe decode the source video segments *frame by frame*, and each decoded frame is a decoded image. *See supra* § IV.A.2(6). Because each of the coding standards disclosed in Sambe (e.g., MPEG-2, MPEG-2 MP@ML, MPEG-4, MPEG-4 ASP) require dividing frames into macroblocks, a POSITA would have understood that each of these decoded frames are subsequently divided into blocks and/or macroblocks for processing purposes. EX1004, 1925; EX1003, ¶¶197-98.

A POSITA would have understood that the blocks or macroblocks of each frame are a plurality of coding units. EX1003, ¶199. As noted above, in video encoding, a coding unit is a block of pixels used as the basic building block for processing and compression. *See supra* § II.A.1; EX1003, ¶199; *see also, e.g.*, EX1014, 37-38 (disclosing that before a discrete cosine transform is performed, a frame is broken down into blocks), EX1019, 17 (disclosing that for motion compensated interframe prediction, a frame is broken down into macroblocks). The blocks and/or macroblocks of Sambe each serve as the basic building blocks for processing and compression—e.g., each block is individually quantized—and thus constitute "coding units." EX1004, 1925; EX1003, ¶200. Further, a POSITA would have understood that each block and/or macro block coding unit is based on the target format, because the target formats determine the required coding units for the format (e.g., MPEG-2 requires macro blocks of size 16x16, while HEVC can have coding units of size 64x64). EX1016, 883; EX1029, 1651; EX1003, ¶200.

B. Ground 2: Claims 5-6 and 15-16 are obvious over Sambe, Vetro, Gu, and the General Knowledge of a POSITA

1. A POSITA would have combined Sambe and Vetro with Gu

As was noted above, a POSITA would have combined the distributed video transcoder of Sambe with the teachings of the Vetro transcoder to take advantage of both the increase in transcoding speed offered by Sambe's distributed transcoding architecture, and improved understanding of coding complexity offered by the transcoding techniques disclosed in Vetro. *Supra* § IV.A.1; EX1003, ¶¶124-138. A POSITA would also have been motivated to combine the above-described combination of Sambe and Vetro with the encoding techniques taught by Gu. EX1003, ¶¶139-142.

Sambe, Vetro, and Gu also all disclose using a version of the video coding process described above, *e.g.*, compressing the original video to allow for efficient transmission and storage by using predictive techniques, transforms, quantization, and entropy encoding to reduce spatial and temporal redundancy. *Supra* § II.A.1; EX1004, 1923, 1925; EX1005, 2:38-3:10, 12:4-16:67, 22:6-9; EX1006, [0031]–[0038]; EX1003, ¶140. Gu describes these steps explicitly, but a POSITA would have understood these fundamental video coding steps to be implicit in Sambe and Vetro. EX1003, ¶140. For example, Sambe discloses converting video into the MPEG-2 and MPEG-4 ASP coding standards, and both standards require generating prediction images, performing transforms on residual images of coding units to generate sets of transform coefficients, performing quantization on the resulting sets of transform coefficients, and performing entropy encoding on sets of quantized transform coefficients, as part of their respective coding processes. EX1003, ¶¶140-141. Even so, a POSITA could look to Gu to clarify these coding steps of Sambe and Vetro. *Id.*, ¶141-142.

2. Claims 5 and 15

The combination of Sambe in view of Vetro and Gu render the method of Claim 5 and the apparatus of claim 15 obvious. EX1003, ¶¶237-253, 266-270. As discussed above, Sambe in view of Vetro teaches every limitation of claims 1 and 11. *See supra* §§ IV.A.2, 3. And, as noted above, a POSITA would have looked to combine Gu with the combination of Sambe and Vetro. *See supra* § IV.B.1.

(1) Claims [5.pre] and [15.pre] "The [method / system] of claim [1 / 11], wherein the encoding of the plurality of decoded images of the decoded portion of video into an alternate video stream includes performing at least the following operations for images in the plurality of decoded images:"

To the extent the preambles of method claim 5 and apparatus claim 15 are limiting, the combination of Sambe in view of Vetro and Gu disclose it. EX1003, ¶¶238, 266-267. As noted above, Sambe in view of Vetro disclose encoding the plurality of decoded images of the decoded portion of video into an alternate video stream. *See supra* § IV.A.2(8).

(2) Claims [5.a] and [15.a] "generating a prediction image for each of a plurality of coding units of an image in the plurality of decoded images using the scene change information and the scene complexity information within the information based on the media metadata according to the target format;"

The combination of Sambe in view of Vetro and Gu renders method claim [5.a] and apparatus claim [15.a] obvious. EX1003, ¶¶239-243, 268.

As noted above, Sambe in view of Vetro discloses encoding the plurality of decoded images "based on a target format and the information based on the media metadata," where the media metadata comprises scene complexity information and scene change information. *See supra* §§ IV.A.2(8), IV.A.7. Further, as noted above, the blocks and/or macro blocks described in Sambe are basic building blocks, or coding units, used by the transcoding PCs when they process (i.e., encode) video data. *See supra* § IV.A.11. At the time of invention, generating a prediction image for each of a plurality of coding units of an image in the plurality of decoded images was a widely understood and used step in the encoding process. *See supra* § II.A.1; EX1003, ¶241. Indeed, despite not explicitly saying so, both Sambe and Vetro implicitly perform this step as part of their respective transcoding processes. EX1003, ¶242. For example, Sambe discloses converting video into MPEG-2 and MPEG-4 ASP standards which both require generating prediction images as part of their respective encoding processes. EX1004, Table 1; EX1003, ¶242. A POSITA would have, thus, understood that Sambe in view of Vetro discloses "generating a prediction image for each of a plurality of coding units of an image in the plurality of decoded images using the scene change information and the scene complexity information within the information based on the media metadata according to the target format." EX1003, ¶242.

But even if generating a prediction image for each of a plurality of coding units of an image in the plurality of decoded images is not expressly disclosed in Sambe or Vetro, it is in Gu, and a POSITA would have been motivated to look to Gu for more explicit detail about the claimed standard encoding functionalities.

EX1003, ¶243. As noted above, the blocks and/or macro blocks described in Sambe are basic building blocks, or coding units, used by the transcoding PCs when they process video data. *See supra* § IV.A.11. The same is true for Gu. EX1003, ¶242. Gu describes generating prediction images for each of a plurality of blocks and/or macroblocks using associated scene complexity information, including motion information. EX1006, [0031]-[0033]; EX1003, ¶243. To do so, the video encoder of Gu first performs "motion estimation for the inter-path video content 110 with motion estimator 258." EX1006, [0032]. Gu explains that "[t]he motion estimator 258 estimates motion of macroblocks or other sets of samples of the video picture with respect to one or more reference pictures, which represent reconstructions of previously encoded video content frames." *Id.* As a result, "[t]he motion estimator 258 outputs motion information 260 such as motion vector information." *Id.* Further, Gu explains that "[t]he motion compensator 262 applies [the] motion vectors to certain reconstructed video content 266 (stored as reference picture(s)) when forming a motion-compensated current picture 268." *Id.*, [0033]; EX1003, ¶243. A POSITA would have been aware that this processing would occur for each of the coding units forming an image. *See, e.g.*, EX1006, [0031], [0035]; EX1003, ¶243. And because generating prediction images for individual coding units was the norm in encoding at the time of invention, it would also have been obvious to a POSITA that Gu generates a prediction image for each block and/or macroblock of the "motion-compensated current picture." EX1003, ¶243; *see supra* § II.A.1; EX1019, 17, 19, Fig. 8.

(3) Claims [5.b] and [15.b] "performing transforms on residual images of the plurality of coding units to generate sets of transform coefficients based on the target format; and"

The combination of Sambe in view of Vetro and Gu renders method claim [5.b] and apparatus claim [15.b] obvious. EX1003, ¶¶244-249, 269.

As noted above, Sambe in view of Vetro discloses encoding the plurality of decoded images "based on a target format." *See supra* § IV.A.2(8). Further, as noted above, the blocks and/or macro blocks described in Sambe are basic building blocks, or coding units, used by the transcoding PCs when they process (i.e., encode) video data. *See supra* § IV.A.11. At the time of invention, performing transforms on residual images of the plurality of coding units was a widely used step in the encoding process. *See supra* § II.A.1; EX1003, ¶245; *see also* EX1019, 14. Indeed, despite not explicitly saying so, both Sambe and Vetro implicitly perform this step as part of their respective encoding processes. EX1003, ¶246. For example, Sambe discloses converting video into MPEG-2 and MPEG-4 ASP standards. Both require performing transforms on residual images of coding units to generate transform coefficients as part of their respective encoding processes. *Id.*, ¶246. A POSITA would have, thus, understood that Sambe in view of Vetro discloses "performing transforms on residual images of the plurality of coding units to generate sets of transform coefficients based on the target format." *Id.*, ¶246.

But even if performing transforms on residual images of the plurality of coding units is not expressly disclosed in Sambe or Vetro, it is in Gu, and a POSITA would have been motivated to look at Gu to provide more explicit detail about this standard encoding functionality. *Id*, ¶247. The video encoder of Gu generates residual images of the plurality of coding units. *Id.* The encoder first generates a prediction image, referred to as a "motion-compensated current picture." EX1006, [0033]. The encoder then determines "[t]he difference (if any) between a block of the motion-compensated picture 268 and corresponding block of the original inter-path video picture" and identifies this difference as "the prediction residual 270 for the block." *Id.*

The frequency transformer, which is a component of the video encoder, then performs transforms on the residual prediction images of the plurality of coding units (blocks or macroblocks). EX1006, [0034]; EX1003, ¶248. Gu discloses that the frequency transcoder processes "residual images of the plurality of coding units," referred to as the "blocks of … prediction residual data," to generate sets of transform coefficients. EX1006, [0034]. Specifically, Gu discloses that "[f]or block-based video content, the frequency transformer 280 *applies a DCT, variant of DCT, or other forward block transform to blocks of ... prediction residual data*, producing blocks of frequency transform coefficients." *Id.* (emphasis added). A POSITA would have understood "*DCT*" as a discrete cosine transform—among the most effective transforms to reduce bit rate—and would have understood that performing a DCT is "performing a transform" as recited in the claim limitation. *Id.* (emphasis added); EX1019, 14; EX1003, ¶249.

(4) Claims [5.c] and [15.c] "performing entropy encoding on the sets of transform coefficients to generate images for the second plurality of encoded images."

The combination of Sambe in view of Vetro and Gu renders method claim [5.c] and apparatus claim [15.c] obvious. EX1003, ¶¶250-253, 270.

As noted above, Sambe in view of Vetro discloses encoding the plurality of decoded images. *See supra* § IV.A.2(8). At the time of invention, performing entropy encoding on the sets of transform coefficients to generate images for the second plurality of encoded images was a widely understood and used step in the encoding process. *See supra* § II.A.1; *see also* EX1019, 15-16; EX1014, 48. This is because entropy encoding minimizes the amount of data that needs to be transmitted over a network, which is particularly important for applications like streaming video, where efficient bandwidth usage is crucial. *Id.* Thus, despite not explicitly describing it, both Sambe and Vetro implicitly perform this step as part of their respective encoding processes. EX1003, ¶251. For example, Sambe discloses converting video into MPEG-2 and MPEG-4 ASP standards. Both require performing entropy encoding on sets of transform coefficients as part of their respective encoding processes. *Id.*, ¶251. A POSITA would have, thus, understood that Sambe in view of Vetro discloses "performing entropy encoding on the sets of transform coefficients to generate images for the second plurality of encoded images." *Id.*, ¶251.

But even if performing transforms on residual images of the plurality of coding units is not expressly disclosed in Sambe or Vetro, it is in Gu, and a POSITA would have been motivated to look at Gu to provide more explicit detail about this standard encoding functionality. *Id.*, ¶252. Gu discloses performing entropy encoding on the sets of transform coefficients to generate images for the second plurality of encoded images. As noted above, the frequency transformer of Gu, a component of the video encoder, processes "residual images of the plurality of coding units," referred to as the "blocks of … prediction residual data" to generate sets of transform coefficients. *See supra* § IV.B.2(3). Each set of transform coefficients passes through a quantizer which compresses them into transform coefficients based on a certain step size. EX1006, [0035]–[0036], Fig. 2. A POSITA would have understood that quantization compresses values into a discrete set of values that compresses them without changing their overall nature and relationship. EX1003, ¶252; *see also* EX1014, 40 (explaining that quantization is simply dividing the block of transform coefficients by a weighted quantization matrix to reduce the number of bits that must be sent). Next, the entropy coder 284 of Gu, another component of the video encoder of Gu, performs entropy encoding on "the output of the quantizer 282," which, a POSITA would understand to comprise performing entropy encoding on the sets of transform coefficients. EX1003, ¶252; *supra* § III.C; EX1006, [0037]–[0038].

As noted above for claim limitation [1.c(ii)], the combined system "encod[es] the plurality of decoded images of the decoded portion of video into an alternate video stream including a second plurality of encoded images . . . ." EX1004, 1923; EX1005, 23:8-30; EX1003, ¶253. Like the combined system, the video encoder of Gu "receives a sequence of video pictures (frames) as its raw video content input 210 and produces a compressed bit stream 230 as output." EX1006, [0027]. A POSITA would have understood the output bit stream consists of a plurality of encoded images. EX1003, ¶253.

3. Claims 6 and 16

As discussed above, Sambe in view of Vetro teaches every limitation of claims 1 and 11. *See supra* §§ IV.A.2, 3. And, as noted above, a POSITA would have looked to combine Gu with the combination of Sambe and Vetro. *See supra* § IV.B.1. The combination of Sambe in view of Vetro and Gu renders the method of Claim 6 and the apparatus of claim 16 obvious. EX1003, ¶¶254-265, 271-274.

(1) [6.pre] and [16.pre] "The [method / system] of claim [1 / 11], [by further performing / wherein the plurality of transcoding devices are configured to further perform] the following at each of the plurality of transcoding devices in parallel:"

To the extent the preambles of claims 6 and 16 are limiting, the combination of Sambe in view of Vetro and Gu discloses them. EX1003, ¶¶255, 271-272. As explained above, Sambe describes a system including transcoding devices, or PCs, that perform transcoding processes in parallel. *See supra* §§ III.A, IV.A.2.

(2) [6.a] and [16.a] "performing quantization on the sets of transform coefficients for an image in the plurality of decoded images based at least in part on the scene complexity information within the information based on the media metadata; and"

The combination of Sambe in view of Vetro and Gu renders claim limitation [6.a] and [16.a] obvious. EX1003, ¶¶256-260, 273. As noted above, the combined system generates sets of transform coefficients for an image in the plurality of decoded images. *See supra* § IV.B.2(3).

As noted above, Sambe in view of Vetro discloses encoding the plurality of decoded images "based on . . . the information based on the media metadata," where the media metadata comprises scene complexity information. *See supra* § IV.A.2(8). As noted above, at the time of invention, performing quantization on the sets of transform coefficients for an image in the plurality of decoded images was a widely understood and used step in the encoding process. *See supra* § II.A.1; EX1003,

¶257; *see also, e.g.*, EX1019, 17. As a result, despite not explicitly saying so, both Sambe and Vetro implicitly perform this step as part of their respective encoding processes. EX1003, ¶258. For example, Sambe discloses converting video into MPEG-2 and MPEG-4 ASP standards. Both require performing quantization on sets of transform coefficients for an image as part of their respective encoding processes. *Id.*, ¶258. A POSITA would have, thus, understood that Sambe in view of Vetro discloses "performing quantization on the sets of transform coefficients for an image in the plurality of decoded images based at least in part on the scene complexity information within the information based on the media metadata." *Id.*, ¶259.

But even if performing quantization on the sets of transform coefficients for an image in the plurality of decoded images is not expressly disclosed in Sambe or Vetro, it is in Gu, and a POSITA would have been motivated to look at Gu to provide more explicit detail about this standard encoding functionality. *Id.*, ¶260. As noted above, the quantizer of Gu, a component of the video encoder, subsequently "quantizes the blocks of transform coefficients." *See supra* § IV.B.2(4).

(3) [6.b] and [16.b] "quantizing the generated set of transform coefficients according to the target format."

The combination of Sambe in view of Vetro and Gu renders claim limitation [6.b] and [16.b] obvious. EX1003, ¶¶261-265, 274. As noted above for limitation [6.a], Sambe in view of Vetro and Gu teaches quantizing the generated set of transform coefficients. *Supra* § IV.B.2(2).

Sambe discloses this limitation. Sambe's transcoding PC's re-encoding step of its disclosed transcoding process involves first "deciding the target frame size," and then determining the "quantization scale Q of each macro-block … *so that actual coded size will be equal to the target size using virtual buffer memory* d_n." EX1004, 1925 (emphasis added). Because the quantization scale Q of each macroblock is chosen based on the target size of the encoding, a POSITA would have understood that Sambe discloses quantizing the generated set of transform coefficients according to the target format. EX1003, ¶262.

Gu also renders this limitation obvious. EX1003, ¶263. Gu discloses that "[e]ncoding can be performed according to a known video encoding standard, such as Windows Media Video format, SMPTE 421-M format, MPEG-x format (e.g., MPEG-1, MPEG-2, or MPEG-4), H.26x format (e.g., H.261, H.262, H.263, or H.264), or other format." EX1006, [0026].

A POSITA would have understood the video encoder of Gu quantizes the transform coefficients according to the target format. EX1003, ¶264; *see also, e.g.*, EX1016, 884 (describes coding standards specifying quantization tables). Gu explains that "[t]he initial sample quantization step size" for the encoding process "can be chosen depending on the particular codec standard used by the video encoder." EX1006, [0053]. "For example, in the case of the SMPTE 421-M video codec, the initial sample quantization step size may be chosen to be 4." *Id.*

As noted above, at the time of invention, "quantizing the generated set of transform coefficients" was a widely understood and used step in the encoding process. *See supra* § II.A.1; EX1003, ¶265; *see also, e.g*, EX1019, 17. Indeed, despite not explicitly saying so, Vetro implicitly performs this step as part of its encoding process. EX1005, 2:38-60 (describing quantization of transform coefficients in the Background Section); EX1003, ¶265. Therefore, a POSITA would have been motivated to look at Gu to provide more explicit detail about the claimed functionalities. EX1003, ¶265.

V. DISCRETIONARY ANALYSIS

A. *Fintiv* Analysis

The Board should not use its discretion to deny institution under § 314(a). None of the factors set forth in *Apple Inc. v. Fintiv, Inc.*, IPR2020-0019, Paper 11 (PTAB Mar. 20, 2020) favor discretionary denial. The district court has not entered a scheduling order in the Current Litigation, nor has a trial date been set (Factor 2). Neither the parties nor the district court have invested substantial resources in the Current Litigation, as the case has not advanced beyond the pleadings stage (Factor 3). The parties have not exchanged infringement or invalidity contentions. Thus, there is currently no overlap between the issues raised in this petition and in the Current Litigation (Factor 4). Petitioner is the sole defendant in the Current

Litigation (Factor 5). The merits of this Petition are compelling, and no forum has ever considered the grounds raised herein (Factor 6).

B. *Advanced Bionics*

Under *Advanced Bionics* and applying the *Becton Dickinson* factors, the Board should not deny institution under § 325(d). Here, *Advanced Bionics*, Part 1, is not satisfied because the same or substantially the same art or arguments have not been "previously [] presented to the Office" under 35 U.S.C. § 325(d). During the prosecution of the 806 patent, Sambe and Vetro were not before the Examiner, and while Gu was disclosed to the PTO, the Examiner did not analyze Gu with respect to the proposed claim or otherwise discuss it. *Supra* § III.C. Moreover, the Examiner did not analyze either Ground presented in this Petition. *Id.* Additionally, the 806 IPR has not previously been subject to an AIA proceeding. For these reasons, the Board should not deny institution under § 325(d). *See Advanced Bionics, LLC v. MED-EL Elektromedizinische Geräte GmbH*, IPR2019-01469, Paper 6 (Feb. 13, 2020) (precedential).

VI. CONCLUSION

For these reasons, Petitioner requests *inter partes* review and cancellation of the challenged claims.

VII. MANDATORY NOTICES AND FEES

A. Real Party-in-Interest

Petitioner is the real party-in-interest. No other party directed, controlled, or funded this IPR proceeding.

B. Related Matters

Petitioner is aware of the following additional related matters involving the 806 Patent and/or related patents:

Case Caption	Forum	Patents
DivX, LLC v. Amazon.com, Inc. :24-cv-02061-CMH-LRV	U.S. District Court, Eastern District of Virginia	U.S. Patent Nos. 10,412,141; 10,715,806; 9,955,195; 11,611,785; 10,542,303; 11,245,938; and 12,184,943

C. Lead and Backup Counsel

Pursuant to 37 C.F.R. § 42.8(b)(3) and 42.10(a), Petitioner designates J. David Hadden, Reg. No. 40,629, as lead counsel, and Saina Shamilov, Reg. No. 48,266, and Allen Wang, Reg. No. 68,456, as back-up counsel, each of Fenwick & West LLP.

D. Service Information

Petitioner consents to service by electronic mail at:

DivX-IPR@fenwick.com.

Petitioner's counsel may also be served by mail or hand delivery at Fenwick & West LLP, 801 California Street, Mountain View, California 94041. Petitioner's counsel may be reached by telephone at (650) 988-8500.

E. Fees

The Office is authorized to charge fees for this Petition to Deposit Account 19-2555.

Dated: May 27, 2025

FENWICK & WEST LLP

/J. David Hadden/
J. David Hadden
Reg. No. 40,629

Attorneys for Petitioners
Amazon.com, Inc. and
Amazon Web Services, Inc.

CERTIFICATION OF WORD COUNT

The undersigned certifies pursuant to 37 C.F.R. § 42.24 that the foregoing Petition for *Inter Partes* Review, excluding any table of contents, mandatory notices under 37 C.F.R. § 42.8, certificates of service or word count, or appendix of exhibits, contains 13,920 words according to the word-processing program used to prepare this document (Microsoft Word).

Dated: May 27, 2025

FENWICK & WEST LLP

/J. David Hadden/
J. David Hadden
Reg. No. 40,629

Attorneys for Petitioner
Amazon.com, Inc. and
Amazon Web Services, Inc.

**CERTIFICATE OF SERVICE ON PATENT OWNER
UNDER 37 C.F.R. § 42.105**

I hereby certify, pursuant to 37 C.F.R. Sections 42.6 and 42.105, that a complete copy of the attached **PETITION FOR INTER PARTES REVIEW OF U.S. PATENT NO. 10,715,806,** including all exhibits (**Nos. 1001-1030**), is being served via Federal Express on May 27, 2025, upon Patent Owner by serving the correspondence address of record with the USPTO as follows:

71897 - KPPB LLP
3780 Kilroy Airport Way, Suite 320
Long Beach, CA 90806

The foregoing was also served via Federal Express upon counsel of record for Patent Owner in the litigation pending before the U.S. District Court for the Eastern District of Virginia entitled *DivX, LLC v. Amazon.com, Inc., and Amazon Web Services, Inc.*, No. 1:24-cv-020610CMH-LRV as follows:

Charles B. Molster, III
The Law Offices Of Charles B. Molster, III PLLC
2141 Wisconsin Avenue, N.W., Suite M
Washington, D.C. 20007

Dated: May 27, 2025

FENWICK & WEST LLP

/J. David Hadden/
J. David Hadden
Reg. No. 40,629

Attorneys for Petitioner
Amazon.com, Inc. and
Amazon Web Services, Inc.

APPENDIX: LIST OF CHALLENGED CLAIMS:

806 Patent, Claim 1	
1(pre)	A method for transcoding a source video file into a set of multiple alternate video streams, the method comprising:
1(a)	generating, at a computer system configured as a media metadata generation device, media metadata related to the source video file prior to decoding, during a transcoding of, at least a portion of the source video file, where the media metadata comprises scene complexity information;
1(b)	providing information based on the media metadata from the computer system to a plurality of transcoding devices; and
1(c(i))	performing the following at each of the plurality of transcoding devices in parallel:
1(c(ii))	receiving the at least a portion of the source video file, including a first plurality of encoded images encoded according to a source format, from a media content source;
1(c(iii))	decoding the at least a portion of the source video file based on the source format to generate a decoded portion of video including a plurality of decoded images;
1(c(iv))	receiving the information based on the media metadata from the computer system; and
1(c(v))	encoding the plurality of decoded images of the decoded portion of video into an alternate video stream including a second plurality of encoded images based on a target format and the information based on the media metadata, the alternate video stream being one of the set of multiple alternate video streams.

806 Patent, Claim 2	
2	The method of claim 1, wherein a group of the plurality of transcoding devices perform processes in parallel for the same alternate video stream from the set of multiple alternate video streams.
806 Patent, Claim 3	
3	The method of claim 1, wherein a group of the plurality of transcoding devices perform processes in parallel for different alternate video streams in the set of multiple alternate video streams.
806 Patent, Claim 4	
4	The method of claim 1, wherein the media metadata further comprises scene change information indicating the start and end of a scene.
806 Patent, Claim 5	
5(pre)	The method of claim 1, wherein the encoding of the plurality of decoded images of the decoded portion of video into an alternate video stream includes performing at least the following operations for images in the plurality of decoded images:
5(a)	generating a prediction image for each of a plurality of coding units of an image in the plurality of decoded images using the scene change information and the scene complexity information within the information based on the media metadata according to the target format;
5(b)	performing transforms on residual images of the plurality of coding units to generate sets of transform coefficients based on the target format; and
5(c)	performing entropy encoding on the sets of transform coefficients to generate images for the second plurality of encoded images.

806 Patent, Claim 6	
6(pre)	The method of claim 1, by further performing the following at each of the plurality of transcoding devices in parallel:
6(a)	performing quantization on the sets of transform coefficients for an image in the plurality of decoded images based at least in part on the scene complexity information within the information based on the media metadata; and
6(b)	quantizing the generated set of transform coefficients according to the target format.
806 Patent, Claim 7	
7(pre)	The method of claim 4, by further performing the following at each of the plurality of transcoding devices in parallel:
7(a)	determining a number of bits to encode a group of pictures (GOP) based at least in part on a number of frames between the start and end of a scene as indicated by the information based on the media metadata.
806 Patent, Claim 8	
8	The method of claim 1, wherein the source format and the target format have different resolutions.
806 Patent, Claim 9	
9	The method of claim 1, wherein the source format and the target format correspond to different video encoding standards.
806 Patent, Claim 10	
10(pre)	The method of claim 1, by further performing the following at each of the plurality of transcoding devices in parallel:
10(a)	dividing an image in the plurality of decoded images into a plurality of coding units based on the target format.

806 Patent, Claim 11	
11(pre)	A system for transcoding video data, the system comprising:
11(a)	a computer system configured as a media metadata generation device, wherein the computer system is configured to:
11(a)(i)	generate media metadata related to the source video file prior to decoding, during a transcoding of, at least a portion of the source video file, where the media metadata comprises scene complexity information; and
11(a)(ii)	provide information based on the media metadata to a plurality of transcoding devices; and
11(b)	the plurality of transcoding devices, configured to perform the following at each of the plurality of transcoding devices in parallel:
11(b)(i)	receive the at least a portion of the source video file, including a first plurality of encoded images encoded according to a source format, from a media content source;
11(b)(ii)	decode the at least a portion of the source video file based on the source format to generate a decoded portion of video including a plurality of decoded images;
11(b)(iii)	receive the information based on the media metadata from the computer system; and
11(b)(iv)	encode the plurality of decoded images of the decoded portion of video into an alternate video stream including a second plurality of encoded images based on a target format and the information based on the media metadata, the alternate video stream being one of the set of multiple alternate video streams.
806 Patent, Claim 12	
12	The system of claim 11, wherein a group of the plurality of transcoding devices perform processes in parallel for the same alternate video stream from the set of multiple alternate video streams.

806 Patent, Claim 13	
13	The system of claim 11, wherein a group of the plurality of transcoding devices perform processes in parallel for different alternate video streams in the set of multiple alternate video streams.
806 Patent, Claim 14	
14	The system of claim 11, wherein the media metadata further comprises scene change information indicating the start and end of a scene.
806 Patent, Claim 15	
15(pre)	The system of claim 11, wherein encoding the plurality of decoded images of the decoded portion of video into an alternate video stream is performed using at least the following operations for images in the plurality of decoded images:
15(a)	generating a prediction image for each of a plurality of coding units of an image in the plurality of decoded images using the scene change information and the scene complexity information within the information based on the media metadata according to the target format;
15(b)	performing transforms on residual images of the plurality of coding units to generate sets of transform coefficients based on the target format; and
15(c)	performing entropy encoding on the sets of transform coefficients to generate images for the second plurality of encoded images.
806 Patent, Claim 16	
16(pre)	The system of claim 11, wherein the plurality of transcoding devices are configured to further perform the following at each of the plurality of transcoding devices in parallel:
16(a)	performing quantization on the sets of transform coefficients for an image in the plurality of decoded images based at least in

	part on the scene complexity information within the information based on the media metadata; and
16(b)	quantizing the generated set of transform coefficients according to the target format.
806 Patent, Claim 17	
17(pre)	The system of claim 14, wherein the plurality of transcoding devices are configured to further perform the following at each of the plurality of transcoding devices in parallel:
17(a)	determining a number of bits to encode a group of pictures (GOP) based at least in part on a number of frames between the start and end of a scene as indicated by the information based on the media metadata.
806 Patent, Claim 18	
18(pre)	The system of claim 11, wherein the plurality of transcoding devices are configured to further perform the following at each of the plurality of transcoding devices in parallel:
18(a)	dividing an image in the plurality of decoded images into a plurality of coding units based on the target format.
806 Patent, Claim 19	
19	The system of claim 11, wherein the source format and the target format have different resolutions.
806 Patent, Claim 20	
20	The system of claim 11, wherein the source format and the target format correspond to different video encoding standards.
806 Patent, Claim 21	
21(pre)	A method for transcoding a source video file into a set of multiple alternate video streams, the method comprising:
21(a)	generating, at a computer system configured as a media metadata generation device, media metadata related to the source video file prior to decoding, during a transcoding of, at

	least a portion of the source video file, where the media metadata comprises scene change information indicating the start and end of a scene, and scene complexity information;
21(b)	providing information based on the media metadata from the computer system to a plurality of transcoding devices; and
21(c)	performing the following at each of the plurality of transcoding devices in parallel:
21(c)(i)	receiving the at least a portion of the source video file, including a first plurality of encoded images encoded according to a source format, from a media content source;
21(c)(ii)	decoding the at least a portion of the source video file based on the source format to generate a decoded portion of video including a plurality of decoded images;
21(c)(iii)	receiving the information based on the media metadata from the computer system;
21(c)(iv)	dividing an image in the plurality of decoded images into a plurality of coding units based on a target format, wherein the source format and the target format have different resolutions;
21(c)(v)	determining a number of bits to encode a group of pictures (GOP) based at least in part on a number of frames between the start and end of a scene as indicated by the information based on the media metadata; and
21(c)(vi)	encoding the plurality of decoded images of the decoded portion of video into an alternate video stream including a second plurality of encoded images based on the target format and the information based on the media metadata, the alternate video stream being one of the set of multiple alternate video streams.