

Declaration of Henry Houh, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 11,126,853 B2

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,
Petitioner,

v.

CELLULAR SOUTH, INC.,
Patent Owner.

Case IPR2025-00877
Patent 11,126,853 B2
Issue Date: September 21, 2021

Title: VIDEO TO DATA

DECLARATION OF HENRY HOUH, PH.D.

Declaration of Henry Houh, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 11,126,853 B2

TABLE OF CONTENTS

	Page
I. INTRODUCTION AND QUALIFICATIONS	5
A. Qualifications and Experience	5
B. Materials Considered	13
II. PERSON OF ORDINARY SKILL IN THE ART	15
III. STATEMENT OF LEGAL PRINCIPLES	17
A. Claim Construction	17
B. Obviousness (§ 103)	19
IV. THE '853 PATENT	25
A. Overview of the Specification	25
B. The Challenged Claims	29
V. APPLICATION OF THE PRIOR ART TO CHALLENGED CLAIMS	30
A. Brief Summary and Overview of Prior Art	31
1. References Common to All Grounds	31
(a) Zhao (EX1003)	31
(b) Kritt (EX1004)	34
(c) Steinberg (EX1005)	37
(d) Kouzani (EX1006)	39
2. Additional References	42
(a) Yang (EX1007)	42
(b) Romdhani (EX1008)	44
(c) Trivedi (EX1009)	46
(d) Singer (EX1010)	49
B. Ground 1: Claims 1-4 Are Obvious Over Zhao in View of Kritt, Steinberg, and Kouzani	54

TABLE OF CONTENTS
(continued)

	Page
1. Independent Claim 1: “A system for generating data from a video, comprising:” (Claim 1[pre]).....	54
(a) “a coordinator communicatively coupled to a splitter and to a plurality of demultiplexer nodes, wherein the splitter is configured to segment the video, wherein the demultiplexer nodes are configured to extract audio files from the video and to extract still frame images from the video;” (Claim 1[a])	56
(b) “an image detector configured to detect an image of an object in the still frame images, wherein the image detector is adjustable to increase detection of non-primary images in the video; and” (Claim 1[b]).....	74
(c) “an object recognizer configured to compare the image of the object to a fractal, wherein the fractal includes a representation of the object based on landmarks associated with the object, wherein the recognizer is further configured to update the fractal with the image.” (Claim 1[c]).....	82
2. Claim 2: “The system of claim 1, wherein the coordinator is configured to increase the plurality of demultiplexer nodes when a threshold of processing capacity is reached.”	93
3. Claim 3: “The system of claim 1, wherein the recognizer is configured to determine distinguishing geometric features of the object.”	95
4. Claim 4: “The system of claim 3, wherein the distinguishing geometric features comprise a contour of eye sockets, a nose, and a chin.”	96
C. Ground 2: Claims 5-7 Are Obvious Over Ground 1 Prior Art in Further View of Yang and Romdhani	97

TABLE OF CONTENTS
(continued)

	Page
5. Claim 5: “The system of claim 4, wherein the recognizer is configured to determine skin textures of the object.”.....	97
6. Claim 6: “The system of claim 5, wherein the distinguishing geometric features or the skin textures are determined based on a three-dimensional model.”	103
7. Claim 7: “The system of claim 6, wherein the three-dimensional model is a three-dimensional morphable model.”	105
D. Ground 3: Claims 8-9 Are Obvious Over Ground 1 Prior Art in Further View of Trivedi	105
8. Claim 8: “The system of claim 1, further comprising a processor and a camera for capturing the video.”	105
9. Claim 9: “The system of claim 8, wherein the processor is configured to transform the video into a rectilinear format.”	109
E. Ground 4: Claims 10-11 Are Obvious Over Ground 1 Prior Art in Further View of Singer	112
10. Claim 10: “The system of claim 1, wherein the coordinator is configured to embed metadata about the object into the video, and wherein the metadata comprises a timestamp and a coordinate location of the object in the still frame images.”.....	112
11. Claim 11: “The system of claim 1, wherein the coordinator is configured to generate a metadata stream corresponding to the image, wherein the metadata stream includes one or more timestamps corresponding to the image, and wherein the coordinator is configured to embed the metadata stream in the video.”	121
V. NO SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS	123
VI. CONCLUSION	124

Declaration of Henry Houh, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 11,126,853 B2

I, Henry Houh, Ph.D., declare as follows:

I. INTRODUCTION AND QUALIFICATIONS

A. Qualifications and Experience

1. My complete qualifications and professional experience are described in my *Curriculum Vitae*, a copy of which is attached as Appendix A. The following is a brief summary of my relevant qualifications and professional experience.

2. I received a Ph.D. in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology (“MIT”) in 1998. Beforehand, I received a Master of Science degree in Electrical Engineering and Computer Science in 1991, a Bachelor of Science degree in Electrical Engineering and Computer Science in 1989, and a Bachelor of Science degree in Physics in 1990, all from MIT.

3. I am currently self-employed as an independent technical consultant. From its founding in 2012 until October 2022, I was also president of a company that provides supplemental science, technology, engineering, and mathematics (“STEM”) education to children of all ages.

4. I first entered telecommunications in 1987 when I worked as a summer intern at AT&T Bell Laboratories as part of a five-year dual degree program at MIT. I continued to work at AT&T Bell Laboratories as part of this MIT program. While at MIT, I was a teaching assistant (“TA”) in the Electrical Engineering and Computer

Declaration of Henry Houh, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 11,126,853 B2

Science Department's core Computer Architectures course. I first was a TA as a senior for a role typically reserved for graduate students. I later became head TA. The course covered various topics in computer architectures. As a TA, I helped write homework assignments, lab assignments, and exams. I also taught in some of the recitation sections.

5. Later, as part of my doctoral research at MIT from 1991-1998, I was a research assistant in the Telemedia Network Systems ("TNS") group at the Laboratory for Computer Science. The TNS group built a high-speed gigabit network and created applications that ran over the network. Example applications included ones for remote video capture, processing, and display of video on computer terminals. In addition to working on the design of core network components, designing and building the high-speed links, and designing and writing the device drivers for the interface, I also set up the group's web server. Also, I helped to maintain, install, and upgrade the networking devices used within the group, along with other graduate students.

6. I also helped to build the web pages that initiated the above-mentioned video sessions via a web interface. Vice President Al Gore visited our group in 1996 and received a demonstration of—and remotely drove—a radio-controlled toy car with a wireless video camera mounted on it that was built by our group. This toy car

device received commands transmitted over a network from a remote computer, and video data from the toy car was transmitted wirelessly then over a computer network back to the user controller. On occasion, we allowed users visiting our web site to drive the toy car from their remote computer while they watched the video on their computer. The video stream was encoded by TNS-designed hardware, streamed over the TNS-designed network, and displayed using TNS-designed software.

7. Our video environment included an extensive amount of video processing. The closed captions were extracted from the analog video broadcasts, and our group built an extensive toolkit to perform video processing functions on the live video as it was being captured and optionally recorded. Our system included features such as motion detection, template matching, model matching, live green screening, face detection, histogram generation and matching, feature matching, edge detection, scene change detection, image transformation, filtering, image segmentation, object motion tracking, and other features.

8. I defended and submitted my Ph.D. thesis, titled “Designing Networks for Tomorrow’s Traffic,” in January 1998. As part of my thesis research, I analyzed local area and wide area flows to show a more efficient method for routing packets in a network, based on traffic patterns at the time. The traffic flow data included Ethernet, IP, TCP or UDP, and RTP header information, which I analyzed to come

Declaration of Henry Houh, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 11,126,853 B2

to the conclusions in my thesis.

9. While I was in graduate school, I started a company that provided web site design services and targeted advertising opportunities. We also sold targeted advertising placements. As the web became popular first at colleges and universities, we focused initially on building college recruiting web sites and sold advertisements to companies wishing to target students to recruit at specific colleges. We developed technology that utilized the web user's college affiliation to customize advertisements targeted to students at that college.

10. From 1997 to 1999, I was a Senior Scientist and Engineer at NBX Corporation, a start-up that made business telephone systems for streaming packetized audio over data networks instead of using traditional telephone lines. NBX was later acquired by 3Com Corporation and the phone system is still used today by numerous businesses.

11. As part of my work at NBX, I designed the core audio reconstruction algorithms for the telephones, as well as the packet transmission algorithms. I also designed and validated the core packet transport protocol used by the phone system. The protocol was used for all signaling in the phone system, including for the setup of conference calls. The NBX system also featured a computer interface for initiating phone calls, which could also initiate conference calls. The NBX system also

Declaration of Henry Houh, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 11,126,853 B2

supported the Telephony Application Programming Interface (“TAPI”) that allowed other computer programs to integrate with our system telephony features. We obtained U.S. Patent No. 6,967,963, entitled “Telecommunication method for ensuring on-time delivery of packets containing time-sensitive data,” as part of this work. The first release of the phones had integrated Ethernet repeater hubs embedded in them, and I worked with the documentation and field support people to develop sets of guidelines so that customers would not build their networks to violate the Ethernet specifications, typically by chaining too many hub repeaters and NBX phones together. The first release of the phones also worked using Ethernet-layer packets (without the use of IP). I also programmed the first prototype of our phone which communicated using IP, and I demonstrated our IP phones working over the Internet when we attended a trade show in California. The phone connected over the Internet to our headquarters in Andover, MA.

12. From 1999-2004, I was employed by Empirix or its predecessor company, Teradyne. Teradyne had traditionally made products to test chips and hardware, but the founder and board chairman, Alex d’Arbeloff wanted to expand the company’s products to include software systems testing. Teradyne made a number of acquisitions to achieve this goal, and I was hired to work for Mr. d’Arbeloff to work with all the acquired software test divisions. These divisions

were later spun-out of Teradyne into a separate company, Empirix. Empirix was a leader in test tools for software systems including telecommunications protocols and systems, providing functional testing tools as well as load testing tools. From 2000-2001, I conceived and built a test platform for testing Voice-over-IP (VoIP). The first application on this new test platform was a cloud emulator for simulating the effects of transmitting VoIP over a busy network. The test platform was based on a network processor chip, which could be programmed to cut-through packets while processing packet data such as various protocol layer headers including addresses included therein and even packet data contents. I also designed a protocol analyzer built on the same platform. The application captured and performed protocol decoding at various layers in the protocol stacks of captured packets, including detailed Ethernet header decoding, IP header decoding, TCP header decoding, UDP header decoding, RTP header decoding, and many other specified protocols. The application was also designed to reconstruct entire conversations that spanned multiple packets.

13. In 2006, as part of my role at BBN Technologies, I helped found PodZinger Inc., now known as RAMP Inc. PodZinger utilized BBN's speech recognition algorithms to search through the spoken words in audio and video segments. Our speech recognition algorithms were also capable of extracting the

Declaration of Henry Houh, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 11,126,853 B2

emotional prosody of the speech input. While I was Vice President of Operations and Technology, PodZinger followed its initial prototype with a full streaming audio and video search solution. The metadata generated from speech recognition process was used in the search result user interface so that the user could see highlights of their search term in the audio transcript, the context around their search term, and the user could select a word in the transcript to jump the video to play from exactly the point where the selected word was uttered.

14. I also created a social networking web site, which BBN sold to a venture-funded startup company. In the process of creating the web site, I designed and specified the authentication and authorization protocols.

15. I applied for and received a number of patents in the above areas. In particular, U.S. Patent No. 9,697,230 describes “an automated method and apparatus for generating metadata enhanced for audio/video search-driven applications. The apparatus includes a media indexer that obtains an (sic) media file/stream (e.g., audio/video podcasts), applies one or more automated media processing techniques to the media file/stream, combines the results of the media processing into metadata enhanced for audio/video search, and stores the enhanced metadata in a searchable index or other data repository.” (EX1016, 4:31-39.) A media indexer “applies a predefined set of audio or video processing routines to derive a portion of the

enhanced metadata from the media content.” (*Id.*, 5:19-21.) The indexer utilizes speech recognition, natural language processing, and video frame analyzers, that can recognize scenes, watermarks, objects, faces, and overlay text. (*Id.*, Fig. 1B, 5:22-34, 6:43-57.) Related advertising may then be displayed utilizing the results of such an analysis. (*Id.*, 15:52-16:25.) A partial list of my patents and patent applications includes:

- U.S. Patent No. 9,697,231, H. Houh, J. Stern, R. Spina, M. Meteer [sic], “Methods and apparatus for providing virtual media channels based on media search,” July 4, 2017.
- U.S. Patent No. 9,697,230, H. Houh, J. Stern, “Methods and apparatus for dynamic presentation of advertising, factual, and informational content using enhanced metadata in search-driven media applications,” July 4, 2017. *See also* WO 2007/056485.
- U.S. Patent No. 7,801,910, H. Houh and J. N. Stern, “Method and apparatus for timed tagging of media content,” September 21, 2010. *See also* US 2007/0112837, US 2009/0222442, WO 2007/056535.
- U.S. Patent Application Publication No. 2007/0106685, “Method and apparatus for updating speech recognition databases and reindexing audio and

video content using the same”

- U.S. Patent Application Publication No. 2007/0106660, “Method and apparatus for using confidence scores of enhanced metadata in search-driven media applications”
- U.S. Patent Application Publication No. 2007/0118873, “Methods and apparatus for merging media content”
- U.S. Patent Application Publication No. 2009/0222442, “User-directed navigation of multimedia search results”
- U.S. Patent Application No. 11/395,732, “User-Directed Navigation of Multimedia Search Results”

16. I have been retained by counsel for Petitioner to provide my expert opinion in connection with the above-captioned proceeding as set forth herein.

B. Materials Considered

17. The analysis that I provide in this Declaration is based on my education, research, and experience, as well as the documents I have considered. In forming my opinions, I have read and considered U.S. Patent No. 11,126,853 B2 (“853 patent”) (**EX1001**) and its prosecution history. I have cited to the following documents in my analysis below:

Declaration of Henry Houh, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 11,126,853 B2

Exhibit No.	Description of Document
1001	U.S. Patent No. 11,126,853 B2 to Bartlett Wade Smith, IV (filed February 8, 2019, issued September 21, 2021) (“ 853 ” or “ 853 patent ”)
1003	U.S. Patent App. Pub. No. 2009/0141940 A1 to Liang Zhao et al. (filed December 3, 2008, published June 4, 2009) (“ Zhao ”)
1004	U.S. Patent App. Pub. No. 2014/0181668 A1 to Barry A. Kritt et al. (filed Dec. 20, 2012, published Jun. 26, 2014) (“ Kritt ”)
1005	U.S. Patent No. 7,574,016 B2 to Eran Steinberg et al. (filed June 26, 2003, issued August 11, 2009) (“ Steinberg ”)
1006	Kouzani, A.Z. et al., “Fractal Face Representation and Recognition,” in Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, pp.1609-1613 (1997) (“ Kouzani ”)
1007	Yang, Ming-Hsuan et al., “Detecting Faces in Images: A Survey,” in IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 24, No. 1 (2002) (“ Yang ”)
1008	Romdhani, S. et al., “Face Identification by Fitting a 3D Morphable Model using Linear Shape and Texture Error Functions,” European Conf. on Computer Vision (2002) (“ Romdhani ”)
1009	U.S. Patent App. Pub. No. 2006/0187305 A1 to Mohan M. Trivedi et al. (PCT filed July 1, 2003, published August 24, 2006) (“ Trivedi ”)
1010	U.S. Patent App. Pub. No. 2011/0305394 A1 to David William Singer et al. (filed June 15, 2010, published December 15, 2011) (“ Singer ”)
1011	Peter Lubbers et al., <i>Pro HTML5 Programming: Powerful APIs for Richer Internet Application Development</i> (2010)
1012	Excerpts from Microsoft Computer Dictionary (5th ed. 2002)
1013	Jura van Vliet et al., <i>Programming Amazon EC2</i> (2011)
1014	Xiutao Tang et al., “Facial Image Recognition Based on Fractal Image Encoding,” in Bell Labs Technical Journal 15(1) (2010)
1015	Prosecution History of U.S. Patent No. 11,126,853 B2

Declaration of Henry Houh, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 11,126,853 B2

Exhibit No.	Description of Document
1016	U.S. Patent No. 9,697,230 to Henry Houh et al. (filed Mar. 31, 2006, issued Jul. 4, 2017)
1017	Cellular South's Opposition to Motion to Dismiss (Document 28) in No. 6:24-cv-00245-DAE (W.D. Tex.)

II. PERSON OF ORDINARY SKILL IN THE ART

18. The face page of the '853 patent shows that the earliest patent application to which the '853 patent claims priority was filed June 29, 2016. Therefore, I understand that the claims of the '853 patent are subject to the patent laws in effect under the America Invents Act ("AIA"), including sections 102 and 103 regarding invalidity based on prior art, which apply to patent applications having an effective filing date on or after March 16, 2013. I understand that under the AIA, an assessment of claims of a patent should be undertaken from the perspective of a person of ordinary skill in the art is to be made as of the effective filing date of the patent (i.e., the earliest of the actual filing date of the patent or application, or the filing date of the earliest application with priority). I note that, for patents subject to pre-AIA patent laws, such an assessment should be made as of the earliest claimed priority date (i.e., the "time the invention was made"). For purposes of my analysis, whether the "effective filing date" or "earliest claimed priority date" applies does not affect my analysis. This is because I understand that the alleged "effective filing

date” and “earliest claimed priority date” are the same in this case—June 29, 2016.

19. I have also been advised that to determine the appropriate level of a person having ordinary skill in the art, the following factors may be considered: (1) the types of problems encountered by those working in the field and prior art solutions thereto; (2) the sophistication of the technology in question, and the rapidity with which innovations occur in the field; (3) the educational level of active workers in the field; and (4) the educational level of the inventor.

20. The '853 patent states that it “relates to a method and a system for generating various and useful data from source media, such as videos and other digital content.” ('853, 1:12-14; *see also id.*, 1:34-36 (“The present invention is generally directed to a method to generate data from video content, such as text and/or image-related information.”).)

21. In my opinion, a person of ordinary skill in the art as of June 2016 would have possessed a bachelor’s degree in electrical engineering, computer science, or similar field, with two years of experience in developing and implementing computer software for processing and/or analyzing multimedia content, such as audio, video, or image data. A person could also have qualified as a person of ordinary skill in the art with some combination of (1) more formal education (such as a master’s of science degree) and less technical experience, or (2)

less formal education and more technical or professional experience.

22. My opinions regarding the level of ordinary skill in the art are based on, among other things, my experience in the field of computer science, my understanding of the basic qualifications that would be relevant to an engineer or scientist tasked with investigating methods and systems in the relevant area, and my familiarity with the backgrounds of colleagues, co-workers, and employees, both past and present.

23. Although my qualifications and experience exceed those of the hypothetical person having ordinary skill in the art defined above, my analysis and opinions regarding the '853 patent have been based on the perspective of a person of ordinary skill in the art as of June 2016.

III. STATEMENT OF LEGAL PRINCIPLES

A. Claim Construction

24. I understand that a purpose of claim construction is to determine what a person of ordinary skill in the art would have understood the claim terms to mean. Claim terms are generally given their ordinary and customary meaning, which is the meaning that the term would have to a person of ordinary skill in the art in question as of the effective filing date.

25. I understand that the person of ordinary skill in the art is deemed to read

Declaration of Henry Houh, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 11,126,853 B2

the claim term not only in the context of the particular claim in which the disputed term appears, but in the context of the entire patent, including the specification. I understand that the patent specification, under the legal principles, has been described as the single best guide to the meaning of a claim term, and is thus highly relevant to the interpretation of claim terms. And I understand for terms that do not have a customary meaning within the art, the specification usually supplies the best context of understanding the meaning of those terms.

26. I further understand that other claims of the patent in question, both asserted and unasserted, can be valuable sources of information as to the meaning of a claim term. Because the claim terms are normally used consistently throughout the patent, the usage of a term in one claim can often illuminate the meaning of the same term in other claims. Differences among claims can also be a useful guide in understanding the meaning of particular claim terms.

27. I understand that the prosecution history can further inform the meaning of the claim language by demonstrating how the inventors understood the invention and whether the inventors limited the invention in the course of prosecution, making the claim scope narrower than it otherwise would be. Extrinsic evidence, such as dictionaries, may also be consulted in construing the claim terms.

28. I understand that, in *Inter Partes* Review (IPR) proceedings, a claim of

a patent shall be construed using the same claim construction standard that would be used to construe the claim in a civil action filed in a U.S. district court (which I understand is called the “*Phillips*” claim construction standard), including construing the claim in accordance with the ordinary and customary meaning of such claim as understood by one of ordinary skill in the art and the prosecution history pertaining to the patent.

29. I have been instructed by counsel to apply the “*Phillips*” claim construction standard for purposes of interpreting the claims in this proceeding, to the extent they require an explicit construction. The description of the legal principles set forth above thus provides my understanding of the “*Phillips*” standard as provided to me by counsel.

30. For purposes of my analysis here, I do not believe express claim constructions are necessary because the prior art renders the claims obvious under any reasonable construction.

B. Obviousness (§ 103)

31. I understand that a patent claim is obvious if, as of the effective filing date, it would have been obvious to a person having ordinary skill in the field of the technology (the “art”) to which the claimed subject matter belongs.

32. I understand that the following factors should be considered in

Declaration of Henry Houh, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 11,126,853 B2

analyzing obviousness: (1) the scope and content of the prior art; (2) the differences between the prior art and the claims; and (3) the level of ordinary skill in the pertinent art. I also understand that certain other facts known as “secondary considerations” such as commercial success, unexplained results, long felt but unsolved need, industry acclaim, simultaneous invention, copying by others, skepticism by experts in the field, and failure of others may be utilized as indicia of nonobviousness. I understand, however, that secondary considerations should be connected, or have a “nexus,” with the invention claimed in the patent at issue.

33. I understand that a reference qualifies as prior art for obviousness purposes when it is analogous to the claimed invention. The test for determining what art is analogous is: (1) whether the art is from the same field of endeavor, regardless of the problem addressed, and (2) if the reference is not within the field of the inventor’s endeavor, whether the reference still is reasonably pertinent to the particular problem with which the inventor is involved.

34. I understand that a person of ordinary skill in the art is assumed to have knowledge of all prior art. I understand that one skilled in the art can combine various prior art references based on the teachings of those prior art references, the general knowledge present in the art, or common sense. I understand that a motivation to combine references may be implicit in the prior art, and there is no

requirement that there be an actual or explicit teaching to combine two references. Thus, one may take into account the inferences and creative steps that a person of ordinary skill in the art would employ to combine the known elements in the prior art in the manner claimed by the patent at issue. I understand that one should avoid “hindsight bias” and *ex post* reasoning in performing an obviousness analysis. But this does not mean that a person of ordinary skill in the art for purposes of the obviousness inquiry does not have recourse to common sense.

35. I understand that when determining whether a patent claim is obvious in light of the prior art, neither the particular motivation for the patent nor the stated purpose of the patentee is controlling. The primary inquiry has to do with the objective reach of the claims, and that if those claims extend to something that is obvious, then the entire patent claim is invalid.

36. I understand one way that a patent can be found obvious is if there existed at the time of the invention a known problem for which there was an obvious solution encompassed by the patent’s claims. I understand that a motivation to combine various prior art references to solve a particular problem may come from a variety of sources, including market demand or scientific literature. I understand that a need or problem known in the field at the time of the invention can also provide a reason to combine prior art references and render a patent claim invalid for

Declaration of Henry Houh, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 11,126,853 B2

obviousness. I understand that familiar items may have obvious uses beyond their primary purpose, and that a person of ordinary skill in the art will be able to fit the teachings of multiple prior art references together like the pieces of a puzzle. I understand that a person of ordinary skill is also a person of at least ordinary creativity. I understand when there is a design need or market pressure to solve a problem and there are a finite number of identified, predictable solutions, a person of ordinary skill has good reason to pursue the known options within his or her technical grasp. If this finite number of predictable solutions leads to the anticipated success, I understand that the invention is likely the product of ordinary skill and common sense, and not of any sort of innovation. I understand that the fact that a combination was obvious to try might also show that it was obvious, and hence invalid, under the patent laws. I understand that if a patent claims a combination of familiar elements according to known methods, the combination is likely to be obvious when it does no more than yield predictable results. Thus, if a person of ordinary skill in the art can implement a predictable variation, an invention is likely obvious. I understand that combining embodiments disclosed near each other in a prior art reference would not ordinarily require a leap of inventiveness.

37. I understand that obviousness may be shown by demonstrating that it would have been obvious to modify what is taught in a single piece of prior art to

create the patented invention. Obviousness may also be shown by demonstrating that it would have been obvious to combine the teachings of more than one item of prior art. I understand that a claimed invention may be obvious if some teaching, suggestion, or motivation exists that would have led a person of ordinary skill in the art to combine the invalidating references. I also understand that this suggestion or motivation may come from the knowledge of a person having ordinary skill in the art, or from sources such as explicit statements in the prior art. I understand that when there is a design need or market pressure, and there are a finite number of predictable solutions, a person of ordinary skill may be motivated to apply common sense and his skill to combine the known options in order to solve the problem.

38. I understand the following are examples of approaches and rationales that may be considered in determining whether a piece of prior art could have been combined with other prior art or with other information within the knowledge of a person having ordinary skill in the art:

- (1) Some teaching, motivation, or suggestion in the prior art that would have led a person of ordinary skill to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention;
- (2) Known work in one field of endeavor may prompt variations of it for use in the same field or a different field based on design incentives or other

market forces if the variations would have been predictable to a person of ordinary skill in the art;

(3) Combining prior art elements according to known methods to yield predictable results;

(4) Applying a known technique to a known device, method, or product ready for improvement to yield predictable results;

(5) Applying a technique or approach that would have been “obvious to try” (choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success);

(6) Simple substitution of one known element for another to obtain predictable results; or

(7) Use of a known technique to improve similar products, devices, or methods in the same way.

39. I understand that, when determining whether a claimed combination is obvious, the correct analysis is not whether one of ordinary skill in the art, writing on a blank slate, would have chosen the particular combination of elements described in the claim. Instead, I understand the correct analysis considers whether one of ordinary skill, facing the wide range of needs created by developments in the field of endeavor, would have seen a benefit to selecting the combination claimed.

40. I understand that the test for obviousness is not whether the features of a secondary reference may be bodily incorporated into the structure of the primary reference. The test for obviousness, in other words, is not whether the references could be physically combined but whether the claimed inventions are rendered obvious by the teachings of the prior art as a whole.

IV. THE '853 PATENT

A. Overview of the Specification

41. The '853 patent purports to offer one solution to the common problem of how to describe visual information from audio-visual content. It is in many ways similar to a museum curator providing text guides to an artist's work, or a movie critic explaining the cinematography of a film. The curator or critic may distinguish images in the foreground from those in the background, and explain visual techniques such as object placement, shading, and texturing that make the artwork or scene visually effective. The curator or critic can provide commentary that tells the viewer what to look for, where to look, and (in a film) when to look.

42. The '853 patent describes its purpose at a very high level. It states that it "relates to a method and a system for generating various and useful data from source media, such as videos and other digital content." ('853, 1:12-14.) And the summary of the invention confirms the broad brush the patent purports to take: "The

present invention is generally directed to a method to generate data from video content, such as text and/or image-related information.” (’853, 1:34-36.)

43. Audio-visual data such as videos long predates the ’853 patent, and prior artists similarly had long been working on various ways to analyze and describe video content. By the time of the alleged priority date (June 29, 2016), the fields of audio and image analysis for video content were well-developed. Indeed, the Background section of the ’853 patent acknowledges known image searching techniques to identify matching and similar images, as well as audio-to-text algorithms for transcribing text from audio. (’853, 1:21-29.)

44. The specification of the ’853 patent provides primarily a series of high-level, functional explanations for how to implement the alleged invention, for which it relies on well-known techniques to implement. (*See, e.g.*, ’853, 14:30-15:47 (describing various well-known “recognition models” for image detection).) For example, Figure 10 from the ’853 patent, shown below, “illustrates exemplary system architecture with an exemplary process flow.”

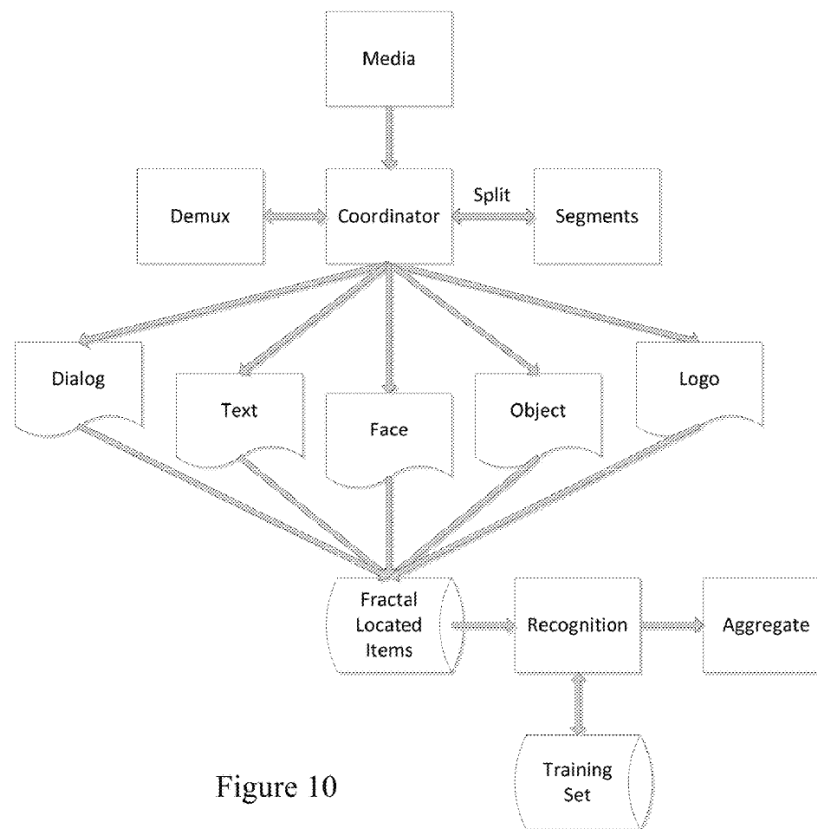


Figure 10

(’853, Fig. 10, 3:50-51.) Referring to Figure 10, the ’853 patent states that “[s]ource media can be provided to a coordinator.” (’853, 16:25-28.) The specification describes the coordinator functionally. The coordinator, among other things, “can be responsible for identifying new job requests and managing the routing of request input and result data output in the system.” (’853, 16:31-33.) “[T]he coordinator can send uploaded source media to a splitter and a demuxer/demultiplexer.” (’853, 16:60-63.) Like the coordinator, the splitter and demuxer/demultiplexer are described functionally. Then, data from the source media, such as faces, text, and

dialog, can be extracted. ('853, 16:63-67.) “[T]he extracted data can be sent through a filter to compare fractal located items...for recognition.” ('853, 17:4-6.) The '853 patent does not purport to have invented fractals; rather, the use of fractals for image recognition is one of the known recognition techniques identified in the '853 specification, including Principal Component Analysis (PCA) and Elastic Bunch Graph Matching (EBGM), among others. ('853, 14:30-15:47 (“[T]he following recognition models can be utilized, alone or in combination, and are provided to illustrate exemplary alternative embodiments.”).) Based on the results of the fractal filtering, the fractal can be refined and the fractal training set updated. ('853, 17:6-10.)

45. The '853 patent further states that, “[o]ften images are segmented only to locate and positively identify one or very few main images in the foreground of a given frame,” and “[t]he non-primary or background images are often treated as noise.” ('853, 9:8-12.) “Nevertheless,” according to the '853 patent, “these can provide useful information, context, and/or branding for example,” and “it may become necessary or desirable to detect more detail from a frame or set of frames.” ('853, 9:12-23.) “In such circumstances, the computational thresholds for identification of an object, face, etc. can be altered according to a then stated need or desire for non-primary, background, obstructed and/or grainy type images. Such

image identification threshold adjustment capability can be implemented, for example, as a user-controlled interface, dial, slider, or button, which enables the user to make adjustments to suit specific needs or preferences.” (’853, 9:23-30.)

46. I discuss additional aspects of the ’853 patent in my analysis of the claims below.

B. The Challenged Claims

47. This Declaration addresses claims 1-11 of the ’853 patent. Independent claim 1 is representative and recites:

1. A system for generating data from a video, comprising:
 - [a] a coordinator communicatively coupled to a splitter and to a plurality of demultiplexer nodes, wherein the splitter is configured to segment the video, wherein the demultiplexer nodes are configured to extract audio files from the video and to extract still frame images from the video;
 - [b] an image detector configured to detect an image of an object in the still frame images, wherein the image detector is adjustable to increase detection of non-primary images in the video; and
 - [c] an object recognizer configured to compare the image of the object to a fractal, wherein the fractal includes a representation of the object based on landmarks associated with the object, wherein the recognizer

is further configured to update the fractal with the image.

(’853, 25:22-26:3 (Claim 1 (bracketed notation (e.g., “[a]”) added)).)

48. I address the claims further in my detailed analysis in **Part V** below.

V. APPLICATION OF THE PRIOR ART TO CHALLENGED CLAIMS

49. I have reviewed and analyzed the prior art references and materials listed in **Part I.B** above. In my opinion the claims of the ’853 patent would have been obvious to a person of ordinary skill in the art based on the following combinations of the prior art.

Ground	Claims	Basis for Challenge Under § 103
1	1-4	Zhao in view of Kritt, Steinberg, and Kouzani
2	5-7	Ground 1 prior art in further view of Yang and Romdhani
3	8-9	Ground 1 prior art in further view of Trivedi
4	10-11	Ground 1 prior art in further view of Singer

50. I am informed by counsel that each of the references cited in the grounds above qualifies as prior art to the challenged claims because each reference was published before the earliest filing date to which the ’853 patent can claim priority (June 29, 2016).

A. Brief Summary and Overview of Prior Art

1. References Common to All Grounds

(a) Zhao (EX1003)

51. **Zhao**, U.S. Patent Application Publication No. 2009/0141940 A1, is entitled “Integrated Systems and Methods for Video-Based Object Modeling, Recognition, and Tracking.” I am informed that Zhao qualifies as prior art because it was both filed (on December 3, 2008) and published (on June 4, 2009) before the earliest priority date of the challenged claims.

52. Like the '853 patent, Zhao discloses a system that performs detection and recognition of objects, such as people and faces, in video images to generate data, including by creating and updating object models used for detection and recognition, and by generating data about the video that can be used for indexing and retrieval. Zhao explains, for example, that it “generally relate[s] to systems and methods for detection, modeling, recognition, and tracking of objects within video content” and to “indexing and retrieval systems for videos based on generated object models.” (Zhao, ¶0051.) “These objects include people, faces, articles of clothing, plants, animals, machinery, electronic equipment, food, and virtually any other type of image that can be captured or presented in video.” (Zhao, ¶0051.) Zhao’s system “may be operated in a computer environment” and “any results or outputs relating

to detection, modeling, recognition, indexing, and/or tracking of object images within videos may be stored in a database” or otherwise output. (Zhao, ¶0055.)

Zhao states that its system is “useful for a wide variety of applications, including video indexing and retrieval, video surveillance and security, unknown person identification, advertising, and many other fields.” (Zhao, ¶0053.)

53. Figure 2 of Zhao illustrates at a high level this overall process of generating data from a video performed by the system:

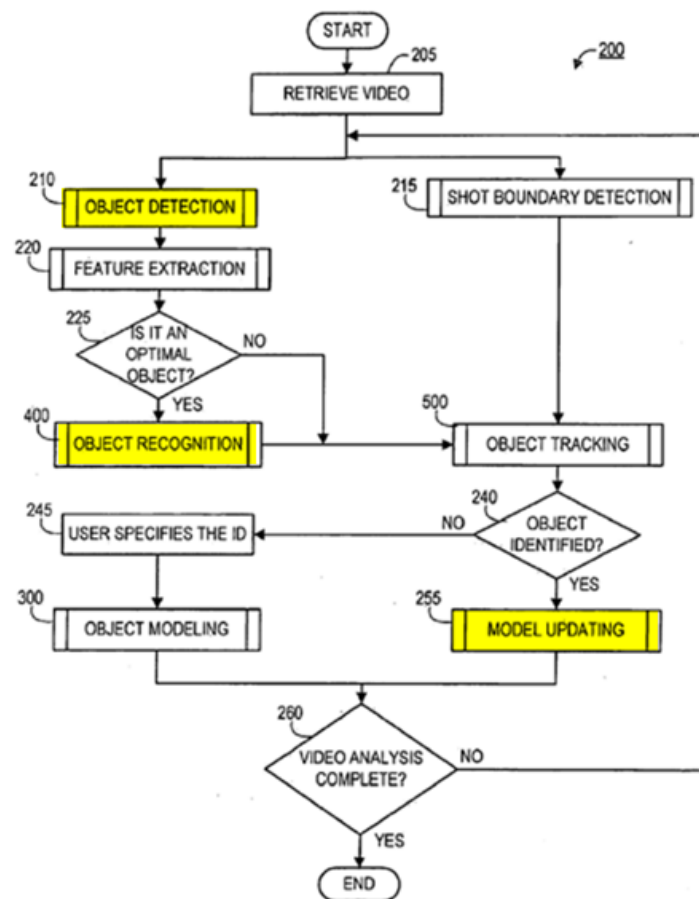


FIG. 2
OVERALL PROCESS

(Zhao, Fig. 2 (highlighting added); *see also id.*, ¶0036 (“FIG. 2 illustrates a flowchart describing the steps involved, from a high-level, in one embodiment of the present system for detecting, modeling, recognizing, and tracking object images throughout one or more videos.”).) As shown in the figure above, like claim 1 of the ’853 patent, Zhao’s process includes steps for object “detection” (210), object “recognition” (400), and model “updating” based on the detection and recognition (255).

54. I will provide more information about Zhao in my discussion of the claim limitations below.

(b) Kritt (EX1004)

55. **Kritt**, U.S. Patent Application Publication No. 2014/0181668 A1, is entitled “Visual Summarization of Video for Quick Understanding.” I am informed that Kritt qualifies as prior art because it was both filed (on December 20, 2012) and published (on June 26, 2014) before the earliest priority date of the challenged claims.

56. I point to Kritt primarily for its teachings regarding distributed/multi processing of video and audio in connection with my analysis of claim 1[a] and the recited “**coordinator**,” “**splitter**,” and “**demultiplexer nodes**.”

57. Like Zhao, Kritt describes a system for processing videos to generate data about the videos. For example:

According to various embodiments, a viewer selects a video, and human, nonhuman, and audio objects of the video are identified. In addition, key words that are spoken by human objects in the video are identified. Human, nonhuman, and audio objects may be used to classify a particular segment of a video as a scene. The objects and key words are then associated with the scenes of the video.

(Kritt, ¶0017.)¹ As Kritt notes in the passage above, both video and audio are processed. (See also Kritt, ¶¶0043 (“visual objects in a shot may be identified by application of one or more known image recognition processes to the shot”), 0059 (“Segments of the video for which the audio feature values are similar to those that are characteristic of speech may be classified as speech.”), 0065 (“In operation **612**, an audio transcript may be determined. An audio transcript may include all of the words spoken in the video....[S]poken words may be determined from audio features classified as speech using any known technique”).)

58. With respect to claim 1[a] specifically, and concepts of distributed/multi processing, Kritt describes modules for processing video and audio of a video file, and that those modules can be distributed on different computer systems:

The memory **104** may store all or a portion of the following: an audio visual file container **150** (shown in FIG. 2 as container **202**), a video processing module **152**, an audio processing module **154**, and a control module **156**. These modules are illustrated as being included within the memory **104** in the computer system **100**, however, in other embodiments, some or all of them may be on different computer

¹ All underlining added to quoted material unless noted.

systems and may be accessed remotely, e.g., via a network.

(Kritt, ¶0022; *see also id.*, ¶0024 (“The video processing module **152** may include various processes that generate visual tags according to one embodiment. The audio processing module **154** may include various processes for generating audio and key word tags according to one embodiment.”).)

59. Kritt further describes how its systems could be provided using “cloud-computing infrastructure.” (Kritt, ¶¶0078-0079.) Kritt explains, as was known to a person of ordinary skill in the art, that “[c]loud computing generally refers to the provision of scalable computing resources as a service over a network.” (Kritt, ¶0078.) “More formally,” Kritt explains, “cloud computing may be defined as a computing capability that provides an abstraction between the computing resource and its underlying technical architecture (e.g., servers, storage, networks), enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.” (Kritt, ¶0078.) Kritt further explains that one benefit of cloud computing is that “[t]ypically, cloud-computing resources are provided to a user on a pay-per-use basis, where users are charged only for the computing resources actually used (e.g., an amount of storage space used by a user or a number of virtualized systems instantiated by the user).” (Kritt, ¶0079.) It also

allows a user to “access any of the resources that reside in the cloud at any time, and from anywhere across the Internet.” (Kritt, ¶0079.)

60. I will provide more information about Kritt in my discussion of the claim limitations below.

(c) Steinberg (EX1005)

61. **Steinberg**, U.S. Patent No. 7,574,016 B2, is entitled “Digital Image Processing Using Face Detection Information.” I am informed that Steinberg qualifies as prior art because it was both filed (on June 26, 2003) and issued (on August 11, 2009) before the earliest priority date of the challenged claims.

62. I point to Steinberg primarily in connection with claim 1**[b]** and its recitation of an image detector that is “**adjustable to increase detection of non-primary images in the video.**”

63. Like Zhao, Steinberg relates to facial detection in digital images. (Steinberg, 1:49-53.) With respect to claim 1**[b]**, Steinberg describes a “face detection process” to find faces in a captured image. (Steinberg, 18:25-31 (“An image-detection process, preferably a face detection process, is applied to the captured image to seek faces in the image, **1020.**”); *see generally id.*, 18:29-44, 19:18-36.) Steinberg explains that the face detection process can be implemented “by changing the sensitivity or threshold of the detection process.” (Steinberg,

18:35-37.) Steinberg describes how “each of the faces is given a weight based on size of the faces, location within the frame, other parameters described herein, etc., which define the importance of the face in relation to other faces detected.” (Steinberg, 18:40-44.) Therefore, for example, “in response to false detection of one or more regions as one or more faces, or in response to a determination to concentrate on less image faces than faces identified in the identifying step,” “faces may be removed by increasing a sensitivity level of the face identifying step[.]” (Steinberg, 11:66-12:7; *see also id.*, 11:3-18 (“An automatically detected face may be removed in response to false detection of regions as faces, or in response to a determination to concentrate on less image faces or images faces that were manually determined to be of higher subjective significance, than faces identified in the identifying step. A face may be removed by increasing a sensitivity level of said face identifying step.”).) As I explain below, although Steinberg focuses on increasing a sensitivity level to *decrease* detection of less important (i.e., non-primary) faces, a person of ordinary skill in the art would have understood that decreasing the sensitivity level would have the opposite effect of *increasing* detection of non-primary faces in the image.

64. I will provide more information about Steinberg in my discussion of the claim limitations below.

(d) Kouzani (EX1006)

65. **Kouzani**, entitled “Fractal Face Representation and Recognition,” is a paper published in the Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics. I am informed that Kouzani qualifies as prior art because it was published in 1997, before the earliest priority date of the challenged claims.

66. I point to Kouzani in my analysis below primarily in connection with claim 1[c] and its recitation of comparing an image of the object to a “**fractal**” and updating “**the fractal**” with the image.

67. The use of fractals for facial recognition was already known before the ’853 patent. In fact, the ’853 patent identifies the use of fractals among a number of known recognition techniques, including Principal Component Analysis (PCA) and Elastic Bunch Graph Matching (EBGM), among others. (’853, 14:30-15:47 (“[T]he following recognition models can be utilized, alone or in combination, and are provided to illustrate exemplary alternative embodiments...”); *see also, e.g.*, Zhao, ¶0072 (Elastic Bunch Graph Matching); Kouzani, p.007 (Principal Component Analysis).) It was known, for instance, that “[a] human face is a surface lying in the fractal space intrinsically. Since many of the parameters are self-similar, the fractal model can better represent faces, especially in handling facial variations, such as

pose, illumination, etc.” (’853, 15:40-44.) For example, a 2010 paper entitled “Facial Image Recognition Based on Fractal Image Encoding” (EX1014) explains:

Fractal image encoding uses self-similarity to encode an image. Utilizing this feature, a new method of facial image recognition based on fractal image encoding is examined. Fractal image encoding can approximate any given image by capturing intrinsic self-similarity within the image. This method has some invariance to rotation, scaling, translation, and luminance.

(EX1014, p.005.) Even as of 2010, the use of fractals for facial recognition was not at all new—Kouzani, for example, dates from 1997. (EX1006, p.007.)

68. Kouzani “presents a face representation and recognition model based on the theory of fractals.” (Kouzani, p.007.) Kouzani explains that “[a] fractal is a fragmented geometric shape that can be subdivided in parts, each of which is a reduced-size copy of the whole[,]” that “[f]ractals model many real-world objects that do not correspond to simple geometric shapes, e.g. clouds, mountains, turbulence, and coastlines [9],” and that “[f]ractals are generally self-similar and independent of scale.” (Kouzani, p.007.) In Kouzani’s system, “[e]ach input face image is represented by its fractal model which is a small collection of contractive transformation parameters.” (Kouzani, p.007.) “For recognition, the input face image is transformed and its fractal model is then compared against the database of

fractal models of known faces.” (Kouzani, p.007.) Kouzani explains a comparison process for recognizing a face performed with a neural network:

The second network has also a feedforward architecture and has been trained to recognise faces. It receives two group of inputs, the fractal model of an input image and the fractal model of a known face from the fractal database. It calculates the degree of similarity between the two fractal face models. If the output is below a threshold, the input face is considered as a known face whose fractal is currently under examination. Similar to the compression stage, more networks with the same architecture can be used to speed up the recognition process.

(Kouzani, p.009.)

69. Kouzani also touts advantages of using fractal models for facial recognition, including being less computationally expensive than prior art principal component analysis (PCA) techniques and providing high data compression:

In the recognition stage, two groups of experiments were performed. In the first experiment, the known face images were presented to the system and the recognition process was carried out. 100% recognition was achieved in this step. The maximum compression ratio obtained for successful recognition of all known faces was 89:1. Figure 6 demonstrates recognition of the known face number 100. In the second experiment, 50 unknown face images were directed to the system and the recognition process was carried out. None of the input images were

recognised by the system. Figure 7 shows the recognition result for an unknown input face image.

The proposed system is less computationally expensive than that for PCA. Moreover, when a new face is added to the database, recalculation of eigenfaces and re-compression of known faces are not necessary.

(Kouzani, p.009; *see also id.* (“Simulation results show that the proposed system overcomes the shortcomings of PCA.”).) Kouzani further notes that “real-time recognition” was achieved. (Kouzani, pp.008-009 (“The main reason for employing neural networks is to take advantage of the highly parallel architecture of neural networks in order to achieve real-time recognition.”).)

70. I will provide more information about Kouzani in my discussion of the claim limitations below.

2. Additional References

(a) Yang (EX1007)

71. **Yang**, entitled “Detecting Faces in Images: A Survey,” is a paper published in IEEE Transactions On Pattern Analysis And Machine Intelligence. I am informed that Yang qualifies as prior art because it was published in 2002, before the earliest priority date of the challenged claims.

72. I point to Yang (along with Romdhani (EX1008)) in my analysis below

primarily in connection with claim 5 and its recitation of an object recognizer that is
“configured to determine skin textures of the object.”

73. I note that Steinberg (EX1005) specifically identifies Yang in connection with face detection techniques using “identification of facial features, texture and skin color” (Steinberg, 1:62-65) and moreover, expressly incorporates Yang by reference (Steinberg, 7:1-3; *see generally id.*, 6:40-7:6 (“Incorporation by Reference” section)). I understand, therefore, that under the patent laws, Yang is effectively part of Steinberg as if Yang were explicitly contained in Steinberg.

74. Yang reviews various techniques for detecting faces in images and provides more detail about facial recognition involving **“skin textures.”** Yang explains that “[h]uman faces have a distinct texture that can be used to separate them from different objects.” (Yang, p.009.) Yang describes how “Augusteijn and Skufca developed a method that infers the presence of a face through the identification of face-like textures” and “[t]o infer the presence of a face from the texture labels, they suggest using votes of the occurrence of hair and skin textures.” (Yang, p.009.) Yang also describes another technique that uses a “face texture model” that incorporates color information. “Using the face texture model, they design a scanning scheme for face detection in color scenes in which the orange-like parts including the face areas are enhanced.” (Yang, p.009.)

75. I will provide more information about Yang in my discussion of the claim limitations below.

(b) Romdhani (EX1008)

76. **Romdhani**, entitled “Face Identification by Fitting a 3D Morphable Model using Linear Shape and Texture Error Functions,” is a paper published in the European Conference on Computer Vision. I am informed that Romdhani qualifies as prior art because it was published in 2002, before the earliest priority date of the challenged claims.

77. I point to Romdhani primarily in connection with claims 5-7 for its teachings regarding “**skin textures**” and a “**three-dimensional model**.”

78. I note that, similar to Yang, Romdhani is also cited by Steinberg (EX1005). (Steinberg, p.005 (references cited).)

79. With respect to “**skin textures**” and claim 5 of the ’853 patent, Romdhani describes techniques for face recognition that involve matching face texture to an image of a known individual. (Romdhani, p.0024 (“Abstract”) (“This paper presents a novel algorithm aiming at analysis and identification of faces viewed from different poses and illumination conditions. Face analysis from a single image is performed by recovering the shape and textures parameters of a 3D Morphable Model in an analysis-by-synthesis fashion.”).) Romdhani explains the

face recognition techniques in great detail, describing how “[t]he fitting procedure provides shape and texture coefficient which ... depends of [sic] the identity only. Identification of a novel face image is performed by matching the coefficients recovered from the image with the coefficients recovered from the image of a known individual.” (Romdhani, p.0036.) Romdhani explains that “[s]tate of the art results were obtained” (Romdhani, p.0039), outperforming “one of the most successful commercial face recognition system[s]” (Romdhani, p.0037).

80. With respect to a **“three-dimensional model”** and claims 6-7 of the ’853 patent, Romdhani presents an algorithm, called “LiST” (for “Linear Shape and Texture,” that involves “fitting a 3D Morphable Model to face images.” (Romdhani, p.0039.) Romdhani explains that the algorithm “exploits the linearity of the model in the recovery of shape and texture parameters leading to a robust and efficient algorithm” and “[s]tate of the art results were obtained” when “LiST was evaluated on the difficult problem of combined pose-illumination variation face identification.” (Romdhani, p.0039.) Romdhani notes that their face identification results outperformed “one of the most successful commercial face recognition system[s]” and that the results **“clearly show that, for pose-invariant face identification, a 3D representation is valuable.”** (Romdhani, p.0037.)

81. I will provide more information about Romdhani in my discussion of

the claim limitations below.

(c) Trivedi (EX1009)

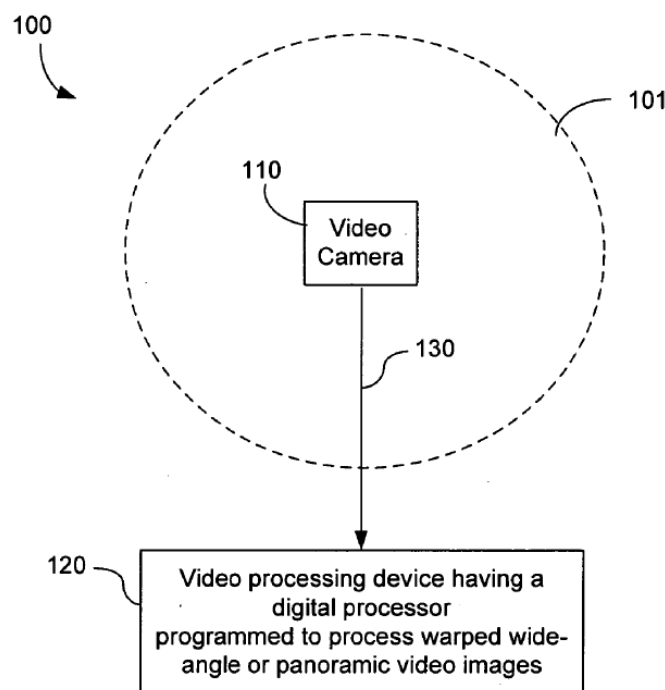
82. **Trivedi**, U.S. Patent Application Publication No. 2006/0187305 A1, is entitled “Digital Processing of Video Images.” I am informed that Trivedi qualifies as prior art because it was both filed (as a PCT application on July 1, 2003) and published (on August 24, 2006) before the earliest priority date of the challenged claims.

83. I point to Trivedi in my analysis below in connection with claims 8-9 of the ’853 patent. I understand that the Examiner cited Trivedi during prosecution in connection with claims that ultimately issued as claims 8 and 9, and was not distinguished by the applicant. (EX1015 (’853 file wrapper), pp.00145 (“Trivedi further teaches further comprising a processor and a camera for capturing the video”; “Trivedi further teaches wherein the processor is configured to transform the video into rectilinear format”), 00167-169 (Applicant Reply to Office Action of October 29, 2020).)

84. With respect to claim 8 of the ’853 patent, which recites that the system of claim 1 further comprises “**a processor and a camera for capturing the video,**” Trivedi confirms that video cameras are used in security/surveillance systems and expressly discloses a security system that includes a camera:

Video cameras have been used to capture video images for various applications such as surveillance, machine vision, security monitoring, inspection, sensing and detection. In these and other applications, the captured video images may be sent to a nearby or remote image processing device to be processed and displayed. Digital image processing techniques may be used to process the video images to extract information from the video images.

(Trivedi, ¶0003.) As shown in Figure 1, reproduced below, Trivedi discloses a video system **100** that includes a video camera **110** and video processing device **120**:



(Trivedi, Fig. 1; *see also id.*, ¶0040 (“FIG. 1 illustrates one example of a video system **100** where a video camera **110** is installed at a location **101** to capture a video of the location **101**....A video processing device **120** is located at a different location

and is linked to the video camera **110** by a communication link **130** to receive the video signal.”.)

85. With respect to claim 9, which recites “**transform[ing] the video to a rectilinear format**,” Trivedi describes capturing video in a warped, non-rectilinear format—for example, video captured with a wide angle lens—and “efficiently transform[ing] warped video images into rectilinear video images”:

Certain video cameras are designed with optical imaging modules to capture wide-angle or panoramic video images. Such video images are distorted due to the designs of various optical imaging modules. Digital imaging processing may be used to transform the distorted images into images that a viewer would normally perceive when directly look [sic] at the scene being captured.

(Trivedi, ¶¶0004, 0005.) Trivedi explains that the transform of video into a rectilinear format is performed by the digital processor of the system’s video processing device. (Trivedi, ¶0042 (“[T]he digital processor may be programmed, among other features and functions, to transform the warped video images into rectilinear video images[.]”); *see also id.*, ¶0041 (“The video processing device **120** includes a digital processor that is programmed to process the warped wide-angle or panoramic video images to extract desired information.”), Fig. 2 (showing device **120**, stating “Video processing device having a digital processor programmed to

process warped wide-angle or panoramic video images”).)

86. I will provide more information about Trivedi in my discussion of the claim limitations below.

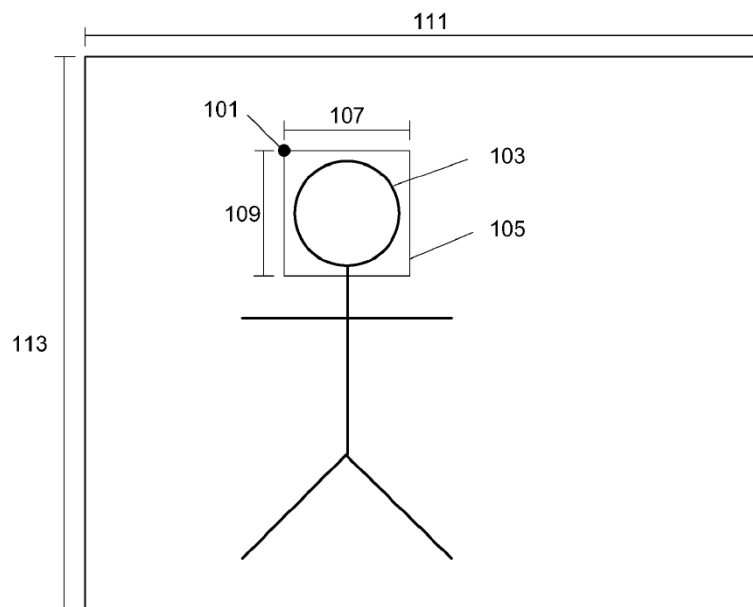
(d) Singer (EX1010)

87. **Singer**, U.S. Patent Application Publication No. 2011/0305394 A1, is entitled “Object Detection Metadata.” I am informed that Singer qualifies as prior art because it was both filed (on June 15, 2010) and published (on December 15, 2011) before the earliest priority date of the challenged claims.

88. I point to Singer in my analysis below in connection with claims 10-11 of the ’853 patent. I understand that the Examiner cited Singer during prosecution in connection with claims that ultimately issued as claims 10 and 11, and was not distinguished by the applicant. (EX1015 (’853 file wrapper), pp.00147-00148 (“Singer...teaches wherein the coordinator is configured to embed metadata about the object into the video, and wherein the metadata comprises a timestamp and a coordinate location of the object in the still frame images”; “Singer further teaches wherein the coordinator is configured to generate a metadata stream corresponding to the image, wherein the metadata stream includes one or more timestamps corresponding to the image, and wherein the coordinator is configured to embed the metadata stream in the video”), 00167-168 (Applicant Reply to Office Action of

October 29, 2020).)

89. Singer describes techniques for generating and storing metadata about objects detected in video image data, including face detection metadata. (Singer, ¶¶0023-0024.) For example, Figure 1 of Singer illustrates detection of a face in an image of a video:



(Singer, Fig. 1.) Singer explains that “the location of each face in a frame of image data may be described with a simple rectangle.” (Singer, ¶0025.) For example, in Figure 1, “[f]ace detection rectangle **105**,” which bounds face **103**, “has dimensions of height **109** and width **107** and an initial point **101**.” (Singer, ¶0026.) Singer in figure 9 illustrates an example of metadata for a face detection rectangle:

Detected-face entry 903

X = 15
Y = 20
Width = 32
Height = 26
ID = 3
Layer = 2
DetectWeight = 243
MatchWeight = 173

(Singer, Fig. 9.) Singer explains that “Detected-object entry **903** provides one example of a detected-face entry in a metadata track of image data.” (Singer, ¶0053.) The X and Y values, and Width and Height values, specify the location and dimensions, respectively, of the detected face in the image. (Singer, ¶0053.)

90. I note that Singer’s depiction and description of detection rectangles matches Zhao’s depiction of using rectangles to identify detected objects in a still frame image of a video:

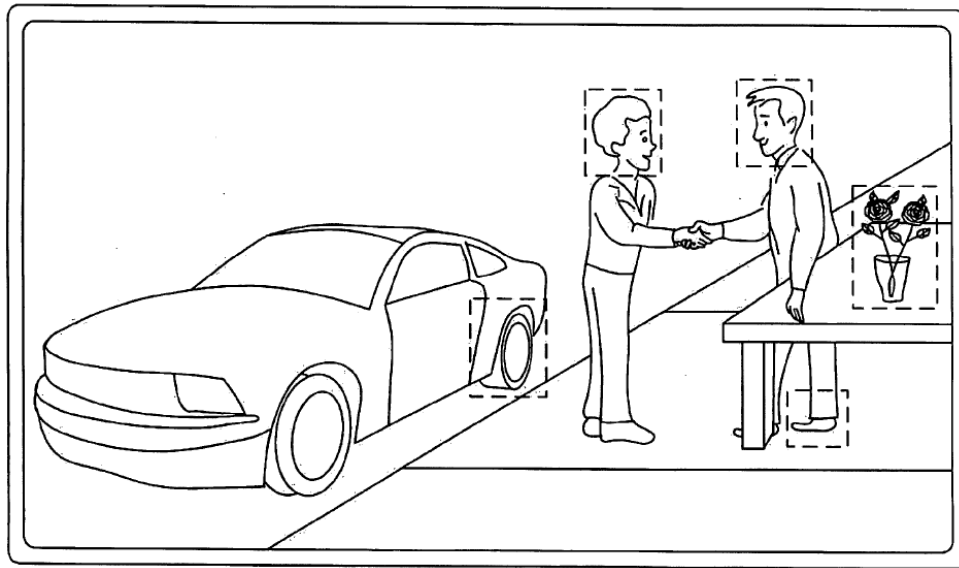


FIG. 1

(Zhao, Fig. 1, ¶0003 (“For example, FIG. 1 illustrates a sample frame of a video scene. Exemplary objects that are capable of being recognized within the illustrated video include characters’ faces, a plant in a vase, a shoe, and an automobile tire, each of which is shown within a dashed box to indicate its detection and recognition within the frame.”).) This suggests that Zhao collects similar coordinate location metadata for detected (and recognized) objects.

91. Singer further explains that the object detection metadata may be stored “as timed metadata associated with a video track,” such as a track of a video in QuickTime File Format. (Singer, ¶0023.) “In time-based metadata, each detected-object entry applies to one or more temporally aligned frames of video data.”

(Singer, ¶0056.) Figure 13 of Singer, reproduced below, illustrates an example of temporally-aligned, timed metadata for a detected face embedded in a video:

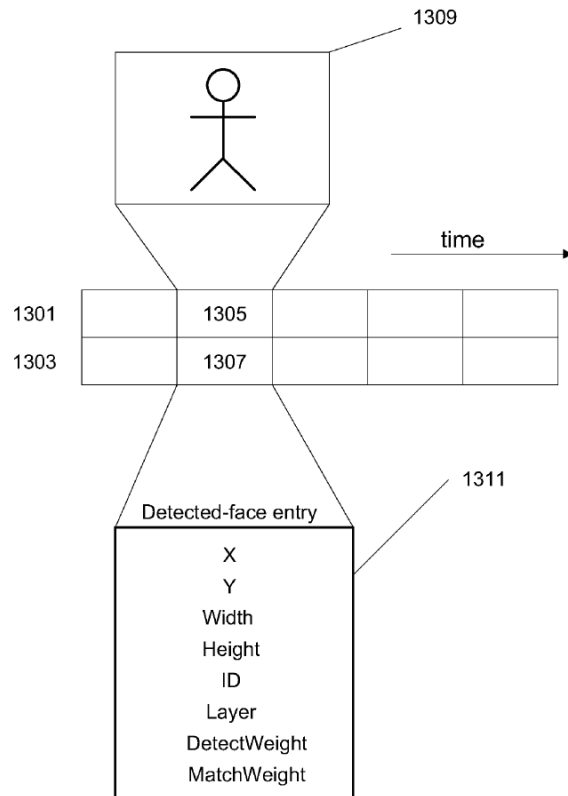


Fig. 13

(Singer, Fig. 13.) Referring to figure 13, Singer explains that “[t]rack **1301** includes frames of image data over time which may be replayed as a video” and “track **1303** includes detected object metadata entries.” (Singer, ¶0068.) “Frame **1305** in track **1301** corresponds to image frame **1309**,” while “metadata entry **1307** corresponds to detected face entry **1311**.” (Singer, ¶0068.) Detected face entry **1311** “includes

metadata describing a detected object in frame **1305**, such as the face of the person captured in image data **1309**.” (Singer, ¶0068.) The metadata can be embedded into a video that is “streamed to a remote client.” (Singer, ¶0040.)

92. I will provide more information about Singer in my discussion of the claim limitations below.

B. Ground 1: Claims 1-4 Are Obvious Over Zhao in View of Kritt, Steinberg, and Kouzani

1. Independent Claim 1: “A system for generating data from a video, comprising:” (Claim 1[pre])

93. The preamble of claim 1 recites “[a] **system for generating data from a video.**” Assuming the preamble provides a claim limitation, Zhao discloses it. Zhao discloses a system that performs detection and recognition of objects, such as people and faces, in video images to generate data, including by creating and updating object models used for detection and recognition, and by generating data about the video that can be used for indexing and retrieval. Figure 2 of Zhao, which I will discuss in more detail in my analysis below, illustrates at a high level this overall process of generating data from a video performed by the system:

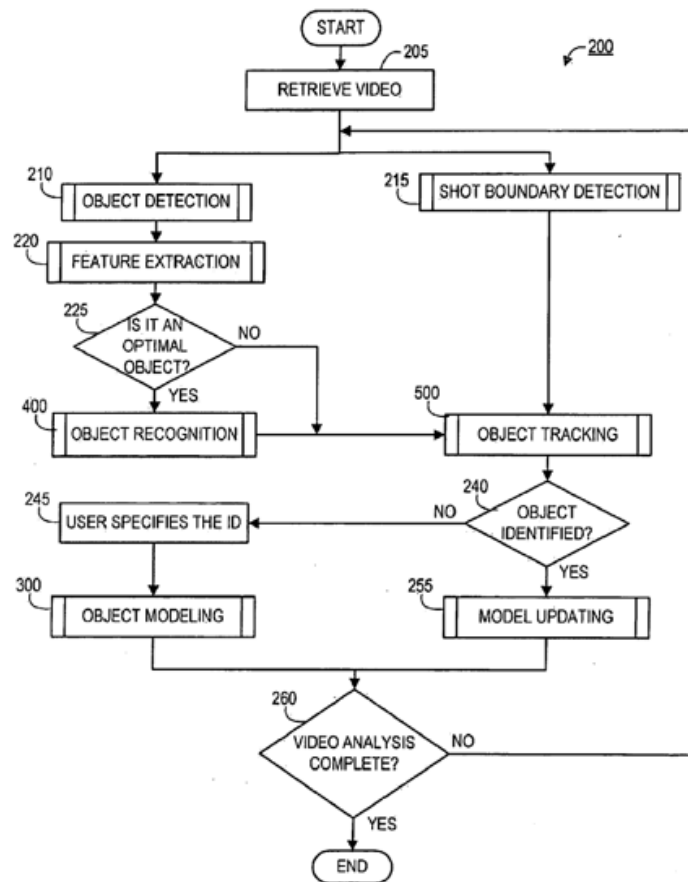


FIG. 2
OVERALL PROCESS

(Zhao, Fig. 2; *see also id.*, ¶0036 (“FIG. 2 illustrates a flowchart describing the steps involved, from a high-level, in one embodiment of the present system for detecting, modeling, recognizing, and tracking object images throughout one or more videos.”).)

94. Zhao explains, for example, that it “generally relate[s] to systems and methods for detection, modeling, recognition, and tracking of objects within video content” and to “indexing and retrieval systems for videos based on generated object

models.” (Zhao, ¶0051.) “These objects include people, faces, articles of clothing, plants, animals, machinery, electronic equipment, food, and virtually any other type of image that can be captured or presented in video.” (Zhao, ¶0051.) Zhao’s system “may be operated in a computer environment” and “any results or outputs relating to detection, modeling, recognition, indexing, and/or tracking of object images within videos may be stored in a database” or otherwise output. (Zhao, ¶0055.) Zhao states that its system is “useful for a wide variety of applications, including video indexing and retrieval, video surveillance and security, unknown person identification, advertising, and many other fields.” (Zhao, ¶0053.)

95. Therefore, Zhao discloses “**a system for generating data from a video.**” And as explained below, the system as claimed is obvious over Zhao, Kritt, Steinberg, and Kouzani.

- (a) “**a coordinator communicatively coupled to a splitter and to a plurality of demultiplexer nodes, wherein the splitter is configured to segment the video, wherein the demultiplexer nodes are configured to extract audio files from the video and to extract still frame images from the video;**” (Claim 1[a])

96. As explained below, claim 1[a] is disclosed by and obvious over Zhao in view of **Kritt**.

97. For context, the ’853 patent describes a “**coordinator,**” “**splitter,**” and

“**demultiplexer**” in terms of functions they may perform, and does not state any particular form or structure. The ’853 patent states, for instance, that “[s]ource media can be provided to a coordinator, which can be a non-visual component that can perform one or more of several functions.” (’853, 16:26-27.) For example, “[t]he coordinator can direct distributed processing and aggregation.” (’853, 18:16-17.) Referring to figure 10, reproduced below, the ’853 patent states that “the coordinator can send uploaded source media to a splitter and a demuxer/demultiplexer”:

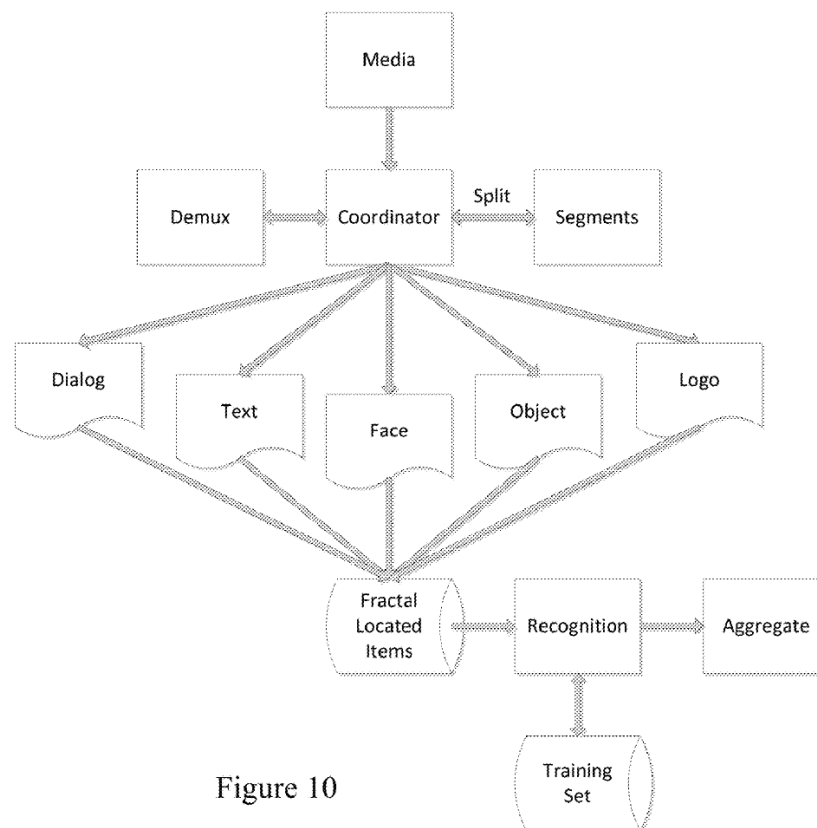


Figure 10

(’853, Fig. 10, 16:60-63; *see also id.*, 3:50-51.) A “**splitter**” can be invoked to “‘slice’ [media] assets into media segments and/or multiple sub assets comprising the entire stream.” (’853, 17:32-37.) A “**demultiplexer**” can “extract audio files from the video and/or to extract still frame images from the video.” (’853, 1:45-47.) For example, “a demultiplexer component can strip audio data from segments and store audio streams for future analysis.” (’853, 17:61-62.)

98. A person of ordinary skill in the art would recognize that the arrangement of a “**coordinator**,” “**splitter**,” and “**plurality of demultiplexer nodes**” recited in claim 1[a] is merely an application of the well-known concept of distributed processing (or multiprocessing) to video processing, to obtain the obvious benefits that processing time can be reduced, and more processing overall can be performed, by sharing the workload and performing multiple operations in a parallel fashion. The ’853 patent likewise explains that “the splitter can be configured to segment the video,” then later states, for example, that “[t]he video-to-data engine can segment the video into chunks for distributed, or parallel, processing....Distributed processing in this context can mean that the processing time for analyzing a video from beginning to end is a fraction of the play time of the video. This can be accomplished by breaking the processes into sections and processing them simultaneously.” (’853, 2:35-36, 6:15-21.) By 2016,

58

distributed/multi processing was well-known and neither novel nor non-obvious. (EX1012 (Microsoft Computer Dictionary 5th ed. 2002), pp.005 (entry for “distributed processing” explaining that “distributed processing shares the workload among computers that can communicate with one another”), 006 (entry for “multiprocessing” explaining “objective is increased speed or computing power”).) And as explained below, claim 1[a] is obvious over Zhao in view of Kritt.

99. Turning to Zhao, I will first explain how Zhao teaches both “**splitter**” operations and “**demultiplexer**” operations. I will then explain how Zhao discloses and renders obvious a “**coordinator**” before turning to the combination with Kritt.

100. Zhao itself discloses a “**splitter**” “**configured to segment the video**” in the form of processes that perform “shot/scene boundary detection” to split a video into shots or scenes that are extracted for processing. Zhao explains:

Once the video has been retrieved, the video undergoes a process of detection of objects and extraction of object features from video images (processes **210**, **220**). Concurrently, shots or scenes are detected in the video via a shot/scene boundary detection procedure (process **215**). The shots or scenes of the video are extracted to provide a sequence of images for analysis by the system, and each detected shot or scene is individually analyzed. Generally, the term “shot” or “scene” refers to a grouping of frames or images recorded by either a stationary or smoothly-moving camera, with little or no background change between

the frames, corresponding to one continuous time period.

(Zhao, ¶0057.) Figure 2 of Zhao illustrates this process with shot boundary detection process 215:

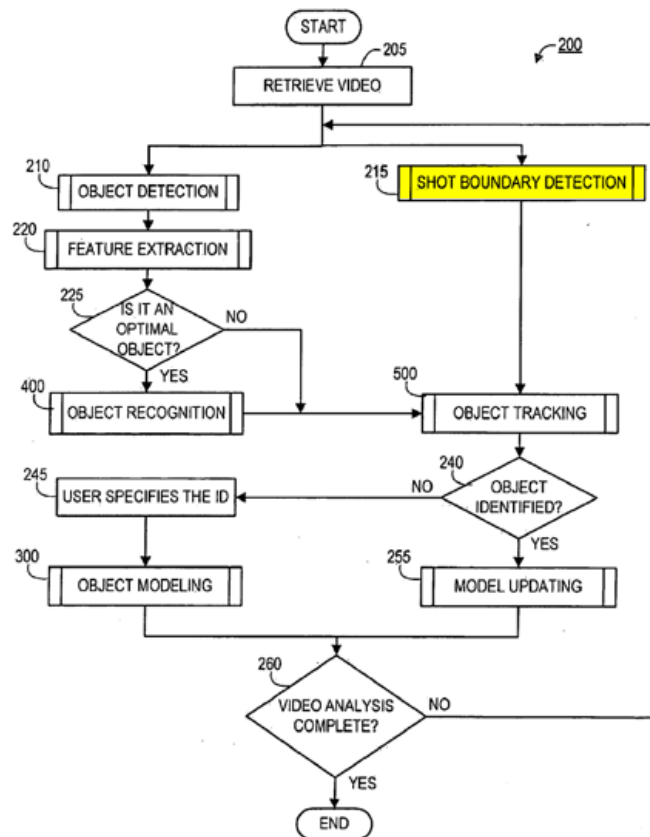


FIG. 2
OVERALL PROCESS

(Zhao, Fig. 2 (highlighted added).)

101. With respect to the “**demultiplexer**,” Zhao also discloses at least some of the claimed “**demultiplexer**” functionality—i.e., “**extract[ing] still frame images from the video.**” Zhao describes “receiving a video file, wherein the video

file comprises a plurality of video frames.” (Zhao, ¶0023.) Zhao teaches that the object detection and recognition processes involve extracting still frame images from the video. For example, Zhao explains that the object recognition process is performed on an “image or frame extracted from a video.” (Zhao, ¶0075 (“The process **400** shown in FIG. 4 is for one image or frame extracted from a video, but as will be appreciated, this process is typically repeated for each optimal object image in the video.”).) Zhao similarly explains that the object detection process involves “detection of objects and extraction of object features from video images (processes **210, 220**).” (Zhao, ¶0057; *see also id.*, ¶0059 (“Once an object has been detected within a video frame...”).)

102. With respect to “**demultiplexer**” functionality to “**extract audio files from the video**,” I note that the claims of the ’853 patent do not recite any additional limitations related to “audio” or the extracted “audio files.” Zhao likewise does not discuss audio information, which is not surprising because audio processing is not a focus of Zhao’s purported invention. But it would have been obvious for Zhao’s system to further include processes for “**extracting audio files from the video**.” For example, Zhao discloses performing its techniques on videos such as episodes of television shows such as the Gilmore Girls (Zhao, e.g., ¶¶0040-0041, 0093, Fig. 6), which a person of ordinary skill in the art would have understood and found

obvious to have included one or more audio tracks. Zhao more generally teaches that its techniques are “useful for a wide variety of applications, including video indexing and retrieval, video surveillance and security, unknown person identification, advertising, and many other fields.” (Zhao, ¶0053.) Given Zhao’s broad applicability, a person of ordinary skill in the art would have readily appreciated and found it obvious that videos processed by Zhao’s system would include audio tracks. A person of ordinary skill in the art would have found it obvious to extract audio tracks from the video into separate audio files—for example, to separately store or analyze—but again, those sorts of preparatory operations are not a focus of Zhao. For example, a person of ordinary skill in the art would have understood and found it obvious to perform operations such as speech recognition or speaker recognition on the audio data. In a security and surveillance context (and numerous other contexts), for instance, these operations could be used to further confirm a person’s identity or to make a record of conversations associated with the captured video. For example, **Kritt** describes using audio to “validate that the human object was correctly identified” in a video. (Kritt, ¶0033 (“[A] visual tag may indicate that a particular human object appears in a shot and an audio tag identifies an audio signature of the human object is associated with the shot. In these examples, if the tags that are compared indicate the same object, the positive or

consistent result of the comparison may be used...to validate that the human object was correctly identified.”).)

103. A step of extracting audio tracks from a video file in Zhao would have resulted in “**extract[ing] audio files**” in at least two separate ways. First, it was well known to a person of ordinary skill in the art that video file formats took the form of “container” files (or “wrapper” files) that in turn stored video tracks, audio tracks, and other data as files within that container—much like how a single “.zip” archive file can store multiple files. For example, Kritt explains in its Background section:

A television show, movie, internet video, or other similar content may be stored on a disc or in other memory using a container or wrapper file format. The container format may be used to specify how multiple different data files are to be used. The container format for a video may identify different data types and describe how they are to be interleaved when the video is played. A container may contain video files, audio files, subtitle files, chapter-information files, metadata, and other files. A container also typically includes a file that specifies synchronization information needed for simultaneous playback of the various files.

(Kritt, ¶0002.) Similarly, the 2010 book, *Pro HTML5 Programming*, by Peter Lubbers et al., explains that “[a]n audio or video file is really just a *container* file, similar to a ZIP archive file that contains a number of files.” (EX1011, p.0013.) Figure 3-1 from the book, reproduced below, “shows how a video file (a video

container) contains audio tracks, video tracks, and additional metadata. The audio and video tracks are combined at runtime to play the video.” (EX1011, p.0013.)

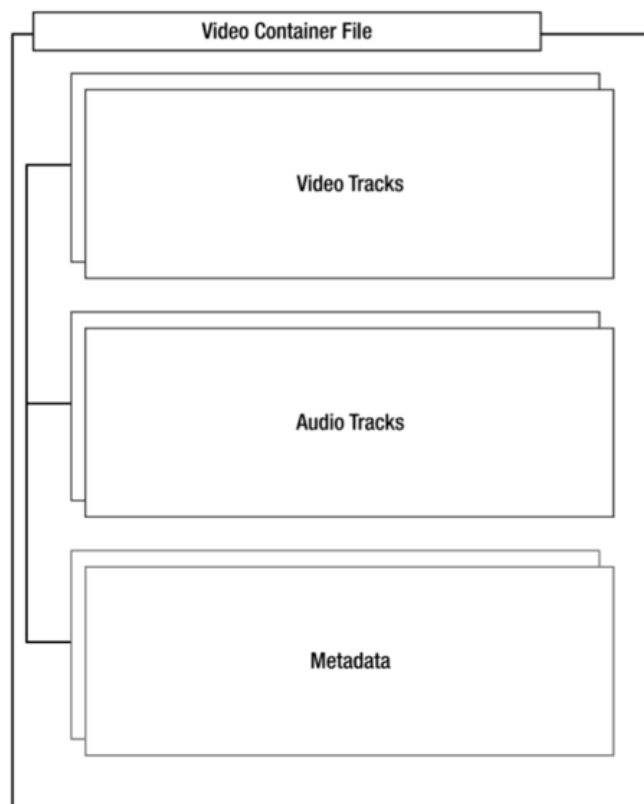


Figure 3-1. Overview of the video container

(EX1011, p.0014.) Therefore, a person of ordinary skill in the art would have understood that, by extracting audio tracks from a video file, a step of “**extract[ing] audio files from the video**” would be performed.

104. Second, by extracting audio from the video for separate analysis or storage, a person of ordinary skill in the art would have understood and found it obvious that the resulting extracted audio would take the form of “files.” This is

because, for starters, file storage was one of a finite number of ways of storing data. Moreover, it was routine to use files—known as “scratch files” or “temporary files”—for at least temporary storage while the associated data was processed. (*See, e.g.*, EX1012 (Microsoft Computer Dictionary (5th ed. 2002)), pp.007 (entries for “scratch” and “scratch file”), 008 (entry for “temporary file”).)

105. Turning back to the claimed “**coordinator**,” a person of ordinary skill in the art would have further understood that, in order to manage the various video file retrieval, splitter, and demultiplexer operations, Zhao’s system would also include “**coordinator**” functionality. For example, referring to figure 2, Zhao shows and explains how **splitter operations** (e.g., process **215**) and **demultiplexer operations** (e.g., processes **210, 400**) take place concurrently:

Once the video has been retrieved, the video undergoes a process of detection of objects and extraction of object features from video images (processes **210, 220**). Concurrently, shots or scenes are detected in the video via a shot/scene boundary detection procedure (process **215**).

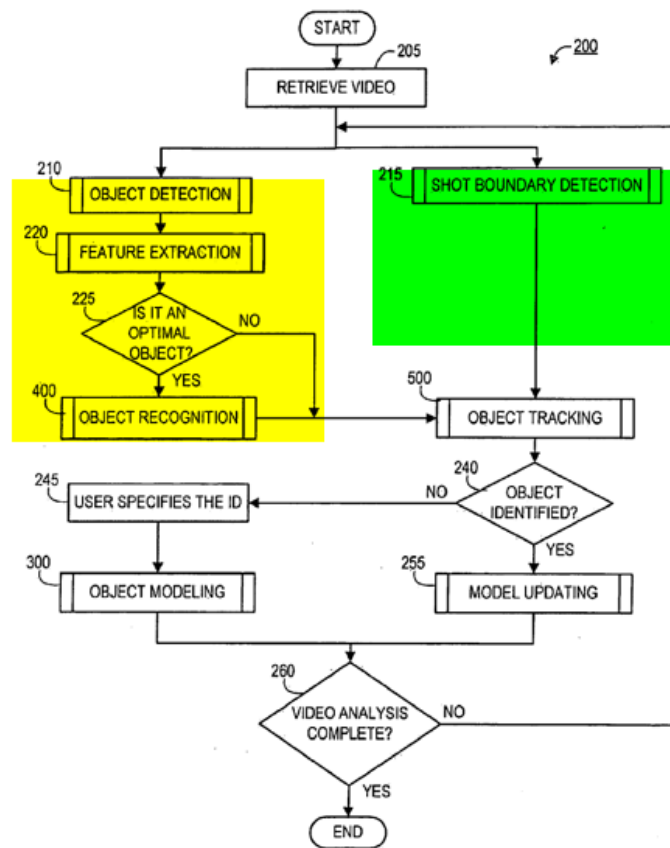


FIG. 2
OVERALL PROCESS

(Zhao, Fig. 2 (annotations added), ¶0057.) A person of ordinary skill in the art would have understood that various operations would need to be performed in order to retrieve a video file and then distribute the file for concurrent processing by processes 215 and 210, for example. Those operations, which I have highlighted in figure 2 of Zhao below, would form at least part of a “**coordinator.**”

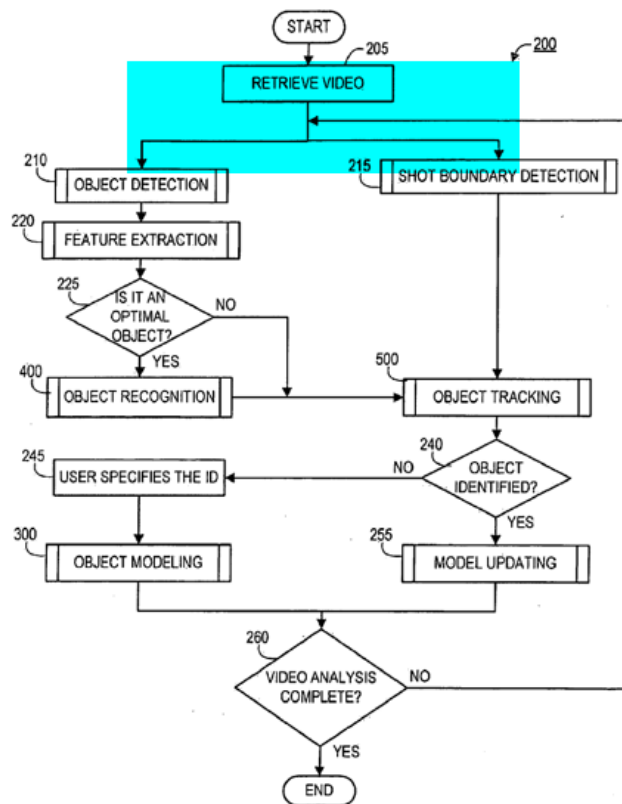


FIG. 2
OVERALL PROCESS

(Zhao, Fig. 2 (annotations added).) A person of ordinary skill in the art would have further understood that, for Zhao’s “coordinator” functionality to distribute the video file that it retrieved, it would need to be “**communicatively coupled**” to the system’s “splitter” and “demultiplexer” functionality in order to notify those components in some manner. In fact, Zhao’s disclosure of “concurrent” processing (Zhao, ¶0057) suggests that the video file is actually distributed to separate nodes (such as separate processors or computer systems) with which Zhao’s “coordinator” would be communicatively coupled. This is just like “coordinator” functionality described in

the '853 patent, which describes receiving a media file and distributing it to a splitter and demultiplexer. ('853, 16:60-63 (“the coordinator can send uploaded source media to a splitter and a demuxer/demultiplexer”).)

106. Lastly, before turning to the combination with Kritt, with respect to the claimed “**plurality of demultiplexer nodes**,” this too is suggested by Zhao. Although Zhao focuses on describing its system with respect to a single video file, Zhao explains that “the process may be extrapolated to analyze a plurality of videos.” (Zhao, ¶0057.) Because as noted above, Zhao also describes “concurrent” processing, a person of ordinary skill in the art would have therefore understood and found it obvious for Zhao’s system, in an implementation for processing a higher volume of video files, to comprise a “**plurality**” of nodes providing the claimed demultiplexer functionality. As discussed above, the benefits of distributed computing were well known and therefore would have been obvious to try. (*E.g.*, EX1012 (Microsoft Computer Dictionary 5th ed. 2002), pp.005 (entry for “distributed processing”), 006 (entry for “multiprocessing”).)

107. Turning to the combination with **Kritt**, like Zhao, Kritt describes a system for processing videos to generate data about the videos. For example:

According to various embodiments, a viewer selects a video, and human, nonhuman, and audio objects of the video are identified. In

addition, key words that are spoken by human objects in the video are identified. Human, nonhuman, and audio objects may be used to classify a particular segment of a video as a scene. The objects and key words are then associated with the scenes of the video.

(Kritt, ¶0017.) As Kritt notes in the passage above, both visual and audio data are processed. (*See also* Kritt, ¶¶0043 (“visual objects in a shot may be identified by application of one or more known image recognition processes to the shot”), 0059 (“Segments of the video for which the audio feature values are similar to those that are characteristic of speech may be classified as speech.”), 0065 (“In operation **612**, an audio transcript may be determined. An audio transcript may include all of the words spoken in the video....[S]poken words may be determined from audio features classified as speech using any known technique.”).)

108. With respect to claim 1[a] specifically, and concepts of distributed/multi processing, Kritt describes modules for processing video and audio of a video file, and that those modules can be distributed on different computer systems:

The memory **104** may store all or a portion of the following: an audio visual file container **150** (shown in FIG. 2 as container **202**), a video processing module **152**, an audio processing module **154**, and a control module **156**. These modules are illustrated as being included within the

memory **104** in the computer system **100**, however, in other embodiments, some or all of them may be on different computer systems and may be accessed remotely, e.g., via a network.

(Kritt, ¶0022; *see also id.*, ¶0024 (“The video processing module **152** may include various processes that generate visual tags according to one embodiment. The audio processing module **154** may include various processes for generating audio and key word tags according to one embodiment.”).)

109. Kritt further describes how its systems could be provided using “cloud-computing infrastructure.” (Kritt, ¶¶0078-0079.) Kritt explains, as was known to a person of ordinary skill in the art, that “[c]loud computing generally refers to the provision of scalable computing resources as a service over a network.” (Kritt, ¶0078.) “More formally,” Kritt explains, “cloud computing may be defined as a computing capability that provides an abstraction between the computing resource and its underlying technical architecture (e.g., servers, storage, networks), enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.” (Kritt, ¶0078.) Kritt further explains that one benefit of cloud computing is that “[t]ypically, cloud-computing resources are provided to a user on a pay-per-use basis, where users are charged only for the

computing resources actually used (e.g., an amount of storage space used by a user or a number of virtualized systems instantiated by the user).” (Kritt, ¶0079.) It also allows a user to “access any of the resources that reside in the cloud at any time, and from anywhere across the Internet.” (Kritt, ¶0079.) As discussed above for Zhao, a person of ordinary skill in the art would have recognized that in such a distributed/multi processing system, “**coordinator**” functionality would be needed to manage the operations and workloads among the various video and audio processing nodes.

110. **Rationale and Motivation to Combine (Zhao with Kritt)**. It would have been obvious to a person of ordinary skill in the art to incorporate Kritt’s teachings regarding using a distributed/multi processing system (such as a pool of nodes in a cloud-computing system) for processing images and audio into Zhao’s video processing system. The combination would have involved the conventional application of distributed processing techniques to Zhao’s system and would have predictably resulted in Zhao’s system in which multiple nodes were used for processing images and audio of a retrieved video file, along with functionality that managed that processing. The combination therefore would have resulted in “**a coordinator**”—i.e., functionality for managing the overall distributed process of analyzing images and audio of a video file—“**communicatively coupled to a**

splitter”—i.e., functionality to split a video into shots/segments—“and to a plurality of demultiplexer nodes”—e.g., multiple nodes in a cloud computing infrastructure for performing image and audio recognition. The combination therefore would have resulted in a “splitter [] configured to segment the video,” and “demultiplexer nodes [] configured to extract audio files from the video and to extract still frame images from the video.”

111. Zhao and Kritt are analogous references to the '853 patent. Like the '853 patent, both Zhao and Kritt are in the field of video processing and generating data from video content. ('853, 1:34-36 (“The present invention is generally directed to a method to generate data from video content, such as text and/or image-related information.”); e.g., Zhao, ¶0036, Fig. 2; e.g., Kritt, ¶0017.) Zhao and Kritt would have also been reasonably pertinent to problems facing the '853 inventors, including techniques for recognizing objects in video.

112. A person of ordinary skill in the art would have been motivated to combine. As noted above, although Zhao focuses on describing its system with respect to a single video file, Zhao explains that “the process may be extrapolated to analyze a plurality of videos.” (Zhao, ¶0057.) Zhao also teaches that its techniques are “useful for a wide variety of applications, including video indexing and retrieval, video surveillance and security, unknown person identification, advertising, and

many other fields.” (Zhao, ¶¶0053.) A person of ordinary skill would have therefore recognized that an application of Zhao’s system could involve high-volume video processing, along with potential time constraints—for example, a large-scale video indexing and retrieval website, or a commercial video surveillance and security service. A person of ordinary skill in the art would have therefore looked to an analogous reference such as Kritt in the same field as Zhao that provides further teachings regarding system architectures capable of providing such a larger-scale service. Kritt also provides express motivation to combine, describing numerous advantages of a cloud-based system, including “convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction,” “resources [] provided to a user on a pay-per-use basis,” and “access any of the resources that reside in the cloud at any time, and from anywhere across the Internet.” (Kritt, ¶¶0078-0079.) Moreover, with respect to audio processing, as explained above, Kritt also explains integrating audio and image processing would beneficially enable confirmation of recognized visual objects in the video, as well as a transcript that could be used for further analysis of the video. (Kritt, e.g., ¶¶0033 (using audio to “validate that the human object was correctly identified” in a video), 0065 (audio transcript).)

113. A person of ordinary skill in the art would have had a reasonable expectation that the combination would have been successful. It would have involved the straightforward application of conventional techniques for distributed processing (and audio processing). As Kritt notes, cloud computing for example was already well-established, as were techniques for speech and speaker recognition.

(b) **“an image detector configured to detect an image of an object in the still frame images, wherein the image detector is adjustable to increase detection of non-primary images in the video; and” (Claim 1[b])**

114. As explained below, claim 1[b] is obvious over Zhao in view of **Steinberg**.

115. The '853 patent describes image “detection” and “recognition” as a two-step process, first “detection,” then “recognition”:

FIG. 8 illustrates a flow diagram of an embodiment of image recognition. Images can be classified as faces and/or objects. Image recognition can include two components: image detection **800** and image recognition **810**. Image detection can be utilized to determine if there is a pattern or patterns in an image that meet the criteria of a face, image, or text. If the result is positive, the detection processing then moves to recognition, i.e. image matching.

(’853, 13:39-46.) Zhao similarly discloses separate detection and recognition steps.

116. With respect to image detection, Zhao discloses **“an image detector**

configured to detect an image of an object in the still frame images”—such as processes to detect images of objects such as faces in extracted still frame images of a video. Referring to Figure 2, reproduced below, Zhao explains that “[o]nce the video has been retrieved, the video undergoes a process of detection of objects and extraction of object features from video images (processes 210, 220)” (Zhao, ¶0057):

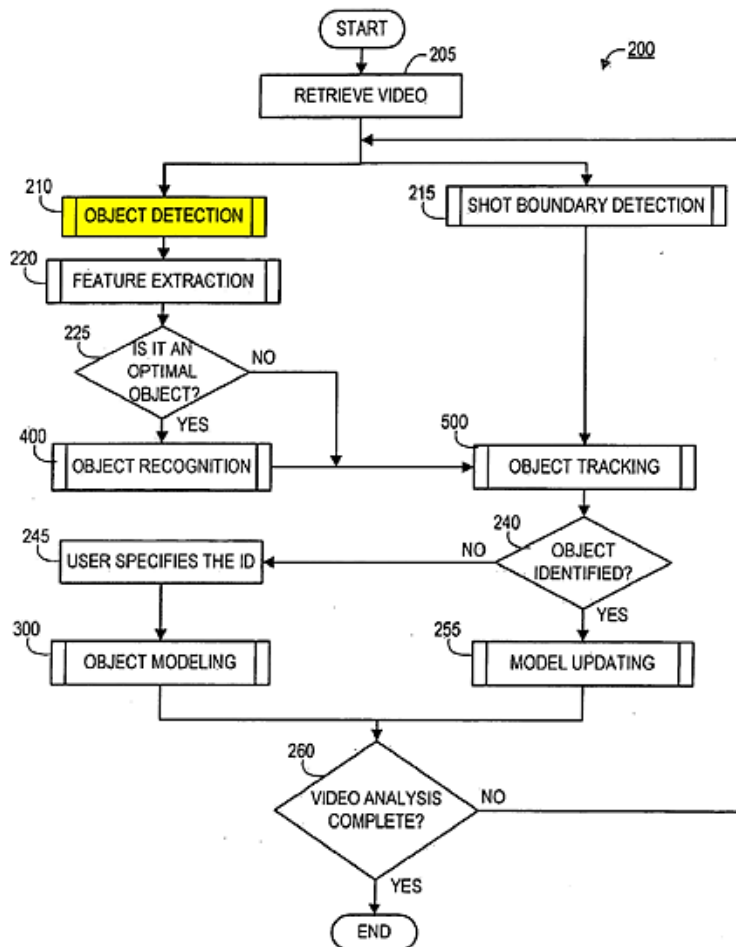


FIG. 2
OVERALL PROCESS

(Zhao, Fig. 2 (highlighting added).) Zhao describes in detail image detection process

210 to detect an object image from video images:

At process 210, object images are detected from video scenes using wavelet features of the images in combination with an Adaboost classifier.... As described herein a “feature” or “local feature” generally describes an element of significance within an image that enables recognition of an object within the image. A feature typically describes a specific structure in an image, ranging from simple structures such as points, edges, corners, and blobs, to more complex structures such as entire objects. For example, for facial recognition, features include eyes, noses, mouths, ears, chins, etc., of a face in an image, as well as the corners, curves, color, overall shape, etc., associated with each feature. Features are described in greater detail below and throughout this document. Generally, an AdaBoost (short for “Adaptive Boosting”) classifier refers to a machine-learning algorithm, and may be used in conjunction with other algorithms to improve overall performance. The AdaBoost classifier assists the system in learning to identify and detect certain types of objects, as described in greater detail below.

(Zhao, ¶0058.)

117. Claim 1[b] further recites “**wherein the image detector is adjustable to increase detection of non-primary images in the video.**” Zhao does not disclose image detection functionality that is “**adjustable to increase detection of non-primary images in the video,**” but this would have been obvious in view of

Steinberg.

118. Zhao does describe techniques for a system operator to adjust settings based on desired accuracy and/or precision of recognition and/or object tracking, and also acknowledges the issue of false facial recognitions produced by extras or crowds in a movie. For example, in connection with an object tracking/recognition process, Zhao explains how a system operator can adjust a similarity threshold:

At step **525**, each average confidence score or similarity measure is compared to a predetermined threshold. In one embodiment, the predetermined threshold is set by a system operator based on a desired accuracy and/or precision of recognition and/or tracking. For example, a higher threshold yields more accurate results because a similarity measure must exceed that threshold to be identified as matching the given model. Thus, if a detected object group and an identified object model have a high similarity value, there is therefore a higher percentage chance that the object group was correctly recognized. However, a higher threshold value also produces a lower recall, as some images or groups that were in fact correctly recognized are discarded or ignored because they do not meet a high similarity threshold (i.e. the group was correctly recognized, but had too many differences from the model based on its images' poses, resolutions, occlusions, etc.). Alternatively, a low threshold leads to higher recognition of images as matching a given model, but it is also likely to produce a higher percentage of false identifications. As will be understood, the threshold

value may be raised or lowered by a system operator depending on whether the operator would rather have higher accuracy, or more identifications.

(Zhao, ¶0089.) Zhao also teaches that whether a high or low similarity threshold is desirable depends on the video content—for example, “in a facial recognition embodiment, extras or crowds in a movie often produce false recognitions”:

In addition to the purposes and advantages of object tracking described above, a further benefit is that false positives of object recognitions are reduced because they likely do not fit into clearly established groups in a given video. For example, in a facial recognition embodiment, extras or crowds in a movie often produce false recognitions. However, because these recognitions are likely to be random and cannot be tracked over time (based on inconsistent or infrequent occurrences in a video), these false recognitions can be discarded as not belonging to a distinct group, and thus can be assumed to be false positives.

(Zhao, ¶0091.)

119. Like Zhao, **Steinberg** also relates to facial detection in digital images. Steinberg notes that “[t]he problem of face detection has not received a great deal of attention from researchers. Most conventional techniques concentrate on face recognition, assuming that a region of an image containing a single face has already been extracted and will be provided as an input.” (Steinberg, 1:49-53.)

120. With respect to claim 1[b], Steinberg describes a “face detection process” to find faces in a captured image. (Steinberg, 18:25-31 (“An image-detection process, preferably a face detection process, is applied to the captured image to seek faces in the image, **1020**.”); *see generally id.*, 18:29-44, 19:18-36.) Steinberg explains that the face detection process can be implemented “by changing the sensitivity or threshold of the detection process.” (Steinberg, 18:35-37.) Steinberg describes how “each of the faces is given a weight based on size of the faces, location within the frame, other parameters described herein, etc., which define the importance of the face in relation to other faces detected.” (Steinberg, 18:40-44.) Therefore, for example, “in response to false detection of one or more regions as one or more faces, or in response to a determination to concentrate on less image faces than faces identified in the identifying step,” “faces may be removed by increasing a sensitivity level of the face identifying step[.]” (Steinberg, 11:66-12:7; *see also id.*, 11:3-18 (“An automatically detected face may be removed in response to false detection of regions as faces, or in response to a determination to concentrate on less image faces or images faces that were manually determined to be of higher subjective significance, than faces identified in the identifying step. A face may be removed by increasing a sensitivity level of said face identifying step.”).) Although Steinberg focuses on increasing a sensitivity level to *decrease* detection of less

important (i.e., non-primary) faces, a person of ordinary skill in the art would have understood that decreasing the sensitivity level would have the opposite effect of *increasing* detection of non-primary faces in the image.

121. **Rationale and Motivation to Combine (Zhao and Kritt with Steinberg)**. It would have been obvious to a person of ordinary skill in the art to combine Steinberg’s teachings about an adjustable sensitivity threshold for detecting faces in images with Zhao’s video processing system (and under the proposed combination with Kritt). The combination would have involved the straightforward incorporation of an adjustable object detection setting as taught by Steinberg into Zhao’s processes for object detection. The combination would have predictably resulted in Zhao’s video processing system in which the facial detection sensitivity could be adjusted, for example, to focus only on more important or prominent faces, or to additionally be able to detect less prominent faces (such as faces in the background or in a crowd). The combination would have therefore resulted in “**an image detector configured to detect an image of an object in the still frame images, wherein the image detector is adjustable to increase detection of non-primary images in the video**” as claimed.

122. Steinberg is an analogous reference to the ’853 patent. Steinberg, like the ’853 patent, relates to object detection in images. Steinberg also would have

been reasonably pertinent to problems facing the '853 patent inventors, including techniques for adjusting sensitivity of object detection.

123. A person of ordinary skill in the art would have been motivated to combine Zhao (and under the proposed combination with Kritt) with Steinberg. Steinberg discloses that its techniques are applicable to video and provides motivation to incorporate an adjustable face detection step in Zhao, such as selective compression and sharpening. (Steinberg, 25:18-32 (“An image can be locally compressed so that specific regions will have a higher quality compression which equates to lower compression rate.... By doing so, selected or determined regions of importance may maintain low compression or high resolution compared with regions determined to have less importance or non-selected regions in the image.”).) Other benefits would have been apparent to a person of ordinary skill in the art such as processing time and efficiency improvements based on selectively increasing or decreasing face detection sensitivity based on needs—such as real-time processing—and subject matter—for example, decrease sensitivity for video involving a smaller group of individuals (thereby increasing detection of non-primary images) and increase sensitivity for video involving larger crowds such as sporting events. (Steinberg, 25:33-43 (“For example, such technology may be used for identification of faces in video sequences, particularly when the detection is to

be performed in real-time.”).)

124. A person of ordinary skill in the art would have had a reasonable expectation that the combination would have been successful. Because facial detection techniques were already well known, incorporating the ability to adjust the sensitivity of facial detection would have involved conventional programming techniques.

(c) **“an object recognizer configured to compare the image of the object to a fractal, wherein the fractal includes a representation of the object based on landmarks associated with the object, wherein the recognizer is further configured to update the fractal with the image.” (Claim 1[c])**

125. As noted for claim 1[b], the ’853 patent describes image “detection” and “recognition” as a two-step process. (’853, 13:39-46.) Zhao also describes a second, image “recognition,” step after performing image “detection.”

126. As explained below, claim 1[c] is obvious over Zhao in view of **Kouzani**. Zhao discloses processes to **“compare the image of the object”** to an object model that **“includes a representation of the object based on landmarks associated with the object,”** and to **“update”** an object model with the image. It would have been obvious, in view of Kouzani, for an object model to take the form of a **“fractal.”** Therefore, as explained below, the combination would have

predictably resulted in an “**object recognizer**”—corresponding to Zhao’s processes under the proposed combination—configured as claimed.

127. To start, Zhao discloses processes to “**compare the image of the object**” to an object model and “**update**” the object model with the image. As shown in Figure 2, reproduced below, Zhao discloses an object recognition process **400** and a model updating process **255**:

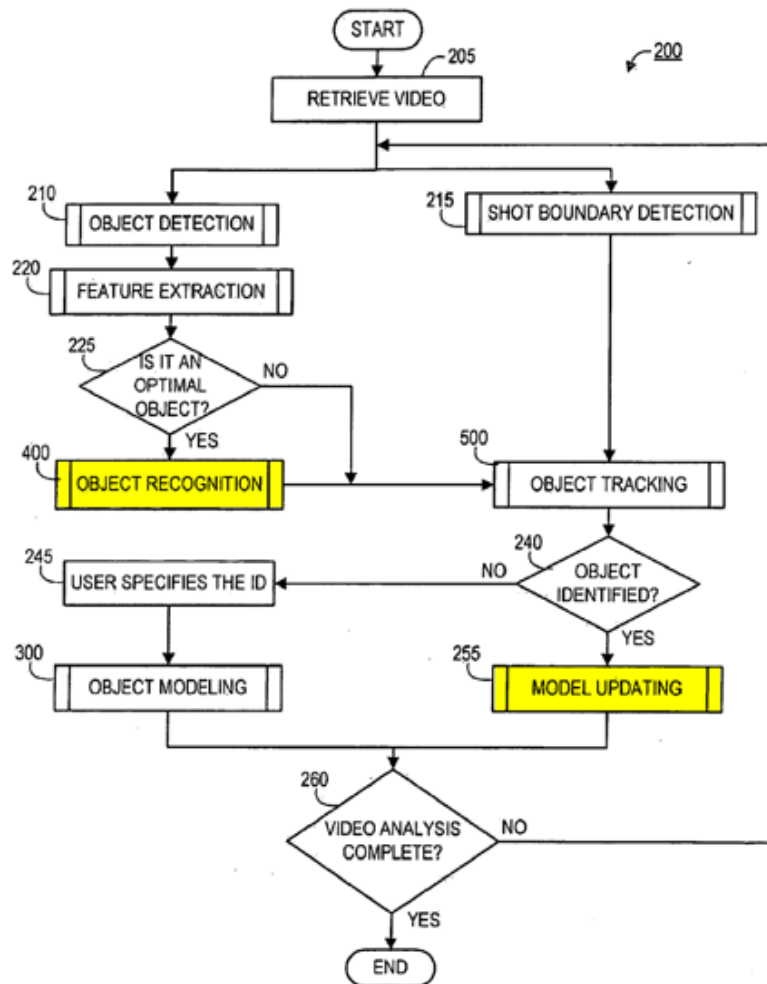


FIG. 2
 OVERALL PROCESS

(Zhao, Fig. 2 (highlighting added).) Referring to Figure 2, Zhao explains that “the object recognition process compares an extracted object image, or set of object images, to a database of predefined object models in order to ‘recognize’ the extracted image as one of the models,” where an object model “comprises a set of data, wherein the data is further comprised of features associated with an object image.” (Zhao, ¶¶0061, 0065.) “[F]or facial recognition, features include eyes,

noses, mouths, ears, chins, etc., of a face in an image, as well as the corners, curves, color, overall shape, etc., associated with each feature.” (Zhao, ¶0058.) “Depending on whether the image is recognized or not, either an existing object model is updated or a new object model is created.” (Zhao, ¶0061.) “For example, the recognized object image could include an actor or actress, a political figure, an animal, or some type of inanimate object.” (Zhao, ¶0063.) If the image is recognized, an object model associated with the recognized object—such as a particular person—is updated with the image by using extracted features of the image:

[T]he system calculates the similarity between the extracted features and those already existing in the model (process **410**), as described in greater detail below. For features that are determined to be similar to those already existing in the model, the relative weights of those features are updated, as it is assumed that these features are prominent in the object because they are reoccurring (step **325**). Features that are not similar to those already existing are added to the model as new features (step **325**), and thus the model becomes more robust as it develops additional features, that can be used to recognize object images in subsequent recognition procedures.

(Zhao, ¶0071.) “For example,” Zhao explains, “in a facial recognition embodiment, a given character’s or actor’s appearance often changes over time, due to weight gain, hair loss, aging, or other similar reasons.” (Zhao, ¶0073.) With Zhao’s system,

“each model is updated over time as new episodes or videos are released, and thus each model adapts and changes in response to changes in a given actor’s appearance.” (Zhao, ¶0073.)

128. Zhao further discloses that an object model **“includes a representation of the object based on landmarks associated with the object.”** As just noted, Zhao explains how an object model includes weighted features of the object, and in a facial recognition application, those features include facial **“landmarks”** such as “eyes, noses, mouths, ears, chins, etc., of a face in an image, as well as the corners, curves, color, overall shape, etc., associated with each feature.” (Zhao, ¶¶0058 (facial features), 0065 (object model “comprises a set of data, wherein the data is further comprised of features associated with an object image”), 0071 (object model updating).) Zhao’s example facial **“landmarks”** are just like those identified in the ’853 patent. (’853, e.g., 13:53-57 (“The system can utilize facial recognition algorithms to identify facial fractals by extracting landmarks from an image of the subject’s face. For example, the algorithm can analyze the relative position, size, and/or shape of the eyes, nose, cheekbones, and jaw.”).)

129. Zhao therefore discloses every aspect of claim 1[c], except possibly **“fractal”**:

an object recognizer configured to compare the image of the object

to a[n object model], wherein the [object model] includes a representation of the object based on landmarks associated with the object, wherein the recognizer is further configured to update the [object model] with the image.

(Claim 1[c] (replacing “fractal” with “object model”).)

130. Turning to “**fractal**,” the ’853 patent states, for example, that “[r]ecognition processing can include creating a fractal representation of the face or object that was detected” and that “[f]or detection and/or recognition requests, fractal datasets can be compared to a reference dataset and/or scored with a probability of a match based on comparing individual fractal values against one another.” (’853, 13:47-49, 19:13-16.)

131. Zhao does not expressly describe comparing an image of the object to a “**fractal**” or “**updat[ing] the fractal**” with the image. However, I have reviewed Patent Owner’s opposition to Petitioner’s motion to dismiss in the related district court litigation, and Patent Owner appears to argue that a “fractal” is nothing more than a shape or polygon, or potentially a shape or polygon used to identify objects or faces. (EX1017, p.007.) Under Patent Owner’s apparent interpretation, Zhao’s object model would disclose a “**fractal**,” since the object model comprises one or more shapes. (Zhao, e.g., ¶0058 (“[F]or facial recognition, features include eyes,

noses, mouths, ears, chins, etc., of a face in an image, as well as the corners, curves, color, overall shape, etc., associated with each feature.”).) And updating Zhao’s object model would accordingly disclose “**updat[ing] the fractal.**” (Zhao, ¶¶0071, 0073 (object model updating).) Zhao therefore discloses every aspect of claim 1[c] under Patent Owner’s apparent interpretation of “fractal.”

132. Putting aside Patent Owner’s apparent interpretation, it would have been obvious to compare an image of the object to a “**fractal**” or “**updat[ing] the fractal**” with the image. The use of fractals for facial recognition was already known before the ’853 patent. In fact, the ’853 patent identifies the use of fractals among a number of known recognition techniques, including Principal Component Analysis (PCA) and Elastic Bunch Graph Matching (EBGM), among others. (’853, 14:30-15:47 (“[T]he following recognition models can be utilized, alone or in combination, and are provided to illustrate exemplary alternative embodiments[.]”); *see also, e.g.*, Zhao, ¶0072 (Elastic Bunch Graph Matching); Kouzani, p.009 (Principal Component Analysis).) It was known, for instance, that “[a] human face is a surface lying in the fractal space intrinsically. Since many of the parameters are self-similar, the fractal model can better represent faces, especially in handling facial variations, such as pose, illumination, etc.” (’853, 15:40-44.) For example, a 2010 paper entitled “Facial Image Recognition Based on Fractal Image Encoding” (EX1014)

explains:

Fractal image encoding uses self-similarity to encode an image. Utilizing this feature, a new method of facial image recognition based on fractal image encoding is examined. Fractal image encoding can approximate any given image by capturing intrinsic self-similarity within the image. This method has some invariance to rotation, scaling, translation, and luminance.

(EX1014, p.005.) Even as of 2010, the use of fractals for facial recognition was not at all new—Kouzani (EX1006), for example, dates from 1997. (EX1006, p.007.)

133. It would have been obvious in view of Kouzani for Zhao’s “object model” for a face to take the form of a “**fractal**.” Kouzani, entitled “Fractal Face Representation and Recognition,” “presents a face representation and recognition scheme based on the theory of fractals.” (Kouzani, p.007.) Kouzani explains that “[a] fractal is a fragmented geometric shape that can be subdivided in parts, each of which is a reduced-size copy of the whole[,]” that “[f]ractals model many real-world objects that do not correspond to simple geometric shapes, e.g. clouds, mountains, turbulence, and coastlines [9],” and that “[f]ractals are generally self-similar and independent of scale.” (Kouzani, p.007.) In Kouzani’s system, “[e]ach face image is represented by its fractal model which is a small collection of transformation parameters.” (Kouzani, p.007.) “For recognition, the input face image is

transformed and its fractal model is then compared against the database of fractal models of known faces.” (Kouzani, p.007.) Kouzani explains a comparison process for recognizing a face performed with a neural network:

The second network has also a feedforward architecture and has been trained to recognise faces. It receives two group of inputs, the fractal model of an input image and the fractal model of a known face from the fractal database. It calculates the degree of similarity between the two fractal face models. If the output is below a threshold, the input face is considered as a known face whose fractal is currently under examination. Similar to the compression stage, more networks with the same architecture can be used to speed up the recognition process.

(Kouzani, p.009.)

134. Kouzani also touts advantages of using fractal models for facial recognition, including being less computationally expensive than prior art principal component analysis (PCA) techniques and providing high data compression:

In the recognition stage, two groups of experiments were performed. In the first experiment, the known face images were presented to the system and the recognition process was carried out. 100% recognition was achieved in this step. The maximum compression ratio obtained for successful recognition of all known faces was 89:1. Figure 6 demonstrates recognition of the known face number 100. In the second experiment, 50 unknown face images were directed to the system and

the recognition process was carried out. None of the input images were recognised by the system. Figure 7 shows the recognition result for an unknown input face image.

The proposed system is less computationally expensive than that for PCA. Moreover, when a new face is added to the database, recalculation of eigenfaces and re-compression of known faces are not necessary.

(Kouzani, p.009; *see also id.* (“Simulation results show that the proposed system overcomes the shortcomings of PCA.”).) Kouzani further notes that “real-time recognition” was achieved. (Kouzani, pp.008-009 (“The main reason for employing neural networks is to take advantage of the highly parallel architecture of neural networks in order to achieve real-time recognition.”).)

135. **Rationale and Motivation to Combine (Zhao, Kritt, and Steinberg with Kouzani)**. It would have been obvious to a person of ordinary skill in the art to combine Kouzani’s techniques for facial recognition using fractals with Zhao’s system (and under the proposed combination with Kritt and Steinberg). The combination would have involved the use of Kouzani’s known fractal-based facial recognition techniques to improve Zhao’s facial recognition system in a similar way (by for example, providing a less computationally expensive system). The combination would have predictably resulted in Zhao’s system for facial detection

and recognition in which recognition was performed by processes that compared a fractal of a detected facial image against a fractal model, and that updated the fractal model with the facial image. The combination would have therefore resulted in “**an object recognizer configured to compare the image of the object to a fractal, wherein the fractal includes a representation of the object based on landmarks associated with the object, wherein the recognizer is further configured to update the fractal with the image**” as claimed.

136. Kouzani is an analogous reference to the '853 patent. Kouzani is in the same field of image recognition as the '853 patent. And moreover, Kouzani would have been reasonably pertinent to problems facing the '853 patent inventors, including techniques for performing facial recognition.

137. A person of ordinary skill would have been motivated to combine. As discussed above, Kouzani touts the value of fractal-based facial recognition, including recognition results, real-time performance, high data compression, and being less computationally expensive compared to other techniques. (Kouzani, p.009.) A person of ordinary skill would have therefore been motivated to implement Kouzani's techniques in Zhao's system to obtain the same benefits and advantages. Moreover, as one of a number of known facial recognition techniques, the use of fractals would have been obvious to try.

138. A person of ordinary skill in the art would have had a reasonable expectation that the combination would be successful. As discussed above, the use of fractals for facial recognition and image processing generally was well known in the art. The combination would have involved the straightforward incorporation of known fractal coding techniques to Zhao's system.

2. Claim 2: "The system of claim 1, wherein the coordinator is configured to increase the plurality of demultiplexer nodes when a threshold of processing capacity is reached."

139. The additional limitations of claim 2 are disclosed by and obvious over Zhao in view of Kritt.

140. As discussed for claim 1, it would have been obvious to implement Zhao's video processing system in a cloud-based environment involving a number of nodes configured to perform demultiplexing operations.

141. It would have further been apparent and obvious to a person of ordinary skill in the art that Zhao's system, in a cloud-based implementation, would include "coordinator" functionality "to increase the plurality of demultiplexer nodes when a threshold of processing capacity is reached" as claimed. The ability to dynamically scale capacity based on workload was a key objective, and selling point, of cloud-based infrastructure. As Kritt explains, cloud computing "generally refers to the provision of scalable computing resources as a service over a network" and

“enable[es] convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.” (Kritt, ¶0078.) One example of a commercial service providing such functionality that would have been well-known to a person of ordinary skill in the art is Amazon’s AWS and Elastic Compute Cloud (EC2) service and its Auto Scaling feature. For example, the 2011 book, *Programming Amazon EC2*, by Jurg van Vliet and Flavia Paganelli explains that “[t]he cloud’s biggest promise is elasticity: running with just the amount of computing resources your workload demands at any given time.” (EX1013, p.0042.) “[F]or some applications, you need to be able to dynamically scale EC2 capacity, or instances. And that is what Auto Scaling does. It allows you to define conditions for increasing or decreasing the number of instances.” (EX1013, p.0042.) The authors further explain that “[a]ccording to AWS, ‘Auto Scaling is particularly well suited for applications that experience hourly, daily, or weekly variability in usage.’... At peak hours, we want to increase our capacity to handle the load, and when it’s not so busy, we want to run idle.” (EX1013, p.0042.)

142. Therefore, as discussed for claim 1, in an application of Zhao involving a large-scale system (such as a commercial video indexing and retrieval website or security/surveillance service), it would have been obvious to a person of ordinary

skill for the system to include functionality to scale computing resources based on whether a processing capacity threshold had been reached. At times when that threshold is reached, more computing resources—such as demultiplexer nodes—could be provisioned in order to continue to provide a responsive service. And at times with less demand, demultiplexer nodes could be released. Therefore, **“the coordinator is configured to increase the plurality of demultiplexer nodes when a threshold of processing capacity is reached”** as claimed would have been obvious.

3. **Claim 3: “The system of claim 1, wherein the recognizer is configured to determine distinguishing geometric features of the object.”**

143. The additional limitations of claim 3 are disclosed by and obvious over Zhao and Kouzani.

144. The **“recognizer”** recited in claim 3 refers back to the **“object recognizer configured to compare the image of the object to a fractal”** recited in claim 1. As discussed for claim 1[c], the claimed **“object recognizer”** is obvious over Zhao and Kouzani.

145. With respect to **“determining distinguishing geometric features of the object,”** claim 4 of the ’853 patent identifies some examples: “a contour of eye sockets, a nose, and a chin.” As discussed for claim 1[c], Zhao determines these

sorts of geometric features of a face in its recognition process, by both identifying them in the object image and comparing them against those features in an object model. (Zhao, e.g., ¶¶0071 (“[T]he system calculates the similarity between the extracted features and those already existing in the model (process **410**)”), 0058 (“[F]or facial recognition, features include eyes, noses, mouths, ears, chins, etc., of a face in an image, as well as the corners, curves, color, overall shape, etc., associated with each feature.”).) Under the proposed combination with Kouzani, these geometric features would have been embodied in fractal representations.

146. Therefore, the proposed combination of Zhao and Kouzani would have predictably resulted in the “**object recognizer**” “**configured to determine distinguishing geometric features of the object.**”

4. Claim 4: “The system of claim 3, wherein the distinguishing geometric features comprise a contour of eye sockets, a nose, and a chin.”

147. I already largely addressed claim 4 in my discussion of claim 3 above. As explained, Zhao discloses distinguishing geometric features including “**a contour of eye sockets**”—e.g., the overall shape associated with eyes—“**a nose, and a chin.**” (Zhao, ¶0058 (“For example, for facial recognition, features include eyes, noses, mouths, ears, chins, etc., of a face in an image, as well as the corners, curves, color, overall shape, etc., associated with each feature.”).)

C. Ground 2: Claims 5-7 Are Obvious Over Ground 1 Prior Art in Further View of Yang and Romdhani

5. Claim 5: “The system of claim 4, wherein the recognizer is configured to determine skin textures of the object.”

148. As explained below, the additional limitations of claim 5 are disclosed by and obvious in view of Yang and Romdhani.

149. Zhao mentions “skin color” as a feature used in facial/object detection and recognition, explaining that “[a]s the system is provided with additional examples of facial features, the general model learns what a typical face image looks like for many different kinds of faces, and thus develops the capability to take into account variations in these features, such as eye or skin color, overall size or scale of features, etc.” (Zhao, ¶0066.) Zhao provides only exemplary lists of facial features used for object detection and recognition, and does not describe determining or recognizing skin textures. (Zhao, ¶¶0058 (“for facial recognition, features include eyes, noses, mouths, ears, chins, etc., of a face in an image, as well as the corners, curves, color, overall shape, etc., associated with each”), 0066 (“such as a nose, mouth, eyes, ears, chin, and other similar features” for face detection).) But using skin textures for recognition (and detection) would have been obvious.

150. Kritt, for example, describes using “skin texture” for recognizing a facial object in a video image. (Kritt, ¶0044 (“[A] method may be of a type that

employs rules that identify so-called ‘invariant features’ that are present in a frame regardless of the position or pose of the object, the lighting, or camera viewpoint. Methods of this type, especially when employed to identify a human facial object, may employ an image recognition processes that identifies: (i) facial features using edge detectors (e.g., a Sobel filter) and templates; (ii) skin or hair texture using a neural network; and (iii) skin color using a pixel chrominance classifier.”.)

151. As another example, Steinberg discloses that known face detection techniques include identification of facial features, texture and skin color. (Steinberg, 1:65-2:5 (“methods of face detection which may be divided into four main categories: (i) knowledge-based methods; (ii) feature-invariant approaches, including the identification of facial features, texture and skin color; (iii) template matching methods, both fixed and deformable and (iv) appearance based methods, including eigenface techniques, statistical distribution based methods and neural network approaches.”).)

152. **Yang**, which Steinberg both discusses and expressly incorporates by reference (Steinberg, 1:62-65, 7:1-3; *see generally id.*, 6:40-7:6 (“Incorporation by Reference” section)), provides more detail about facial recognition involving “**skin textures.**” Yang, entitled “Detecting Faces in Images: A Survey,” reviews various techniques for detecting faces in images. Yang explains that “[h]uman faces have a

distinct texture that can be used to separate them from different objects.” (Yang, p.009.) Yang describes how “Augusteijn and Skufca developed a method that infers the presence of a face through the identification of face-like textures” and “[t]o infer the presence of a face from the texture labels, they suggest using votes of the occurrence of hair and skin textures.” (Yang, p.009.) Yang also describes another technique by Dai and Nakano that uses a “face texture model” that incorporates color information. “Using the face texture model, they design a scanning scheme for face detection in color scenes in which the orange-like parts including the face areas are enhanced.” (Yang, p.009.)

153. **Romdhani**, also cited in Steinberg (Steinberg, references cited, page 5), describes facial recognition using texture that produced “[s]tate of the art results[.]” (Romdhani, p.0039.) Entitled “Face Identification by Fitting a 3D Morphable Model using Linear Shape and Texture Error Functions,” Romdhani describes techniques for face recognition that involve matching face texture to an image of a known individual. (Romdhani, p.0024 (“Abstract”)) (“This paper presents a novel algorithm aiming at analysis and identification of faces viewed from different poses and illumination conditions. Face analysis from a single image is performed by recovering the shape and textures parameters of a 3D Morphable Model in an analysis-by-synthesis fashion.”).) Romdhani explains the face

recognition techniques in great detail, describing how “[t]he fitting procedure provides shape and texture coefficient which ... depends of [sic] the identity only. Identification of a novel face image is performed by matching the coefficients recovered from the image with the coefficients recovered from the image of a known individual.” (Romdhani, p.0036.) Romdhani explains that “[s]tate of the art results were obtained” (Romdhani, p.0039), outperforming “one of the most successful commercial face recognition system[s]” (Romdhani, p.0037).

154. **Rationale and Motivation to Combine (Ground 1 Prior Art with Yang and Romdhani)**. It would have been obvious to a person of ordinary skill in the art to combine Yang’s and Romdhani’s teachings regarding using skin and face textures for facial detection and recognition with Zhao’s video processing system (and under the proposed combination). The combination would have involved the incorporation of skin texture feature(s) as one of several features used for face recognition, and it would have predictably resulted in Zhao’s video processing system in which Zhao’s face recognition processes used skin texture as a feature when comparing a detected face image to an object model (or fractal, under the combination with Kouzani). The combination therefore discloses and renders obvious “**the [object] recognizer**” “**configured to determine skin textures of the object**” as claimed.

155. Both Yang and Romdhani are analogous references to the '853 patent. Like the '853 patent, Yang and Romdhani are also directed to the field of image processing and object detection/recognition, including facial detection and recognition. (Yang, e.g., p.004 (Abstract); Romdhani, p.0024 (Abstract) (“This paper presents a novel algorithm aiming at analysis and identification of faces viewed from different poses and illumination conditions.”).) Moreover, both Yang and Romdhani would have been reasonably pertinent to problems facing the '853 patent inventors, including techniques involving the use of skin and face textures to detect and recognize faces in video images.

156. A person of ordinary skill in the art would have been motivated to combine. For starters, as noted above, both Yang and Romdhani were identified in Steinberg. (*See* Steinberg, 1:62-65 (Yang), references cited, p.005 (Romdhani).) Steinberg specifically identifies Yang in connection with face detection techniques using “identification of facial features, texture and skin color” (Steinberg, 1:62-65) and moreover, expressly incorporates Yang (Steinberg, 7:1-3; *see generally id.*, 6:40-7:6 (“Incorporation by Reference” section)). I understand, therefore, that under the patent laws, Yang is effectively part of Steinberg as if Yang were explicitly contained in Steinberg. Based on at least their express identification by title (and incorporation, in Yang’s case), a person of ordinary skill in the art would have

appreciated that both references were directly relevant to techniques for face detection and recognition. Because, as discussed, Zhao provides only exemplary lists of facial features (which largely overlap) to use for detection and recognition, a person of ordinary skill in the art would have been motivated to investigate additional features that could provide improved recognition results. (Zhao, ¶¶0058 (“for facial recognition, features include eyes, noses, mouths, ears, chins, etc., of a face in an image, as well as the corners, curves, color, overall shape, etc., associated with each”), 0066 (“such as a nose, mouth, eyes, ears, chin, and other similar features” for face detection).)

157. In fact, Zhao teaches that, for face recognition, “more discriminative and detailed features than those used for general model detection [e.g., face detection] should be extracted from the given image.” (Zhao, ¶0069.) A person of ordinary skill in the art would have therefore been motivated to investigate additional “more discriminative and detailed features,” and would have appreciated that the incorporation of face textures, such as skin textures, would provide a further way to discriminate among potential recognition matches. At a minimum, a person of ordinary skill in the art would have appreciated that, as one of a number of identified features for facial detection and recognition, using skin textures would have at least been obvious to try.

158. Moreover, as noted Romdhani explains that “[s]tate of the art results were obtained” (Romdhani, p.0039), outperforming “one of the most successful commercial face recognition system[s]” (Romdhani, p.0037). A person of ordinary skill in the art would have therefore been motivated to incorporate the use of face textures, such as skin textures, to try to achieve similar state of the art results.

159. A person of ordinary skill in the art would have had a reasonable expectation that the combination would have been successful. The combination would have involved the straightforward incorporation of another face feature, skin textures, to use for facial recognition. As Yang and Romdhani both make clear systems making use of face textures, including skin textures, was known in the art.

6. Claim 6: “The system of claim 5, wherein the distinguishing geometric features or the skin textures are determined based on a three-dimensional model.”

160. Zhao does not describe determining distinguishing geometric features or skin textures “**based on a three-dimensional model**,” but this would have been obvious.

161. **Romdhani** discloses using a “**three-dimensional model**” for face recognition. Romdhani presents an algorithm, called “LiST” (for “Linear Shape and Texture,” that involves “fitting a 3D Morphable Model to face images.” (Romdhani, p.0039.) Romdhani explains that the algorithm “exploits the linearity of the model

in the recovery of shape and texture parameters leading to a robust and efficient algorithm” and “[s]tate of the art results were obtained” when “LiST was evaluated on the difficult problem of combined pose-illumination variation face identification.” (Romdhani, p.0039.) Romdhani notes that their face identification results outperformed “one of the most successful commercial face recognition system[s]” and that the results “clearly show that, for pose-invariant face identification, a 3D representation is valuable.” (Romdhani, p.0037.)

162. It would have been obvious to a person of ordinary skill in the art to combine Romdhani’s teachings regarding using a three-dimensional facial model for facial recognition with Zhao’s video processing system. The combination would have predictably resulted in Zhao’s object recognition processes using a three-dimensional model for facial recognition, and accordingly, in determining distinguishing geometric features—such as eyes, nose, and chin—or skin textures based on a three-dimensional model. The combination with Romdhani therefore discloses and renders obvious “**wherein the distinguishing geometric features or the skin textures are determined based on a three-dimensional model**” as claimed.

163. I discussed the rationale and motivation to further combine with Romdhani in my analysis of claim 5 above, which applies here as well. With respect

to claim 6, Romdhani provides express motivation to incorporate a three-dimensional model for facial recognition, explaining that state of the art results were obtained, and that “a 3D representation is valuable.” (Romdhani, pp.0037-0039.)

7. Claim 7: “The system of claim 6, wherein the three-dimensional model is a three-dimensional morphable model.”

164. As discussed for claim 6, Romdhani expressly discloses a “**three-dimensional model**” that is a “**three-dimensional morphable model**.” (Romdhani, e.g., pp.0024, (Title: “Face Identification by Fitting a 3D Morphable Model using Linear Shape and Texture Error Functions”), 0039 (“fitting a 3D Morphable Model to face images”); *see also id.*, 0025-0026 (describing “3D Morphable Face Model”).)

165. The rationale and motivation to further combine with Romdhani was discussed above in connection with claims 5 and 6, which applies here as well.

D. Ground 3: Claims 8-9 Are Obvious Over Ground 1 Prior Art in Further View of Trivedi

8. Claim 8: “The system of claim 1, further comprising a processor and a camera for capturing the video.”

166. Claim 8 is obvious in view of Trivedi.

167. Zhao discloses “**a processor**.” (Zhao, e.g., ¶¶0029 (“The system comprises a processor and a computer program product...”), 0055 (“[E]mbodiments of the present system may be operated in a computer environment including

databases and other storage apparatuses, servers, processors, terminals and displays, computer-readable media, algorithms, and other computer-related components.”.)

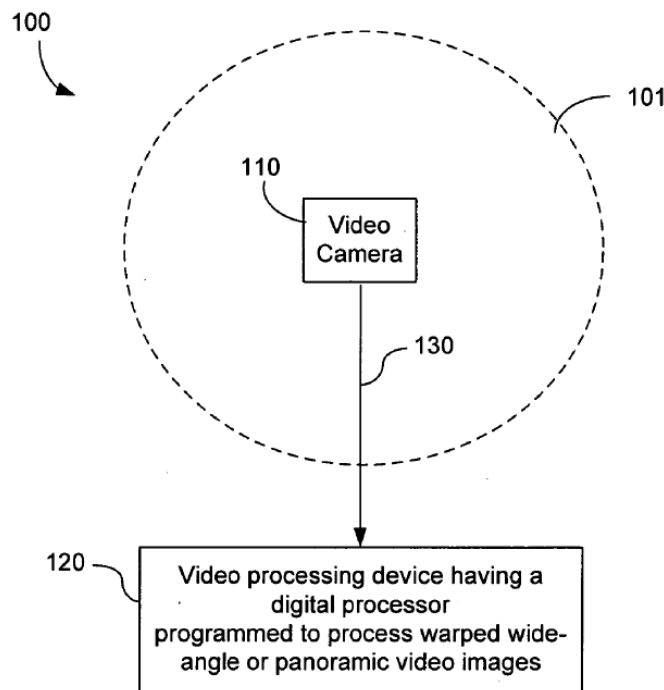
168. With respect to “**a camera for capturing the video**,” Zhao also teaches using its techniques in systems that a person of ordinary skill in the art would have understood and found obvious would include a camera, such as video surveillance systems. For example, Zhao explains that “[e]mbodiments of the present system are useful for a wide variety of applications, including video indexing and retrieval, video surveillance and security, unknown person identification, advertising, and many other fields.” (Zhao, ¶0053; *see also id.* ¶¶0003 (“security and surveillance”), 0007 (“security systems can benefit from accurate and consistent identification of perpetrators or victims within surveillance videos”).)

169. To the extent there is any question that Zhao alone discloses and renders obvious “**a camera for capturing the video**,” **Trivedi** confirms that video cameras are used in security/surveillance systems and expressly discloses a security system that includes a camera. Trivedi, entitled “Digital Processing of Video Images,” explains:

Video cameras have been used to capture video images for various applications such as surveillance, machine vision, security monitoring, inspection, sensing and detection. In these and other applications, the

captured video images may be sent to a nearby or remote image processing device to be processed and displayed. Digital image processing techniques may be used to process the video images to extract information from the video images.

(Trivedi, ¶0003.) As shown in Figure 1, reproduced below, Trivedi discloses a video system **100** that includes a video camera **110** and video processing device **120**:



(Trivedi, Fig. 1; *see also id.*, ¶0040 (“FIG. 1 illustrates one example of a video system **100** where a video camera **110** is installed at a location **101** to capture a video of the location **101**....A video processing device **120** is located at a different location and is linked to the video camera **110** by a communication link **130** to receive the video signal.”).) I understand that the Examiner cited Trivedi during prosecution in

connection with claims that ultimately issued as claims 8 and 9, and was not distinguished by the applicant. (EX1015 ('853 file wrapper), pp.00145 (“Trivedi further teaches further comprising a processor and a camera for capturing the video”; “Trivedi further teaches wherein the processor is configured to transform the video into rectilinear format”), 00167-00168 (Applicant Reply to Office Action of October 29, 2020).)

170. **Rationale and Motivation to Combine (Ground 1 Prior Art with Trivedi)**. It would have been obvious to a person of ordinary skill in the art to combine Trivedi’s teachings regarding a camera for capturing video. The combination would have involved the straightforward incorporation of a video camera into Zhao’s system (to the extent a camera was not already part of Zhao’s system), and would have predictably resulted in Zhao’s system having a video camera for capturing video. The combination therefore discloses a system having **“a camera for capturing the video”** as claimed.

171. Trivedi is an analogous reference to the '853 patent. Trivedi similarly describing a video processing system that involves “real-time tracking of persons and objects” and “face recognition of persons.” (Trivedi, e.g., ¶0005.) Moreover, Trivedi would have been reasonably pertinent to problems facing the '853 patent inventors, including techniques for transforming video images into a rectilinear

format.

172. A person of ordinary skill in the art would have been motivated to further combine with Trivedi. As discussed, Zhao teaches that its system has a wide range of applications, including video surveillance and security. (Zhao, ¶0053; *see also id.* ¶¶0003, 0007.) Therefore, for starters, Trivedi simply provides express confirmation of what was well-known to laypersons and persons of ordinary skill alike—video surveillance and security systems include one or more video cameras for capturing video. Moreover, a person of ordinary skill in the art, in view of Zhao’s teachings, a person of ordinary skill in the art when implementing Zhao’s system would have looked to a reference such as Trivedi for techniques for capturing video—including by incorporating a video camera.

173. A person of ordinary skill in the art would have had a reasonable expectation that the combination would have been successful, as it would have involved the straightforward incorporation of a video camera into Zhao’s video processing system.

9. Claim 9: “The system of claim 8, wherein the processor is configured to transform the video into a rectilinear format.”

174. Claim 9 is obvious in view of Trivedi.

175. Trivedi describes capturing video in a warped, non-rectilinear format—

for example, video captured with a wide angle lens—and “**transforming the video to a rectilinear format.**” Trivedi explains:

Certain video cameras are designed with optical imaging modules to capture wide-angle or panoramic video images. Such video images are distorted due to the designs of various optical imaging modules. Digital imaging processing may be used to transform the distorted images into images that a viewer would normally perceive when directly look at the scene being captured.

(Trivedi, ¶0004.) Trivedi’s system includes digital processing techniques to “efficiently transform warped video images into rectilinear video images.” (Trivedi, ¶0005.) Trivedi explains that the transform of video into a rectilinear format is performed by the digital processor of the system’s video processing device. (Trivedi, ¶0042 (“[T]he digital processor may be programmed, among other features and functions, to transform the warped video images into rectilinear video images[.]”); *see also id.*, ¶0041 (“The video processing device **120** includes a digital processor that is programmed to process the warped wide-angle or panoramic video images to extract desired information.”), Fig. 2 (showing device **120** “Video processing device having a digital processor programmed to process warped wide-angle or panoramic video images”).)

176. As noted in my analysis of claim 8, I understand that the Examiner cited

Trivedi during prosecution in connection with claims that ultimately issued as claims 8 and 9, and was not distinguished by the applicant. (EX1015 ('853 file wrapper), pp.00145 (“Trivedi further teaches further comprising a processor and a camera for capturing the video”; “Trivedi further teaches wherein the processor is configured to transform the video into rectilinear format”), 00167-168 (Applicant Reply to Office Action of October 29, 2020).)

177. It would have been obvious to further incorporate Trivedi’s teachings regarding transforming the video, such as video captured by a wide-angle camera of a surveillance system, into a rectilinear format. The combination would have predictably resulted in Zhao’s video processing system having a processor **“configured transform the video into a rectilinear format”** as claimed.

178. I discussed the rationale and motivation to combine Ground 1 Prior Art with Trivedi in my analysis of claim 8 above, which applies here as well. With respect to claim 9, incorporating Trivedi’s teachings regarding transforming the video, such as video captured by a wide-angle camera of a surveillance system, into a rectilinear format would serve two purposes. First, it would advantageously enable a viewer of the video to see the scene in an undistorted way that the viewer would normally perceive when directly looking at the scene. (Trivedi, ¶0004 (“Digital imaging processing may be used to transform the distorted images into images that

a viewer would normally perceive when directly look at the scene being captured.”.)
Second, like Zhao, Trivedi describes using its system for face recognition. (Trivedi, e.g., ¶¶0005, 0092-0094.) A person of ordinary skill in the art would have readily appreciated that, in order to best perform face recognition, video should be unwarped into a rectilinear format to better recognize distinguishing facial features. Trivedi explains, for example, how the face video stream used for face recognition “may be obtained from the unwarped rectilinear video stream produced by the DTV module of the digital processor.” (Trivedi, ¶0095.) Moreover, because video input into Trivedi’s system for face recognition could comprise both rectilinear and non-rectilinear formatted video (Trivedi, ¶0095), a person of ordinary skill in the art would have also understood that it would be desirable for all video input to be converted to a common, rectilinear format to help ensure better, more consistent face recognition.

E. Ground 4: Claims 10-11 Are Obvious Over Ground 1 Prior Art in Further View of Singer

- 10. Claim 10: “The system of claim 1, wherein the coordinator is configured to embed metadata about the object into the video, and wherein the metadata comprises a timestamp and a coordinate location of the object in the still frame images.”**

179. Zhao describes various uses for “**metadata about the object**” generated as a result of its image processing techniques, including indexing and

retrieval in order to locate and retrieve videos or portions of videos with the associated recognized images. For example:

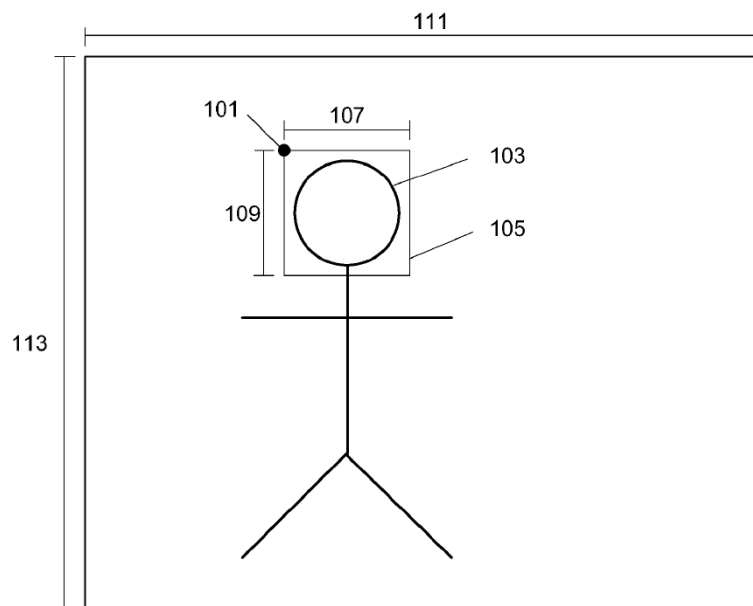
As mentioned previously, one application of embodiments of the present system is video indexing and retrieval. In order to accomplish such indexing and retrieval, in one embodiment, images in videos that have been identified and labeled as known objects (i.e. corresponding to model) are encoded with metadata attributes identifying those images. These encoded metadata attributes are generally stored in index files and are configured to be searched in order to locate and retrieve videos or portions of videos with the associated images. The system automatically encodes the video images based on identifiers associated with the corresponding object models. Accordingly, vast amounts of videos may be embedded with metadata and indexed according to object images in the videos by implementing embodiments of the present system.

(Zhao, ¶0092; *see also id.*, ¶¶0053 (“Embodiments of the present system are useful for a wide variety of applications”), 0007 (“The ability to effectively and efficiently index, store, and retrieve videos, or portions of videos, based on objects in those videos is important for a variety of fields.”).)

180. As shown in the passage above, Zhao describes “metadata attributes identifying those images” that can be “searched in order to locate and retrieve videos or portions of videos with the associated images.” (Zhao, ¶0092.) Zhao does not

further detail what exactly the “metadata attributes” are, but because they are used to locate portions of videos with the associated object images, a person of ordinary skill in the art would have understood and found it obvious that the metadata attributes would “**comprise[] a timestamp and a coordinate location of the object in the still frame images.**” It would have also been obvious to “**embed metadata about the object into the video.**” As explained below, both of these aspects of claim 10 are disclosed in Singer.

181. Singer, entitled “Object Detection Metadata,” describes techniques for generating and storing metadata about objects detected in video image data, including face detection metadata. (Singer, ¶¶0023-0024.) For example, Figure 1 of Singer illustrates detection of a face in an image of a video:



(Singer, Fig. 1.) Singer explains that “the location of each face in a frame of image data may be described with a simple rectangle.” (Singer, ¶0025.) For example, in Figure 1, “[f]ace detection rectangle **105**,” which bounds face **103**, “has dimensions of height **109** and width **107** and an initial point **101**.” (Singer, ¶0026.) Singer in figure 9 illustrates an example of metadata for a face detection rectangle:

Detected-face entry 903	
X =	15
Y =	20
Width =	32
Height =	26
ID =	3
Layer =	2
DetectWeight =	243
MatchWeight =	173

(Singer, Fig. 9.) Singer explains that “Detected-object entry **903** provides one example of a detected-face entry in a metadata track of image data.” (Singer, ¶0053.) The X and Y values, and Width and Height values, specify the location and dimensions, respectively, of the detected face in the image. (Singer, ¶0053.)

182. I note that Singer’s depiction and description of detection rectangles matches Zhao’s depiction of using rectangles to identify detected objects in a still frame image of a video:

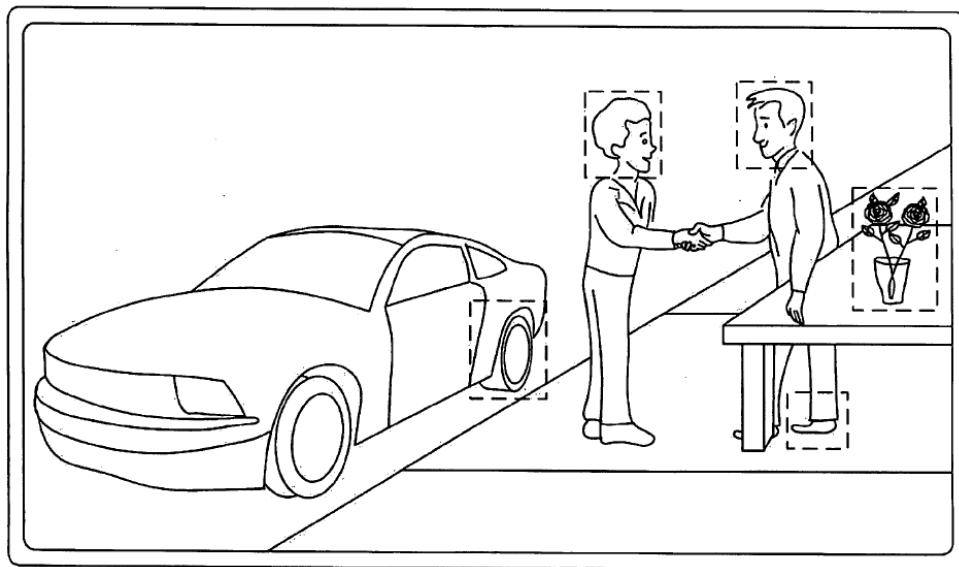


FIG. 1

(Zhao, Fig. 1, ¶0003 (“For example, FIG. 1 illustrates a sample frame of a video scene. Exemplary objects that are capable of being recognized within the illustrated video include characters’ faces, a plant in a vase, a shoe, and an automobile tire, each of which is shown within a dashed box to indicate its detection and recognition within the frame.”).) This suggests that Zhao collects similar coordinate location metadata for detected (and recognized) objects.

183. Singer further explains that the object detection metadata may be stored “as timed metadata associated with a video track[,]” such as a track of a video in QuickTime File Format. (Singer, ¶0023.) “In time-based metadata, each detected-object entry applies to one or more temporally aligned frames of video data.”

(Singer, ¶0056.) Figure 13 of Singer, reproduced below, illustrates an example of temporally-aligned, timed metadata for a detected face embedded in a video:

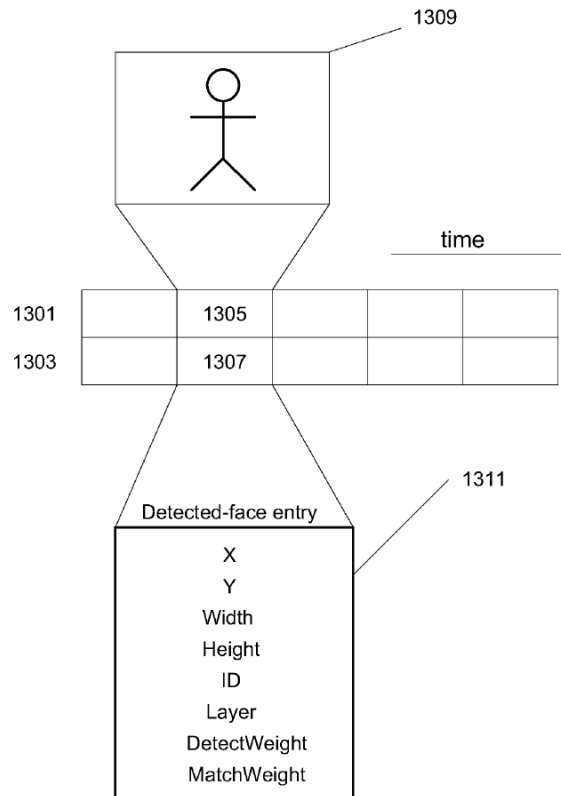


Fig. 13

(Singer, Fig. 13.) Referring to figure 13, Singer explains that “[t]rack **1301** includes frames of image data over time which may be replayed as a video” and “[t]rack **1303** includes detected object metadata entries.” (Singer, ¶0068.) “Frame **1305** in track **1301** corresponds to image frame **1309**[.]” while “[m]etadata entry **1307** corresponds to detected face entry **1311**[.]” (Singer, ¶0068.) Detected face entry

1311 “includes metadata describing a detected object in frame **1305**, such as the face of the person captured in image data **1309**.” (Singer, ¶0068.)

184. Singer therefore discloses both “**a coordinate location of the object in the still frame images**”—i.e., X and Y values in a metadata entry for a detected object—as well as “**a timestamp**”—i.e., information that “temporally aligns” the metadata entry with the associated video frame. Thus, Singer discloses “**embed[ding] metadata about [an] object into [a] video, and wherein the metadata comprises a timestamp and a coordinate location of the object in the still frame images.**”

185. I understand that the Examiner cited Singer during prosecution in connection with claims that ultimately issued as claims 10 and 11, and was not distinguished by the applicant. (EX1015 ('853 file wrapper), pp.00147-00148 (“Singer ... teaches wherein the coordinator is configured to embed metadata about the object into the video, and wherein the metadata comprises a timestamp and a coordinate location of the object in the still frame images”; “Singer further teaches wherein the coordinator is configured to generate a metadata stream corresponding to the image, wherein the metadata stream includes one or more timestamps corresponding to the image, and wherein the coordinator is configured to embed the metadata stream in the video”), 00167-00169 (Applicant Reply to Office Action of

October 29, 2020).)

186. **Rationale and Motivation to Combine (Ground 1 Prior Art with Singer)**. It would have been obvious to a person of ordinary skill in the art to combine Singer’s teachings regarding embedding metadata about detected objects in video, including “**coordinate location**” and “**timestamp**,” with Zhao’s video processing system (and under the proposed combination). The combination would have involved the straightforward incorporation of Singer’s teachings for embedding metadata, and would have predictably resulted in Zhao’s video processing system in which detected metadata was embedded into videos processed by Zhao’s system. In the proposed combination, the functionality to embed metadata would form part of the “**coordinator**” functionality—for example, as a function that received metadata from demultiplexer nodes and embedded that metadata into the retrieved video file. This is consistent with Zhao’s description of storing any “results or outputs” after performing video analysis. (Zhao, ¶0055.) The combination therefore discloses and renders obvious claim 10.

187. Singer is an analogous reference to the ’853 patent. Like the ’853 patent, Singer relates to video processing and object detection in video images, including face detection. In addition, Singer would have been reasonably pertinent to problems facing the ’853 patent inventors, including how to store metadata

associated with processed video.

188. A person of ordinary skill in the art would have been motivated to combine. A person of ordinary skill would have appreciated the embedding metadata in a video, as opposed to only storing metadata separate from the video, would beneficially enable the video to be transferred and shared without losing the associated metadata. The recipient of the video could therefore use the metadata without having to perform the recipient's own processing on the video. This sharing of videos is suggested in Zhao, which notes "online video viewing" (Zhao, ¶0007), and discussed in Singer, which describes both streaming and downloading a video with the embedded metadata:

At block 607, the method transmits the image data, including the added metadata track. For example, the image data may be streamed to a remote client. Alternatively, the image data may be downloaded to a remote client and then played locally on the remote client.

(Singer, ¶0040.) Zhao states how "encoded metadata attributes are generally stored in index files" (Zhao, ¶0092), but a person of ordinary skill in the art would have recognized that this would not diminish the motivation to embed metadata into the video, as doing so serves different purposes. In fact, Singer itself states that the metadata is both embedded into video as well as stored in a separate search index.

(Singer, ¶0065 (“At block **1205**, the method stores the object detection metadata and the object identification metadata in a searchable index. The searchable index may be an index database.”).)

189. A person of ordinary skill in the art would have had a reasonable expectation of success. Zhao already discloses object detection and metadata creation, and Singer acknowledges the existence of commercially-available video formats (such as Apple Inc.’s QuickTime File Format) capable of storing embedded metadata. (Singer, ¶0023.) The combination therefore would have involved at most the routine application of conventional programming techniques and technologies.

11. Claim 11: “The system of claim 1, wherein the coordinator is configured to generate a metadata stream corresponding to the image, wherein the metadata stream includes one or more timestamps corresponding to the image, and wherein the coordinator is configured to embed the metadata stream in the video.”

190. I largely addressed the limitations of claim 11 in my analysis of claim 10 above. Claim 11 differs from claim 10 mostly in its recitation of a “**metadata stream**” that is embedded in the video. The metadata stream, like the metadata in claim 10, includes one or more “**timestamps.**”

191. Claim 11 would also have been obvious in further view of Singer. Singer discloses embedding metadata into a video that is “streamed to a remote

client.” (Singer, ¶0040.) Because the video itself is streamed, a person of ordinary skill in the art would have understood that the embedded metadata track would similarly comprise a “**metadata stream**.”

192. As noted for claim 10, I understand that the Examiner cited Singer during prosecution in connection with claims that ultimately issued as claims 10 and 11, and was not distinguished by the applicant. (EX1015 (’853 file wrapper), pp.00147-00148 (“Singer ... teaches wherein the coordinator is configured to embed metadata about the object into the video, and wherein the metadata comprises a timestamp and a coordinate location of the object in the still frame images”; “Singer further teaches wherein the coordinator is configured to generate a metadata stream corresponding to the image, wherein the metadata stream includes one or more timestamps corresponding to the image, and wherein the coordinator is configured to embed the metadata stream in the video”), 00167-168 (Applicant Reply to Office Action of October 29, 2020).)

193. In the proposed combination, just as with claim 10, the functionality to generate a metadata stream and embed the metadata stream would form part of the “**coordinator**” functionality—for example, as a function that received metadata from demultiplexer nodes and embedded that metadata into the retrieved video file for streaming. This is consistent with Zhao’s description of storing any “results or

outputs” after performing video analysis. (Zhao, ¶0055.)

194. The combination with Singer would have predictably resulted in “**the coordinator is configured to generate a metadata stream corresponding to the image, wherein the metadata stream includes one or more timestamps corresponding to the image, and wherein the coordinator is configured to embed the metadata stream in the video**” as claimed.

V. NO SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS

195. As explained in **Part III.B**, I understand that “secondary considerations” may be used as indicia of nonobviousness, provided that any such consideration has a nexus to the invention claimed. I understand from Petitioner’s counsel that Patent Owner in the related district court litigation has not yet identified any evidence related to secondary considerations with a nexus to the claimed invention. Nor am I aware of information, such as commercial success, unexplained results, long felt but unsolved need, industry acclaim, simultaneous invention, copying by others, skepticism by experts in the field, and failure of others, suggesting that the claims addressed in this Declaration are not obvious. To the extent Patent Owner later provides information it claims relates to secondary considerations, I reserve the right to supplement my analysis and opinions to comment on it.

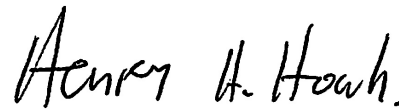
VI. CONCLUSION

196. In signing this Declaration, I recognize that the Declaration will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I also recognize that I may be subject to cross-examination in this proceeding. If required, I will appear for cross-examination at the appropriate time. I reserve the right to offer opinions relevant to the invalidity of the challenged claims at issue and/or offer testimony in support of this Declaration.

197. I hereby declare that all statements made herein of my own knowledge are true and that all statements are made on information and belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. § 1001.

Dated: May 9, 2025

Respectfully submitted,

A handwritten signature in black ink that reads "Henry H. Houh." The signature is written in a cursive, slightly slanted style.

Henry Houh, Ph.D.
Lexington, Massachusetts

APPENDIX A

Henry H. Houh

15 Ballard Ter.
Lexington, MA 02420
henry.houh@gmail.com
cell: 781-325-8273

Education

Massachusetts Institute of Technology, Cambridge, MA

- PhD in Electrical Engineering and Computer Science, February 1998. “Designing Networks for Tomorrow's Traffic,” thesis supervised by Professor David Tennenhouse and Professor John Guttag. GPA 4.7/5.0
- Master of Science in Electrical Engineering and Computer Science, February 1991. “Demonstration of a laser repetition rate multiplier,” thesis. GPA 4.5/5.0
- Bachelor of Science in Electrical Engineering and Computer Science, June 1989. “Boundary element analysis of arbitrarily shaped dielectric structures,” thesis. GPA 4.7/5.0
- Bachelor of Science in Physics, February 1990. GPA 4.7/5.0

MITx, Cambridge, MA

- September 2020 – present: Enrolled in Statistics and Data Science MicroMasters program, a sequence of five MIT-offered graduate-level courses in statistics and machine learning, culminating in a MicroMasters certificate. Have completed four of five required courses.

Experience

Houh Consulting Inc. / Independent Consultant

- 2009 - present: Technical consultant specializing in Social Networking, Web 2.0, Web Site Development, Data Networking, Optical Networking, Telecommunications, Media Streaming and Voice Over IP. Clients include: Akin Gump Strauss Hauer & Feld LLP, Arnold & Porter Kaye Scholer LLP, BBN, Baker Botts LLP, Barnes & Thornburg LLP, BBN Technologies, Beus Gilbert McGroder PLLC, Bracewell LLP, Christensen O'Connor Johnson Kindness PLLC, Cooley LLP, Covington & Burling LLP, Cozen O'Connor P.C., DLA Piper, Davis Polk & Wardwell LLP, Desmarais LLP, Duane Morris LLP, Erise IP, P.A., Faegre Baker Daniels LLP, Finnegan, Henderson, Farabow, Garrett & Dunner, LLP, Fish & Richardson P.C., Goldman Ismail Tomaselli Brennan & Baum LLP, Haynes and Boone, LLP, Kaye Scholer, Kellogg Huber Hanson Todd Evans & Figel PLLC, Kirkland & Ellis LLP, Klarquist Sparkman, LLP, McCarthy Tetrault LLP, McGuireWoods LLP, Morgan, Lewis & Bockius LLP, Munger Tolles & Olson LLP, Orrick, Herrington & Sutcliffe LLP, O'Melveny & Myers LLP, Orrick, Herrington & Sutcliffe LLP, Paul Hastings LLP, Paul, Weiss, Rifkind, Wharton & Garrison LLP, Perkins Coie LLP, Quinn Emanuel Urquhart & Sullivan, LLP, Sidley Austin LLP, Smith Baluch LLP, Sterne, Kessler, Goldstein & Fox PLLC, Troutman Sanders LLP, Venable LLP, Weil, Gotshal &

Manges LLP, Wiley Rein LLP, Wilmer Cutler Pickering Hale and Dorr LLP, and Winston & Strawn LLP.

BlocksCAD Inc. (H3XL spin-out)

- 2014 - present: Founder and member of Board of Directors. BlocksCAD is a visual Computer-Aided Design (CAD) program, designed to teach children mathematics and geometric visualization. BlocksCAD is being used in U.S. schools and worldwide in maker spaces, with 350,000+ users. BlocksCAD was selected to be part of the LearnLaunch 2017 Boost Accelerator, the MIT Play Labs 2018 Summer Accelerator, and the MassChallenge 2018 Boston Cohort. Recipient of DARPA DSO grant in 2014, and SBIR grant from the USDA in 2019.

H3XL Inc. d/b/a Einstein's Workshop (formerly Lexington Robotics)

- 2009 - 2022: Founder and President. Started a science, technology, engineering and math enrichment program and creative/maker space, in 7,000 square feet of space. Serve 2,000+ kids and families annually. As of end of 2022, have delivered an estimated 250,000 student-instruction-hours of STEM courses. Principle Investigator for 2-year DARPA grant to improve 3D Computer Aided Design tool developed in-house for the purposes of teaching 8 year olds and up 3D CAD.

Eons

- 2008 - 2009: Chief Technology Officer. Created product that Eons acquired from BBN Technologies. Integrated BBN product with Eons social networking platform and significantly increased the Eons group creation rate. Helped evaluate advertisement platform offerings and rolled out the "Boom Network" advertising network. Eons raised \$32 million from General Catalyst Partners, Charles River Ventures, Sequoia Capital, and Intel Capital.

BBN Technologies

- 2007 - 2008: Delta Division, Vice President of Technology. Grew "Boomerang" counter-sniper project engineering team and significantly de-risked \$10 million worth of product deliveries. Identified sales team leading to \$10 million dollars of new sales within 6 months and \$100+ million in additional orders in the following two years. Boomerang was a significant asset leading to the acquisition of BBN by Raytheon in 2009. Created new business plan and grew team; launched new fully-featured social networking web site in 5 months. Served as lead expert witness in patent infringement lawsuit, resulting in \$58 million jury award to client; verdict for patents I testified on were upheld on appeal which resulted in a \$120 million settlement.
- 2004 - 2007: Delta Division, Director of Technology, responsible for commercializing IP and creating new businesses. Hired and grew division's initial engineering team. Wrote three business plans, two of which were funded. For call center business plan, acted as general manager, hiring and managing engineering team, inside sales team, and identifying and recruiting a new general manager. Identified and recruited other key employees to Delta Division, including senior members of team leading to successful internal sales growth and spin-outs of projects. Contracted by BBN to BBN spin-out PodZinger as VP of Operations and Technology. Additional areas studied include mesh networking, network security, satellite communications, and semantic web.

PodZinger Inc. (BBN spin-out, also known as EveryZing and now RAMP)

- 2006: Vice President of Operations and Technology. Significantly upgraded capability of consumer-facing search site and redeployed web site from company to co-location facility. Identified key portions of infrastructure for upgrading and cost reduction. Helped write business plan, evaluating advertisement models of revenue. Hired in operations replacement and phased back to BBN.

Commonwealth Capital

- 2004: Entrepreneur-in-residence (informally), performed technical due diligence on business plans, brainstormed ideas for new businesses with venture partner. With venture partner, left for portfolio company BBN to form core of commercialization team.

Empirix, Inc./Teradyne, Inc.

- 2001 - 2004: Chief Technologist, Engineering Manager, Web Application Test Group. Researched potential new product areas; developed product plan and prototype. Responsible for three new and existing products. Managed off-shore development team. Chief architect for all web testing products. Re-architected core testing product, helped write javascript interpreter. Provided technical vision for core product.
- 2000 - 2001: Chief Technologist, Communication Infrastructure Test Group. Responsible for incorporating new technology internally, tracking new technologies, technical evaluation of partnerships and potential acquisitions. Helped develop division strategy. Developed plans which formed core capabilities for successful new products introduced in 2004-5.
- 2000 - 2001: Engineering Manager, Communications Infrastructure Test Group. Execution of new product plan developed in prior role. Grew team from four existing engineers to team of over 30 on immediate team and over 40 on project. Delivered new platform in one year. Platform and derivatives accounted for large portion of booked products for the division within 2 years and is currently (2008) a key portion of new product offerings.
- 2000: Empirix was formed as a spin-out of Teradyne in January 2000. Reported to CEO in carve-out of Empirix from Teradyne.
- 1999: (Teradyne) Director of Business Development, Software Test Units. Reported directly to Chairman of the Board/Founder and then to general manager of software test unit (6 divisions of Teradyne). Evaluated and researched acquisition and partnership candidates. Internally assessed technology position in market and gaps in product lines. Worked with senior division staff to develop new product strategies. Attended internal Teradyne divisional board meetings. Chairman served as my mentor.

3Com Corporation/NBX Corporation

- 1999: Software Engineer 5. Continued work after 3Com acquisition of NBX. Built cross-division relationships for new products and research directions. NBX was acquired by 3Com in March 1999.
- 1997 - 1999: (NBX) Senior Scientist and Engineer. Work in IP Telephony. Architected next-generation product. Protocol design and validation for core protocol now used tens of millions of times daily. Led team in integration of IP protocols into current product. Designed audio reconstruction algorithms. Developed applications for bug analysis and diagnosis of system

problems. Implementation of network simulator. Work on collaborative projects with external partners. Worked to identify gaps in product. Representative at numerous trade shows. Innovated novel methods of using product.

MIT Department of Electrical Engineering and Computer Science, Cambridge, MA

- 1991 - 1998: Research assistant, Telemedia Network Systems Group. Design, development, and implementation of Gigabit ATM network for distributed multimedia system. Studied host interface design issues. ATM network simulation.
- Spring 1989, Fall 1990, Spring 1995: Teaching assistant, Computation Structures digital systems course. (Spring 1995 Head TA)
- 1988 - 1989: Head laboratory teaching assistant for Computation Structures. Responsibilities included writing and revising lab assignments, and maintaining the lab.
- 1987: Laboratory teaching assistant for Computation Structures.
- 1987: Design, construction, and programming of 16-bit computer.

Agora Technology Group, Incorporated

- 1994 - 1996: Founder and CEO. Conceived and oversaw development of targeted advertising-supported Web sites. Responsible for company's vision and direction. Attended the first two WWW Conferences, presenting a workshop and paper at the first, and appearing on the "Commercialization and Economics of the Web" panel and chairing the "Where Commercial Services and the Web Are Headed" panel at the second. Sold company intact; is currently an operating stand-alone company.

AT&T Bell Laboratories, Holmdel, NJ

- 1989 - 1990: Implementation of cascadable all-optical fiber logic gate. Modelocking of all-fiber erbium laser. Construction of modelocked laser repetition rate booster. Strong optics laboratory and fiber optic experience.
- Summer 1988, 1987: Research in integrated optics. Analysis of rectangular waveguides using microwave modeling. Fabrication of integrated optical components.

Honors

- MIT Alumni Association Great Dome Award, 2010, Baker 60th Anniversary Reunion Co-Chair (highest group award given by MIT Alumni Association)
- MIT Alumni Association Presidential Citation Award (now known as Great Dome), 2008, Member of MIT Chairman's Salon committee
- MIT Alumni Association Bronze Beaver Award, 2007 (highest individual award given by MIT Alumni Association)
- MIT Alumni Association Volunteer Honor Roll, February 2004
- MIT Alumni Association Lobdell Award, 1999
- Boston Museum of Science Gold Pin for 1000 hours of Volunteer Service, April 1999
- MIT Alumni Association Presidential Citation Award (now known as Great Dome), 1997, Member of Alumni Online Communications Committee

Patents and Patent Publications

- U.S. Patent #9,697,231, H. Houh, J. Stern, R. Spina, M. Meteer, “Methods and apparatus for providing virtual media channels based on media search,” July 4, 2017.
- U.S. Patent #9,697,230, H. Houh, J. Stern, “Methods and apparatus for dynamic presentation of advertising, factual, and informational content using enhanced metadata in search-driven media applications,” July 4, 2017. See also WO2007056485.
- U.S. Patent #7,975,296, L. Apfelbaum, H. Houh, T. Mayberry and G. Friedman, “Automated security threat testing of web pages,” July 5, 2011. See also US20030159063, WO2003067405.
- U.S. Patent #7,877,736, H. Houh and J. N. Stern, “Computer language interpretation and optimization for server testing,” January 25, 2011. See also US20050138104, WO2005043300.
- U.S. Patent #7,801,910, H. Houh and J. N. Stern, “Method and apparatus for timed tagging of media content,” September 21, 2010. See also US20070112837, US20090222442, WO2007056535.
- U.S. Patent #7,590,542, D. C. Williams, W. C. Hand, H. Houh, A. R. Seeley, “Method of Generating Test Scripts Using a Voice-Capable Markup Language,” September 15, 2009. See also EP1530869, US20030212561, WO2003096663.
- U.S. Patent #6,967,963, H. H. Houh, P. Anderson, C. Gadda, “Telecommunication method for ensuring on-time delivery of packets containing time-sensitive data,” November 22, 2005. See also EP1060400, WO2000033092, CA2318774.
- U.S. Patent #5,144,375, M. C. Gabriel, H. H. Houh, N. A. Whitaker, “Sagnac Optical Logic Gate,” September 1, 1992. Also issued as European Patent # EP0456422, July 23, 1997, German Patent #DE69126913, August 28, 1997
- U.S. Patent Application Publication No. 20020015387, “Voice Traffic Packet Capture and Analysis Tool for a Data Network” (Abandoned in 2007)
- U.S. Patent Application Publication No. 20020016708, “Method and Apparatus for Utilizing a Network Processor as Part of a Test System” (Abandoned in 2007)
- U.S. Patent Application Publication No. 20020016937, “Method and Apparatus for Utilizing a Network Processor as Part of a Test System.” (Abandoned in 2007) See also WO2002011413.
- U.S. Patent Application Publication No. 20070106646, “User-directed navigation of multimedia search results” (Abandoned in 2009)
- U.S. Patent Application Publication No. 20070106660, “Method and apparatus for using confidence scores of enhanced metadata in search-driven media applications” (Abandoned in 2009)
- U.S. Patent Application Publication No. 20070106685, “Method and apparatus for updating speech recognition databases and reindexing audio and video content using the same” (Abandoned in 2015)
- U.S. Patent Application Publication No. 20070118873, “Methods and apparatus for merging media content” (Abandoned in 2015)
- U.S. Patent Application Publication No. 20090222442, “User-directed navigation of multimedia search results” (Abandoned in 2016)
- U.S. Patent Application 11/395,732, “Search snippet creation for audio and video data” (Abandoned in 2009)
- U.S. Patent Application 11/774,931, “Methods and apparatus for managing a social networking web site”
- U.S. Patent Application 11/774,947, “Methods and apparatus for organizing media files”
- U.S. Patent Application 11/774,956, “Methods and apparatus for managing an online event”

- U.S. Provisional Patent Application 61/086,909, “Measuring and ranking relationship activity”
- U.S. Provisional Patent Application 61/086,914, “Detecting media object commonality”
- U.S. Provisional Patent Application 61/086,904, “Message categorization based on message characteristics”
- U.S. Provisional Patent Application 61/086,905, “Photo tagging to request action”

Trials, Testimony and Depositions

- Case No. 1:06CV682 (CMH/BRP), Verizon vs. Vonage, U.S. District Court for the Eastern District of Virginia, filed expert report, was deposed and testified at trial.
- Case No. 1:08CV157 (CMH/TRJ), Verizon vs. Cox, U.S. District Court for the Eastern District of Virginia, was deposed and testified at trial.
- Case No. 5:09-cv-476, Two-Way Media vs. AT&T, U.S. District Court for the Western District of Texas, filed expert report, testified at trial.
- Case No. 2:10-cv-248 (RAJ/FBS), ActiveVideo Networks vs. Verizon, U.S. District Court for the Eastern District of Virginia, filed expert report and was deposed as an expert witness.
- Case No. 1:11-cv-00880-TSE-JFA, Bear Creek Technologies, Inc. vs. Verizon Services Corp., et al, U.S. District Court for the Eastern District of Virginia, was deposed as an expert witness.
- Case No. 3:10-CV-298-BBC, AlmondNet, Inc. vs. Microsoft Corp., U.S. District Court for the Western District of Wisconsin, filed expert report.
- Case No. 6:10-cv-00597, Guardian Media Technologies, Ltd. Vs. AT&T Operations, Inc. et al., U.S. District Court for the Eastern District of Texas, Tyler Division, filed expert report.
- Case No. ESCV2010-02282C, The Octopus Solution LLC v. Gary Brown et al., Essex, MA Superior Court, testified at trial.
- Investigation No. 337-TA-882, In the matter of Certain digital media devices, including televisions, Blu-ray disc players, home theater systems, tablets and mobile phones, components thereof and associated software, U.S. International Trade Commission, filed expert reports, was deposed and testified at hearing.
- Investigation No. 337-TA-995, In the matter of Certain communications or computing devices, and components thereof, U.S. International Trade Commission, filed expert reports, and was deposed.
- Case No. 8:12-cv-122-LES-TDT, Prism Technologies LLC v. AT&T Mobility LLC, U.S. District Court for the District of Nebraska, filed expert report, was deposed and testified at trial.
- Case No. 11-2684 (JWL), Sprint Communications Co., L.P., v. Comcast Cable Communications, LLC, et al., U.S. District Court for the District of Kansas, filed expert report and was deposed.
- Case No. 11-2686 (JWL), Sprint Communications Co., L.P., v. Time Warner Cable, Inc., et al., LLC, et al., U.S. District Court for the District of Kansas, filed expert report, was deposed, and testified at trial.
- Case No. 12-487 (SLR), Cox Communications, Inc., et al., v. Sprint Communications Company L.P., et al., U.S. District Court for the District of Delaware, filed expert report and was deposed.
- Case IPR 2014-00039, Microsoft Corporation v. B.E. Technology LLC, U.S. Patent No. 6,628,314. Submitted declaration and cross-examined in deposition.
- Case IPR 2014-00086, Apple, Inc. v. Evolutionary Intelligence, LLC, U.S. Patent No. 7,010,536. Submitted declaration and cross-examined in deposition.
- Case No. 6:11-CV-421, Stragent, LLC v. Intel Corporation, U.S. District Court for the Eastern District of Texas, Tyler Division, deposed and testified at trial as a fact witness.

- IPR2014-01366, IPR2014-01367, IPR2014-01368, Samsung et al. v. Straight Path, submitted IPR declarations and was deposed.
- IPR2015-1006, IPR2015-1007, Cisco v. Straight Path, submitted IPR declarations.
- IPR2014-01457, IPR2014-01459, Microsoft v. Biscotti, submitted IPR declaration and was deposed.
- IPR2016-00302, Apple v. Nonend, submitted IPR declaration.
- IPR2014-00812, Yelp and Twitter v. Evolutionary Intelligence, submitted IPR declaration.
- IPR2015-00307, Cisco v. AIP, submitted IPR declaration.
- Microsoft v. Acacia, submitted IPR declaration.
- IPR2015-00342, AT&T Mobility LLC and Cellco Partnership d/b/a Verizon Wireless v. Solocron, submitted IPR declaration.
- Mediatube v. Bell Canada, Canadian Federal Court, submitted expert report and testified at trial.
- CBM2016-00036, Google v. At Home Bondholder's Liquidating Trust, submitted CBM declaration.
- IPR2016-01198, IPR2016-01201, Apple, Inc. v. VoIP-Pal.com, Inc., submitted IPR declaration and was deposed.
- Case No. 3:14-cv-2824-VC, Trend Micro Incorporated v. Rpost Holdings, Inc., et al, submitted expert report and was deposed
- Akamai v. Limelight Networks, IPR2016-01631, IPR2016-01711, IPR2017-00249, IPR2017-00349, submitted IPR declarations and was deposed
- Hyperlync Technologies v. Verizon Sourcing, No. 650151/2015 New York Supreme Court, submitted declaration and was deposed
- Cascades Streaming Technologies v. Big Ten Networks, submitted expert reports and was deposed
- Klaustech, Inc. v. Google LLC., Case No. 4:10-CV-05899-JSW (N.D. Cal.), submitted expert report and was deposed
- Cisco v. Uniloc, IPR2017-00198, IPR2017-00058, submitted IPR declarations and was deposed.
- Cisco v. Uniloc, IPR2017-00597, submitted IPR declaration and was deposed
- Twitter v. Youtoo Technologies, LLC, IPR2017-01133, IPR2017-01131, submitted IPR declarations and was deposed
- Taser International Inc. (now Axon Enterprise, Inc.) v. Digital Ally Inc., IPR2017-00375, submitted IPR declaration and was deposed
- Taser International Inc. (now Axon Enterprise, Inc.) v. Digital Ally Inc., IPR2017-00376, IPR2017-00515, IPR2017-00775, submitted IPR declarations
- HTC Corporation v. IPA Technologies, IPR2018-00306, IPR2018-00307, submitted IPR declaration
- Hulu v. Soundview, IPR2018-00864, IPR2018-01023, submitted IPR declarations and was deposed
- XpertUniverse v. Cisco, 3:17-cv-3848, U.S. District Court for Northern District of California, submitted declaration in support of claim construction and was deposed
- Improved Search v. Microsoft, C.A. No. 16-cv-650-JFB-SRF (Del.), submitted declaration in support of claim construction
- Apple v. Uniloc, IPR2018-00361, IPR2018-00394, IPR2018-00395, IPR2019-00700, IPR2019-00701, IPR2019-01667, submitted IPR declarations
- LG Electronics v. Uniloc, submitted IPR declaration
- SRC Labs and Saint Regis Mohawk Tribe v. Microsoft, 2:18-cv-00321-JLR, U.S. District Court for Western District of Washington at Seattle, submitted declaration in support of claim construction and was deposed

- Verizon Wireless v. Barkan Wireless IP Holdings LP, submitted IPR declarations
- Verizon Wireless and Samsung Electronics America Inc. v. Barkan Wireless IP Holdings LP, IPR2019-199, IPR2019-200, IPR2019-234, IPR2019-631, IPR2019-632, submitted IPR declarations
- Samsung Electronics America Inc. v. Barkan Wireless IP Holdings LP, IPR2019-100, submitted IPR declaration
- Cisco Systems Inc. v. Meetrix IP LLC, IPR2019-539, IPR2019-540, IPR2019-541, IPR2019-542, IPR2019-543, IPR2019-544, submitted IPR declarations
- GMG Products LLC v. Traeger Pellet Grills LLC, PGR2019-24, PGR2019-34, PGR2019-35, PGR2019-36, submitted PGR declaration and was deposed
- Uniloc 2017 LLC v. Google LLC, 2:18-cv-00502-JRG-RSP, U.S. District Court for the Eastern District of Texas, Marshall Division, submitted declaration in support of claim construction
- Microsoft Corporation v. Uniloc 2017 LLC, IPR2020-00102, IPR2020-00103, submitted IPR declarations
- Microsoft Corporation and HP Inc. v. Syncloud Technologies, LLC, IPR2020-00316, IPR2020-01031, IPR2020-01032, IPR2020-01269, IPR2020-01270, IPR2020-01271, submitted IPR declarations
- Seven Networks, LLC v. Apple Inc, 2:19-cv-115-JRG, U.S. District Court for the Eastern District of Texas, Marshall Division, submitted declaration in support of claim construction
- Unified Patents Inc. v. Ortiz & Associates Consulting, LLC, IPR2019-00743, submitted IPR declaration
- Uniloc 2017 LLC v. Apple Inc., 3:19-cv-01905-JD, U.S. District Court for the Northern District of California, San Francisco Division, submitted declaration in support of claim construction and was deposed
- ClearPlay Inc. v. Dish Network LLC et al.; 2:14-cv-00191, U.S. District Court for the District of Utah, Central Division, submitted expert report and re-examination declarations, was deposed, and testified at trial
- AMC Networks, Inc. et al v. Sound View Innovations, LLC, IPR2020-00482, submitted IPR declaration
- Case No. 1:17-cv-01734-RGA, Sprint Communications Co., L.P., v. Charter Communications, Inc., et al., U.S. District Court for the District of Delaware, filed expert report and was deposed.
- CommScope Technologies LLC et al. v. Barkan Wireless IP Holdings LP et al., IPR2020-00827, IPR2020-00829, IPR2020-00831, IPR2020-00833, IPR2020-00835, IPR2020-00838, submitted IPR declarations
- Cisco Systems Inc. v. Monarch Networking Solutions LLC, IPR2020-01226, submitted IPR declaration
- CoolTV Network.com v. Blackboard, Inc., Facebook, Inc., International Business Machines Corporation, Kaltura, Inc., Limelight Networks, Inc., Microsoft Corporation, Ooyala, Inc., Snap, Inc., and Trapelo Corp., submitted declaration in support of claim construction and was deposed
- Facebook, Inc. v. Onstream Media Corporation, IPR2020-01507, IPR2020-01508, IPR2020-01525, IPR2020-01527, IPR2020-01528, submitted IPR declarations
- Hisense Co LTD. v. Maxell, LTD, IPR2020-01598, IPR2020-01600, submitted IPR declarations
- Microsoft Corp. v. O'Brien, IPR2021-00015, submitted IPR declaration and was deposed
- Sable Networks Inc. v. Cisco Systems, Inc., WA:20-CV-00288-ADA, U.S. District Court for the Western District of Texas, Waco Division, submitted declaration in support of claim construction
- Peloton Interactive, Inc. v. Icon Health and Fitness Inc et al., IPR2021-00342, submitted IPR declaration

- Hulu, LLC v. SITO Mobile R&D IP, LLC et al., CBM2020-00028, IPR2020-00158, IPR2020-00206 , IPR2020-00219 , IPR2020-00265 , IPR2020-00298 , IPR2020-00304 , IPR2020-00308 , submitted CMB and IPR declarations and was deposed
- Twitter, Inc. and Google LLC v. B.E. Technology, LLC, IPR2021-00482, IPR2021-00483, IPR2021-00484, IPR2021-00485, submitted IPR declarations
- MG Freesites LTD v. ScorpCast, LLC, IPR2021-00511, submitted IPR declaration
- Avionte, LLC v. Sandeep Acharya, Samar Basnet, and Aqore LLC, Arbitration, Hon. Jeffrey Keyes (Ret.)
- Microsoft Corporation v. WSOU Investments, LLC d/b/a Brazos Licensing and Development, IPR2021-00751, submitted IPR declaration
- SAP America, Inc. v. Express Mobile, Inc., IPR2021-01144, IPR2021-01145, IPR2021-01146, submitted IPR declarations
- StratosAudio Inc. v. Hyundai Motor America, 6:20-cv-01125-ADA, U.S. District Court for the Western District of Texas, Waco Division, submitted claim construction declaration and was deposed
- Peloton Interactive, Inc. v. Ifit Health and Fitness Inc, IPR2022-00029, IPR2022-00030, submitted IPR declarations and was deposed
- Investigation No. 337-TA-1263, In the matter of Certain televisions, remote controls, and components thereof, U.S. International Trade Commission, filed expert reports, was deposed, and testified at hearing
- Alexander Stross v. Netease, Inc., et al., U.S. District Court (Central District of California), Case 2:20-cv-00861-AB(JC), filed expert reports, and was deposed
- Ralph Ledergerber v. Netease, Inc., et al., U.S. District Court (Central District of California), Case 2:20-cv-00862-AB(JC), filed expert reports and was deposed
- John Walmsley v. Netease, Inc., et al., U.S. District Court (Central District of California), Case 2:20-cv-00863-AB(JC), filed expert reports and was deposed
- Bernhard Kuhmstedt v. Netease, Inc., et al., U.S. District Court (Central District of California), Case 2:20-cv-02044-AB(JC), filed expert reports and was deposed
- Netflix, Inc. v. CA, Inc., IPR2021-01319, submitted IPR declaration and was deposed
- F45 Training Pty Ltd. v. Body Fit Training USA Inc., C.A. No. 20-1194 (LPS) (D. Del.), filed expert report
- Apple v. Fintiv, IPR2022-00976, IPR2022-01149, IPR2022-01150, IPR2023-00399, submitted IPR declarations and was deposed
- Estech Systems, Inc. v. Howard Midstream Energy Partners d/b/a Howard Energy Partners, C.A. No. 6:20-cv-777-ADA (Western District of Texas, Waco Division), filed expert reports and was deposed
- Echelon Fitness Multimedia, LLC. v. Peloton Interactive, Inc., IPR2020-01187, submitted IPR declaration and was deposed
- VIZIO. V. Maxell Ltd., IPR2022-01458, submitted IPR declaration
- Maxell, Ltd. et el. v. VIZIO, Inc., Case No. 2-21-cv-06758 (C.D. Cal.). filed declaration and was deposed
- Microsoft Corporation v. SurfCast, Inc., submitted IPR declarations
- The Walt Disney Company et al. v. WAG Acquisition LLC, IPR2022-01227, IPR2022-01228, IPR2022-01346, submitted IPR declarations and was deposed
- Rakuten Rewards et al. v. IBM, submitted IPR declarations and was deposed
- Fubotv Media Inc. v. SITO Mobile R&D IP, LLC et al., submitted IPR declarations
- Unified Patents, LLC v. Noblewood IP LLC, IPR2022-01111, submitted IPR declaration
- Unified Patents, LLC. v. Pedersen, Peter Henrik, IPR2023-00029, submitted IPR declaration
- Ericsson Inc v. Dali Wireless Inc., submitted IPR declarations

- Universal Electronics, Inc. et al. v. Roku, Inc., IPR2022-00818, submitted declarations and was deposed
- Google LLC v. Wildseed Mobile LLC et al., IPR2023-00244, IPR2023-00245, IPR2023-00246, IPR2023-00247, IPR2023-00248, submitted declarations
- Lululemon Athletica Canada Inc. et al. v. Nike, Inc., IPR2023-00180, IPR2023-00424, submitted declarations
- Zoho Corporation, et al. v. Meetrix IP, LLC, IPR2023-00371, IPR2023-00377, IPR2023-00379, IPR2023-00380, IPR2023-00382, submitted declarations
- WSOU Investments, LLC d/b/a/ Brazos Licensing and Development v. Google LLC, 6:20-cv-572, U.S. District Court for the Western District of Texas, Waco Division, submitted reports and was deposed
- Google LLC v. Safecast Limited LLC, IPR2023-00652, submitted declaration
- Investigation No. 337-TA-1338, In the matter of Certain smart televisions, U.S. International Trade Commission, filed reports and was deposed
- Cisco Systems, Inc. v. Orckit IP LLC, IPR2023-00714, submitted declaration
- SupplyPro, Inc. v. Sandvik Machining Solutions AB, et al., AAA Case No. 01-22-000-7897. submitted declaration
- TrackThings LLC v. NETGEAR, Inc., Case No. 22-00981-RGA-JLH, U.S. District Court for the District of Delaware, submitted declaration
- Correct Transmission, LLC v. Nokia of America Corporation, Case No. 2:22-CV-00343-JRG-RSP, U.S. District Court for the Eastern District of Texas, Marshall Division, submitted export reports, was deposed, and testified at trial
- Walt Disney Parks & Resorts US Inc. v. Agile Journeys LLC; IPR2024-00029, submitted declaration
- Aylo Freesites Ltd v. Dish Technologies LLC; IPR2024-00043, IPR2024-00044, IPR2024-00045, IPR2024-00046, IPR2024-00047, IPR2024-00048, IPR2024-00146, IPR2024-00147; submitted declarations and was deposed
- Arm Inc. et al. v. Icpillar LLC; IPR2024-00476, IPR2024-00566; Retained by Arm; submitted declarations
- Microsoft Corp et al. v. Litl LLC, IPR2024-00454, IPR2024-00455, IPR2024-00456, IPR2024-00457, IPR2024-00458; submitted declarations
- Cisco Systems Inc et al. v. InfoExpress Inc, IPR2024-00539, IPR2024-00540, IPR2024-00677, IPR2024-00678; submitted declarations and was deposed
- Vizio Inc v. Multimedia Technologies Pte Ltd; IPR2024-00722, IPR2024-00723; Retained by Vizio; submitted declarations and was deposed
- Sap America Inc v. Isix IP LLC; IPR2024-00615; Retained by SAP; submitted declarations
- Cisco Systems Inc et al. v. Portsmouth Network Corp; IPR2024-00505, IPR2024-00506; submitted declarations and was deposed
- Aylo Freesites Ltd v. Dish Technologies LLC; IPR2024-00512, IPR2024-00513, IPR2024-00514, IPR2024-00515, IPR2024-00516, IPR2024-00517, IPR2024-00518, IPR2024-00519, IPR2024-00940, IPR2024-00941; submitted declarations and was deposed
- Aylo Freesites Ltd v. WellcomeMat LLC; IPR2024-00710, IPR2024-01101; submitted declarations and was deposed
- Verizon Corporate Service Group et al. v. Headwater Partners I LLC; IPR2024-00809, IPR2024-00945; Retained by Verizon; Technology: Mobile device management; Submitted declarations
- Microchip Technology Inc v. Aptiv Technologies Ltd et al.; IPR2024-00228; Retained by Aptiv; Technology: Universal Serial Bus; submitted declaration and was deposed

- Vusion Group SA v. Hanshow Technology Co Ltd; IPR2024-00963; Retained by VusionGroup; Technology: Video analysis; submitted declaration
- Maxell, Ltd. v. TCL Electronic Holdings Ltd. et al.; Case No. 5:23-cv-00108-RWS-JBB; U.S. District Court for the Eastern District of Texas, Texarkana Division; Retained by TCL; submitted declaration and was deposed
- Cerence Operating Company v. Samsung Electronics Co., LTD. et al., Case No. 2:23-cv-00482-JRG-RSP, U.S. District Court for the Eastern District of Texas, Marshall Division, submitted export report on claim construction and was deposed
- MasterClass and FuboTV v. Dish Technologies LLC; IPR2024-00917, IPR2024-00918, IPR2024-00919; submitted declarations and was deposed
- Microsoft Corp v. Dialect LLC; IPR2025-00655, IPR2025-00656, IPR2025-00657, IPR2025-00658, IPR2025-00659; submitted declarations

Publications

- “IP switching: server driven flow classification,” H. H. Houh, *Proceedings of the Washington University Workshop on Integration of IP and ATM*, November 1996.
- “Aurora at MIT,” D. D. Clark, H. H. Houh, and D. L. Tennenhouse, Editors, *MIT Laboratory for Computer Science Technical Report 673*, December 1995.
- “ViewStation Applications: Implications for Network Traffic,” C. J. Lindblad, D. Wetherall, W. Stasior, J. F. Adam, H. H. Houh, M. Ismert, D. Bacher, B. Phillips, and D. L. Tennenhouse, *IEEE Journal of Selected Areas in Communications*, 1995.
- “The VuNet Desk Area Network: Architecture, Implementation, and Experience,” H. H. Houh, J. F. Adam, M. Ismert, C. J. Lindblad, and D. L. Tennenhouse, *IEEE Journal of Selected Areas in Communications*, 13 (4), May, 1995.
- “Reducing the Complexity of ATM Host Interfaces,” H. H. Houh and D. L. Tennenhouse, Hot Interconnects II Symposium Proceedings, Stanford, August 11-12, 1994.
- “Media-intensive data communications in a ‘desk-area’ network,” J. F. Adam, H. H. Houh, M. Ismert, and D. L. Tennenhouse, *IEEE Communications*, August 1994.
- “ViewStation Applications: Intelligent Video Processing Over A Broadband Local Area Network,” C. J. Lindblad, D. J. Wetherall, W. Stasior, B. Phillips, D. Bacher, J. Adam, H. Houh, M. Ismert, and D. L. Tennenhouse, Proceedings of the 1994 USENIX Symposium on High-Speed Networking, Oakland, CA, August 1994.
- “The Media Gateway: Live Video on the World Wide Web,” H. H. Houh, C. J. Lindblad, J. Soo, D. L. Tennenhouse, and D. J. Wetherall, Workshop at the 1994 World Wide Web Conference, Geneva, Switzerland, May 1994.
- “Active Pages: Intelligent Nodes on the World Wide Web,” H. H. Houh, C. J. Lindblad, and D. J. Wetherall, Proceedings of the 1994 World Wide Web Conference, Geneva, Switzerland, May 1994.
- “Wavelength Division vs. Code Division Access Methods for Optical Networks,” H. H. Houh, Area Exam Paper, May 1993.
- “Experience with the VuNet: A Network Architecture for a Distributed Multimedia System,” J. F. Adam, H. H. Houh, D. L. Tennenhouse, Proceedings of the 18th Conference on Local Computer Networks, pp. 70-76, September 1993
- “The VudBoard: A Simple DMA Interface,” H. H. Houh, Proceedings of the 4th Gigabit Minijam, January 1994.

- “A Software-Oriented Approach to the Design of Media Processing Environments,” D. L. Tennenhouse, J. Adam, D. Carver, H. Houh, M. Ismert, C. Lindblad, W. Stasior, D. Weatherall, D. Bacher, and T. Chang., submitted to the International Conference on Multimedia Computing and Systems, May 1994.
- “A Network Architecture for Distributed Multimedia Systems,” J. F. Adam, H. H. Houh, M. Ismert, and D. L. Tennenhouse, submitted to the International Conference on Multimedia Computing and Systems, May 1994.
- “The Viewstation Collected Papers,” D. L. Tennenhouse, J. Adam, C. Compton, A. Duda, D. Gifford, H. Houh, M. Ismert, C. Lindblad, W. Stasior, R. Weiss, D. Wetherall, D. Bacher, D. Carver, and T. Chang, MIT Laboratory for Computer Science Technical Report, MIT/LCS/TR-590, November 1993.
- “The Viewstation Collected Papers II,” W. F. Stasior, D. L. Tennenhouse, J. F. Adam, D. R. Bacher, H. H. Houh, M. Ismert, C. J. Lindblad, B. M. Phillips, D. J. Wetherall, MIT Laboratory for Computer Science Technical Report, MIT/LCS/TR-696, April 1996.
- “A System's Perspective of the Sagnac Fiber Logic Gates and Their Possible Applications,” A. Huang, N. Whitaker, C. Gabriel, H. Avramopoulos, P. M. W. French, H. H. Houh, and I. Chuang, Applied Optics, September 10, 1994
- “Complete Switching in a Three-Terminal Sagnac Switch,” H. Avramopoulos, P. M. W. French, M. C. Gabriel, H. H. Houh, N. A. Whitaker, T. Morse, IEEE Phot. Tech. Lett. **3** (3), 235
- “Complete Switching in a Three-Terminal Sagnac Switch,” H. Avramopoulos, P. M. W. French, M. C. Gabriel, H. H. Houh, N. A. Whitaker, IEEE/LEOS Annual Meeting, Paper PDP-13, November 1990
- “All-optical phase-locked oscillator,” N. A. Whitaker, Jr., H. H. Houh, H. Avramopoulos, T. F. Morse, IEEE/LEOS Annual Meeting, Paper ELT2.4/MOO3, November 1990
- “Passive modelocking of an all-fiber erbium laser,” H. Avramopoulos, H. H. Houh, N. A. Whitaker, M. C. Gabriel, T. F. Morse, IEEE/LEOS Conference on Optical Amplifiers and their Applications, Paper PDP-8, August 1990
- “Transverse modes, waveguide dispersion, and 30ps recovery in submicron GaAs/AlAs microresonators,” J. L. Jewell, S. L. McCall, A. Scherer, H. H. Houh, N. A. Whitaker, A. C. Gossard, and J. H. English, Appl. Phys. Lett. **55** (1), July 3, 1989

Leadership, Activities and Interests

- Leadership
 - Reading Symphony Orchestra
 - Board of Directors (Treasurer), 2022 – present
 - Discovery Museum (Acton, MA)
 - Board of Directors, 2023 – present
 - Board of Directors, Development Committee, 2023 – present
 - Science and Technology Advisory Council, 2012 – present
 - The Discover Museum received the 2024 National Medal for Museum and Library Service, the nation’s highest honor given to museums and libraries that make significant and exceptional contributions to their communities.
 - Tau Beta Pi National Engineering Honor Society
 - Executive Councilor and Secretary, 2024 – 2026
 - Advisor, MA B Chapter at MIT, 2003 – present
 - District Director (National Officer), Tau Beta Pi, New England Area, 1991 – 2003

- President, MA B Chapter at MIT, Fall 1988 – Spring 1989
- Laureate award, 1989
- MIT Alumni Association Board of Directors
 - MIT Annual Giving Board, 3 year term, 2024 – 2027
 - National Selection Committee, 3 year term, 2020 – 2023
 - K-12 STEM Initiatives Co-chair, 2013 – 2017
 - Awards Committee Chair, 2012 – 2014
 - Awards Committee, 3 year term, 2011 – 2014
 - Vice President, 2 year term, 2004 – 2006
 - Alumni Ad-Hoc Committee, Chair, 2005-2006
 - Board Member, 2 year term, 1997 – 1999
- MIT Club of Boston
 - Board of Directors, 2006 – 2011
 - K-12 Initiatives Chair, 2009 – 2012
 - VP of Communications, MIT Club of Boston, 2003 – 2006
 - Past-President, MIT Club of Boston, 2002 – 2003
 - President, MIT Club of Boston, 2001 – 2002
 - President-Elect, MIT Club of Boston, 2000 – 2001
 - VP of Programs, MIT Club of Boston, 1999 – 2000
 - Activities Super-Chair, MIT Club of Boston, 1998 – 1999
- MIT Enterprise Forum of Cambridge, Inc.
 - Past Chair, 2009 – 2011
 - In-NOW-vation Co-chair, 2010
 - Chair, 2007 – 2009
 - Vice Chair, 2005 – 2007
 - Executive Board Member, 2002 – 2011
 - Winter Workshop Co-Chair, February 2007 – conceived idea for conference, which sold-out and produced largest attendance numbers in recent memory
 - Spring Workshop Co-Chair, Spring 2004
 - Membership Committee Chair, Fall 2003 - 2006
 - 25th Anniversary Dinner Chair, Fall 2003
 - As Membership Chair and Board Member, started Special Interest Groups in 2004; a SIG won the MIT Presidential Citation award, the MIT Alumni Association's highest award for organizations, in 2006
- Estabrook Elementary School PTA
 - Advisory committee to the superintendent on PCB issue, 2010-2011
 - 4th Grade after-school science program co-organizer, 2010-2012
 - 4th and 5th Grade before-school Math Olympiad co-organizer, 2009-2013
 - 5th Grade BBQ and Yearbook Committee, 2011, 2013
 - Family Math Night volunteer, 2008-2012
- MIT Class of 1989
 - 35-year Reunion Gift Committee, Co-chair, 2024
 - Secretary, five consecutive 5 year terms, 1989 - 2014
 - 30-year Reunion Committee, 2019
 - 25-year Reunion Committee, 2014
 - 20-year Reunion Committee and Gift Committee, 2009
 - 15-year Reunion Committee and Gift Committee, 2004
 - 10-year Reunion Committee and Gift Committee, 1999

- 5-year Reunion Committee, 1994
 - Interim Treasurer, 1993 - 1994
 - Instituted annual senior class career fair, now raising over \$100,000 annually for senior class activities, Fall 1988
 - Strong, consistent record of leadership dating to high school
- Acting
 - '21' (Sony Pictures), credited as "Chinatown Dealer," 2007, Kevin Spacey's movie about the MIT Blackjack Team inspired by "Bringing Down the House" by Ben Mezrich, opened nationwide on March 28, 2008. 21 was the number one movie in US for two weeks and number one globally for one week. 21 also topped the DVD sales, Blu-ray sales and DVD rental charts.
 - Spring Lake Theater Company, first New York-area off-broadway production of "A Chorus Line," played role of Mark, Summer 1990
- Former member of the MIT Blackjack team
- Producer for 10,000 Maniacs' 2013 album "Music from the Motion Picture"
- Executive Producer for 10,000 Maniacs' 2015 album "Twice Told Tales"
- Violist, violist, harpist, guitarist, singer, actor: played in many amateur/semi-professional groups including Firebird Pops Orchestra, Reading Symphony Orchestra, MIT Summer Philharmonic Orchestra, Carlisle Chamber Orchestra, Rivers Symphony Orchestra, Opera51 Orchestra, Freisinger Chamber Orchestra, Merrimack Valley Philharmonic, Longwood Symphony, MIT Symphony, and Somerville Community Chorus
- Violist, "Coming Home" album by Karen Phillips, 2014
- Google Scholar citations: 2224, h-index: 22, i10-index: 27 (since 2019: 539/12/12)
- Erdos-Bacon number of 7
- Erdos-Bacon-Sabbath number of 11