

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

VUSIONGROUP SA
Petitioner,

v.

HANSHOW TECHNOLOGY CO., LTD.
Patent Owner.

IPR2024-00963
U.S. Patent No. 10,701,321

**DECLARATION OF DR. HENRY HOUH,
UNDER 37 C.F.R. § 1.68 IN SUPPORT OF PETITION FOR
INTER PARTES REVIEW**

TABLE OF CONTENTS

I.	Introduction.....	1
II.	Qualifications and Professional Experience.....	4
III.	Level of Ordinary Skill in the Art	9
IV.	Relevant Legal Standards	10
V.	SUMMARY OF THE '321 PATENT.....	15
	A. Specification.....	15
	B. Prosecution History	17
VI.	Technology Background.....	19
	A. Early digital video capture systems.....	19
	B. Cloud Computing	20
	C. Network-based cameras	21
	D. Video compression.....	22
	E. State of object recognition/detection.....	24
VII.	Claim Construction.....	25
VIII.	Claims 1-21 are Unpatentable	26
	A. Ground 1A: Laganieri '363 Renders Claims 1-5, 7-21 Obvious in view of Laganieri '817 and Laganieri '461	27
	1. Summary of Laganieri '363	27
	2. Summary of Laganieri '817	30
	3. Summary of Laganieri '461	32
	4. Independent Claim 1	34
	5. Independent Claim 13.....	61

6. Claims 2 and 14: “wherein the object detection system creates cropped images from the image data that include the detected objects of interest and sends the cropped images to the video analytics system.” 64
7. Claims 3 and 15: “wherein the image data include video data and the object detection system provides frames from the video data that include the detected objects of interest and sends the frames to the video analytics system as the image data.”.....65
8. Claims 4 and 16: “The system of claim 1, wherein the video analytics system includes an object analysis system that analyzes the image data corresponding to objects of interest detected by the object detection system by comparing the image data to reference images to identify the objects of interest, by recognizing patterns in the image data, and/or by applying image processing functions, neural network functions, deep learning functions, and/or object classification functions to the image data.” 67
9. Claims 5 and 18: “wherein the object detection system is integrated within the one or more surveillance cameras.” 68
10. Claims 7 and 17: “further comprising a database that receives metadata associated with the identified objects of interest from the video analytics system.” 69
11. Claims 8 and 19: “further comprising a gateway system that provides access to the video analytics system over a public network.”.....71
12. Claims 9 and 20: “wherein the object detection system is integrated within the gateway system.” 72
13. Claims 10 and 21: “wherein the one or more surveillance cameras each include a gateway system that provides access to the video analytics system over a public network.” 74
14. Claim 11: “wherein the video analytics system receives the image data from the object detection system after being transmitted over an enterprise network.” 75

15.	Claim 12: “wherein the video analytics system receives the image data from the object detection system after being transmitted over a cloud network.”.....	77
B.	Ground 1B: Laganieri ’363 Renders Claims 4, 6, 16, Obvious in View of Laganieri ’817, and Chen, with or without Laganieri ’461.....	78
1.	Summary of Chen.....	78
2.	Motivation to Combine.....	80
3.	Claims 4 and 16: “The system of claim 1, wherein the video analytics system includes an object analysis system that analyzes the image data corresponding to objects of interest detected by the object detection system by comparing the image data to reference images to identify the objects of interest, by recognizing patterns in the image data, and/or by applying image processing functions, neural network functions, deep learning functions, and/or object classification functions to the image data.”	82
4.	Claim 6: “The system of claim 1, wherein the image data corresponding to objects of interest detected by the object detection system are compared by the video analytics system to reference images to identify the objects of interest.”	83
C.	Ground 2: Bedros Renders Claims 1-21 Obvious.....	84
1.	Summary of Bedros	84
2.	Independent Claim 1	88
3.	Independent Claim 13.....	118
4.	Claims 2 and 14: “wherein the object detection system creates cropped images from the image data that include the detected objects of interest and sends the cropped images to the video analytics system.”	120
5.	Claims 3 and 15: “wherein the image data include video data and the object detection system provides frames from the video	

	data that include the detected objects of interest and sends the frames to the video analytics system as the image data.”.....	121
6.	Claims 4, 6, and 16: “wherein the video analytics system includes an object analysis system that analyzes the image data corresponding to objects of interest detected by the object detection system by comparing the image data to reference images to identify the objects of interest, by recognizing patterns in the image data, and/or by applying image processing functions, neural network functions, deep learning functions, and/or object classification functions to the image data.”	123
7.	Claims 5 and 18: “wherein the object detection system is integrated within the one or more surveillance cameras.”	125
8.	Claims 7 and 17: “further comprising a database that receives metadata associated with the identified objects of interest from the video analytics system.”	127
9.	Claims 8 and 19: “further comprising a gateway system that provides access to the video analytics system over a public network.”.....	127
10.	Claims 9 and 20: “wherein the object detection system is integrated within the gateway system.”	129
11.	Claims 10 and 21: “wherein the one or more surveillance cameras each include a gateway system that provides access to the video analytics system over a public network.”	129
12.	Claim 11: “wherein the video analytics system receives the image data from the object detection system after being transmitted over an enterprise network.”	130
13.	Claim 12: “wherein the video analytics system receives the image data from the object detection system after being transmitted over a cloud network.”	131
IX.	Conclusion	133

I, Henry Houh, do hereby declare as follows:

I. INTRODUCTION

1. I make this declaration based upon my own personal knowledge and, if called upon to testify, would testify competently to the matters stated herein.

2. I have been asked by VusionGroup SA (“VusionGroup”) to provide my expert opinion connection with the *Inter Partes* Review of U.S. Patent No. 10,701,321 (“the ’321 patent”) to Westmacott *et al.* concerning whether it is unpatentable over certain prior art. This declaration is a statement of my opinions on issues relating to the patentability of claims 1–21 (“the Challenged Claims”) of the ’321 patent. As I explain more fully below, it is my opinion that all of the Challenged Claims would have been obvious to a person of ordinary skill in the art (“POSITA”) at the time of the alleged invention.

3. I am being compensated for my work in this matter at my standard hourly rate of \$820. I am also being reimbursed for reasonable and customary expenses associated with my work and testimony in this investigation. My compensation is not contingent on the outcome of this matter or the specifics of my testimony.

4. I have reviewed and considered the following documents in connection with my analysis of the ’321 patent:

- EX1001: U.S. Patent No. 10,701,321 (“’321 patent” or “’321Pat”)

- EX1002: File History of the '321 patent (“’321Pat File History”)
- EX1005: U.S. Patent Application Publication No. 2012/0195363 (“Laganiere ’363”)
- EX1006: U.S. Patent No. 8,139,817 (“Laganiere ’817”)
- EX1007: U.S. Patent Application Publication No. 2011/0113461 (“Laganiere ’461”)
- EX1008: U.S. Patent No. 8,634,601 (“Bedros”)
- EX1009: U.S. Patent Application Publication No. 2014/0044348 (“Chen”)
- EX1010: Excerpt of THE AUTHORITATIVE DICTIONARY OF IEEE STANDARDS TERMS, 7th ed. (2000)
- EX1011: Excerpt of Merriam-Webster’s Collegiate Dictionary, 11th ed. (2005)
- EX1012: Excerpt of NIST Special Publication 800-53, rev. 4 (Jan. 22, 2015)
- EX1013: Excerpt of REVERSE ACRONYMS, INITIALISMS AND ABBREVIATIONS DICTIONARY, 10th ed., vol. 3, part 1 (Oct. 23, 1985) (“Reverse Acronyms”)
- EX1014: U.S. Patent No. 8,619,150 (“Deever ’150”)
- EX1015: U.S. Patent No. 8,432,965 (“Deever ’965”)
- EX1016: Excerpt of Dictionary of Contemporary Photography (1974)
- EX1017: Excerpt of AXIS 211M Network Camera User’s Manual (Feb. 2010) (“AXIS 211M manual”)
- EX1021: U.S. Patent Application Publication No. 2016/0050396 (“Gali”)
- EX1022: U.S. Patent Application Publication No. 2016/0191779

(“Yang”)

- EX1023: U.S. Patent Application Publication No. 2013/0166711 (“Wang”)
- EX1024: Christopher John Lindblad, A Programming System for the Dynamic Manipulation of Temporally Sensitive Data, (1994) (Ph.D. dissertation, Massachusetts Institute of Technology) (“Lindblad”)
- EX1025: David James Wetherall, An Interactive Programming System for Media Computation (1994) (Master of Science dissertation, Massachusetts Institute of Technology) (“Wetherall”)
- EX1026: William F. Stasior, An Interactive Approach to the Identification and Extraction of Visual Events (1997) (Ph.D. dissertation, Massachusetts Institute of Technology) (“Stasior”)
- EX1027: Jia Deng et al., ImageNet: A Large-Scale Hierarchical Image Database, *2009 IEEE Conference on Computer Vision and Pattern Recognition*, (2009) (“Deng”)
- EX1028: Alex Krizhevsky et al., ImageNet Classification with Deep Convolutional Neural Networks, *Comm’n of the ACM*, vol. 60, issue 6, pp. 84-90 (2017) (“Krizhevsky”)
- EX1029: Andrew Tanenbaum, *Computer Networks* (3rd ed. 1996) (“Tanenbaum”)
- EX1030: Joel F. Adam et al., The Vidboard: A Video Capture and Processing Peripheral for a Distributed Multimedia System, *Multimedia Systems*, vol. 2, pp. 150-56 (1994) (“Adam”)

5. In forming the opinions expressed below, I have considered:

- a) the documents listed above;
- b) the relevant legal standards, including the standard for

obviousness, and any additional authoritative documents as cited in the body of this declaration; and

c) my own knowledge and experience based upon my work in the field of software systems as described below, as well as the following materials

d) any other documents not listed above but cited below.

6. Unless otherwise noted, all emphasis in any quoted material has been added.

II. QUALIFICATIONS AND PROFESSIONAL EXPERIENCE

7. My complete qualifications and professional experience are described in my *Curriculum Vitae*, a copy of which can be found in Exhibit 1019. The following is a brief summary of my relevant qualifications and professional experience.

8. I received a Ph.D. in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology (“MIT”) in 1998. My dissertation advisor was David Tennenhouse. 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.

9. 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.

10. Beginning in 1987, I worked in telecommunications 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 Science Department’s core Computer Architectures course.

11. 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. The software system was designed for building applications in which audio and video input is analyzed and actions may be taken based upon the analysis. Example applications included detecting people that enter a room, capturing images of a whiteboard when it is not obscured by the person writing, and automatically segmenting videos, such as breaking down a talk-show monolog into individual jokes. In addition to working on the design of core network components (which was based on FPGAs), designing and building the high-speed links (which

included programmable logic devices), and designing and writing the device drivers for the interface cards (which included programmable logic devices), 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.

12. 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 to the conclusions in my thesis.

13. 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.

14. 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.

15. 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 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.

16. From 1999-2004, I was employed by Empirix and its predecessor company, Teradyne. 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.

17. 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. While I was Vice President of Operations and Technology, PodZinger

followed its initial prototype with a full streaming audio and video search solution. 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.

III. LEVEL OF ORDINARY SKILL IN THE ART

18. I understand there are multiple factors relevant to determining the level of ordinary skill in the pertinent art, including (1) the levels of education and experience of persons working in the field at the time of the invention; (2) the sophistication of the technology; (3) the types of problems encountered in the field; and (4) the prior art solutions to those problems.

19. I understand that the '321 patent was filed on February 27, 2015. My analysis reflects the views of a person of ordinary skill in the art in the field of the '321 patent as of February 27, 2015. This would have been someone knowledgeable and familiar with the field of video surveillance systems that are pertinent to the '321 patent. That person would have a bachelor's degree in electrical engineering or computer science, or equivalent training, and approximately two years of experience working in the field of video processing systems. Lack of work experience can be remedied by additional education, and vice versa.

20. As of February 27, 2015, I met, and in fact exceeded, the qualifications of a person of ordinary skill in the art.

21. For purposes of this Declaration, in general, and unless otherwise noted, my statements and opinions, such as those regarding my experience and the understanding of a POSITA generally (and specifically related to the references I consulted herein), reflect the knowledge that existed in the field as of the earliest proper priority date of the '321 patent (i.e., February 27, 2015). Unless otherwise stated, when I provide my understanding and analysis below, it is consistent with the level of a POSITA prior to the earliest alleged priority date of the '321 patent.

IV. RELEVANT LEGAL STANDARDS

22. I am not an attorney. In preparing and expressing my opinions and considering the subject matter of the '321 patent, I am relying on certain basic legal principles that counsel have explained to me. These principles are discussed below.

23. I understand that for an invention claimed in a patent to be found patentable, it must be, among other things, new and not obvious in view of what was known before the invention was made.

24. I understand the information that is used to evaluate whether an invention is new and not obvious is generally referred to as “prior art” and generally includes patents and printed publications (e.g., books, journal publications, articles on websites, product manuals, etc.).

25. I understand that in this proceeding the Petitioner has the burden of proving that the claims of the '321 Patent are anticipated by or obvious in view of

the prior art by a preponderance of the evidence. I understand that “a preponderance of the evidence” is evidence sufficient to show that a fact is more likely true than it is not.

26. I understand that prior art to the '321 patent in this proceeding includes patents and printed publications in the relevant art that predate the proper priority date of the alleged invention recited in the '321 patent.

27. I understand that terms appearing in the patent claims are to be interpreted according to their “ordinary and customary meaning” in an IPR proceeding. In determining the ordinary and customary meaning, the words of a claim are first given their plain meaning as they would have been understood by a POSITA at the time of the alleged invention, in light of the specification and file history. I understand that treatises and dictionaries may be consulted, albeit under limited circumstances, to determine the meaning attributed by a POSITA at the time of the alleged invention. I have followed this approach in my analysis and have applied the ordinary and customary meaning of those terms throughout my analysis in this Declaration.

28. I understand that there are two ways in which prior art may render a patent claim unpatentable. First, the prior art can be shown to “anticipate” the claim. Second, the prior art can be shown to have made the claim “obvious” to a person of ordinary skill in the art. In this declaration I focus on “obviousness.” My

understanding of the legal standard for “obviousness” is set forth below.

29. I have been informed that a claimed invention is unpatentable as obvious if the differences between the invention and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which the subject matter pertains. I have also been informed by counsel that the obviousness analysis takes into account factual inquiries including the level of ordinary skill in the art, the scope and content of the prior art, and the differences between the prior art and the claimed subject matter.

30. Specifically, I understand that the obviousness question requires consideration of four factors (although not necessarily in the following order):

- The scope and content of the prior art;
- The differences between the prior art and the claims at issue;
- The knowledge of a person of ordinary skill in the pertinent art; and
- Whatever objective factors indicating obviousness or non- obviousness may be present in any particular case.

31. I have been informed by counsel that the Supreme Court has recognized several rationales for combining references or modifying a reference to show obviousness of claimed subject matter. Some of these rationales include the following: (a) combining prior art elements according to known methods to yield

predictable results; (b) simple substitution of one known element for another to obtain predictable results; (c) use of a known technique to improve a similar device (method, or product) in the same way; (d) applying a known technique to a known device (method, or product) ready for improvement to yield predictable results; (e) choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success; and (f) some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention.

32. I further understand that certain factors may support or rebut the obviousness of a claim. I understand that such secondary considerations include, among other things, commercial success of the patented invention, skepticism of those having ordinary skill in the art at the time of invention, unexpected results of the invention, any long-felt but unsolved need in the art that was satisfied by the alleged invention, the failure of others to make the alleged invention, praise of the alleged invention by those having ordinary skill in the art, and copying of the alleged invention by others in the field. I understand that there must be a nexus—a connection—between any such secondary considerations and the alleged invention. I also understand that contemporaneous and independent invention by others is a secondary consideration tending to show obviousness.

33. I am not aware of any allegations by the named inventor of the '321

patent or any assignee of the '321 patent that any secondary considerations are relevant to the obviousness analysis of any Challenged Claim of the '321 patent.

34. I understand that a petition must identify, for any terms with means-plus-function limitations, the specific portions of the specification that describe the structure, material, or acts corresponding to each claimed function. I also understand a rebuttable presumption exists that without use of the term, “means,” a claim term is not interpreted as means-plus-function, but that other nonce words like “component” have been found to satisfy the rebuttable presumption and invoke a means-plus-function claim construction of such a claim term.

35. I also understand that a petitioner seeking to demonstrate that the prior art discloses a means-plus-function limitation must prove that the corresponding structure—or an equivalent—was present in the prior art and performing the same claimed function.

36. I understand that the standard for determining whether the prior art discloses the corresponding structure or an equivalent is whether there are insubstantial differences, which means that if the assertedly equivalent structure performs the claimed function in substantially the same way to achieve substantially the same result as the corresponding structure described in the specification, it is equivalent.

37. I understand that proving structural equivalence does not require a

component-by-component analysis, but rather that it is sufficient to show structure, materials, or acts that perform the claimed function in substantially the same way to achieve substantially the same result because the claim limitation is the overall structure corresponding to the claimed function and not individual components.

38. I understand that for a claim to be valid, its scope must be definite, and that improperly introducing a term can lead to an unascertainable scope of a term, rendering it indefinite. I understand that any claim element first addressed in the claims by the antecedent basis, “a” may then be identified later in the claims by the antecedent basis, “the.” I understand that failing to follow this convention may lead to ambiguity in the identification of the particular claim element in question, leading to an indefinite scope of the claim’s term. For purposes of this proceeding, I identified and provided my best understanding of any such terms despite such ascertainable scope.

V. SUMMARY OF THE ’321 PATENT

A. Specification

39. By my review, the ’321 patent relates to a “system and method for distributed analysis of image data in a video surveillance system can be deployed in households, businesses, and within corporate entities.” *’321 patent*, Abstract. The ’321 patent acknowledges that existing “[v]ideo surveillance systems can increase the volume of data traffic over the networks,” causing “performance

problems.” *Id.*, 1:47-53. To resolve this issue, the “image data are processed on local video analytics systems located within the networks of the businesses or on remote video analytics systems hosted by a cloud service. To limit the data traffic imposed by the image data on the network, the system divides the image data processing into separate object detection and analysis functions.” *Id.*, Abstract. I have annotated and reproduced Figure 1 of the ’321 patent below to show object detection system 108-1 and 108-2 (yellow) connected to video analytics system 122 (purple) through network 60 (red):

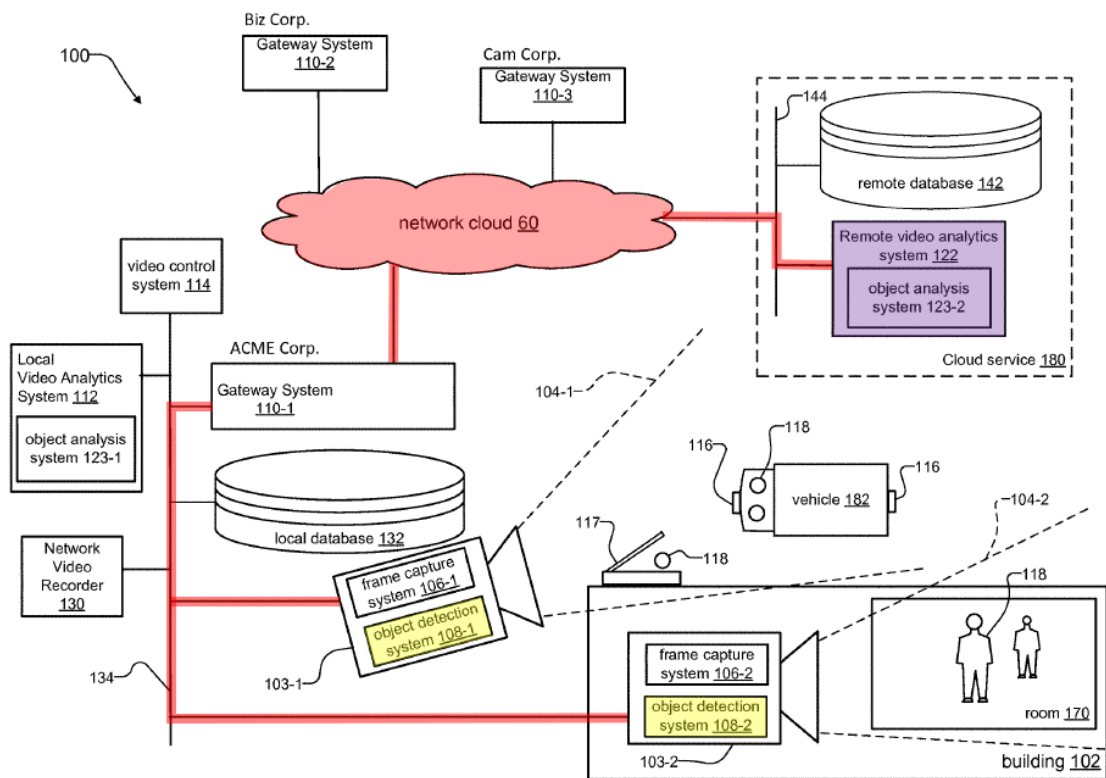


FIG. 1

40. According to the ’321 patent, “[t]he object detection system detects objects of interest” and “only image data that includes the detected objects of interest

needs to be sent over the data network for object analysis by the local or remote video analytics system.” *Id.*, 2:11-20, 2:41-44. This division of image data processing “saves on bandwidth as compared to current video surveillance systems.” *Id.*, 44-47. To further reduce bandwidth, the object detection system can create and send “cropped images from the image data that include only the detected objects of interest” to the video analytics system. *Id.*, 2:48-54.

B. Prosecution History

41. I understand the ’321 patent issued from U.S. Patent Application No. 14/634,428 that was filed on Feb. 27, 2015. The ’321 patent does not claim any earlier priority. For purposes of this declaration, I treat Feb. 27, 2015 as the priority date for the ’321 patent.

42. By my review, I see that the Examiner and applicant exchanged seven Office Actions (non-final and final) and three Requests for Continued Examination over the course of five years. *See ’321Pat File History*, 79-87, 112-21, 147-49, 154-62, 183-93, 223-25, 245-52, 272-81, 311-13, 320-28. The office actions included repeated rejections on all claims because various prior art references independently taught surveillance cameras that detect objects of interest within the image data and transmit the image data over a network. *See id.*, 79-87, 112-21, 154-62, 183-93, 245-52, 272-81, 320-28.

43. In all the office actions, the Examiner noted that U.S. Patent

Application Publication No. 2014/0044348 to Chen et al. (“Chen”) disclosed an object detection system that ranks image data according to the factors listed in as-filed claims 4 and 17¹. (*See id.*, 83-84, 116, 159-60, 187, 249, 276, 324-25.) Specifically, Chen discloses that “the images may be based on a ranking such that a predetermined number of images with the better quality are identified for further processing.” *Chen*, ¶34. The Examiner also consistently noted that Chen disclosed “comparing the image data to reference images to identify the objects of interest.” *Id.* The applicant never contested the Examiner’s rejections based on Chen. *See id.* Rather, in response to the October 11, 2019, office action, applicant amended claims 4 and 17 to recite limitations including “create[ing],” “sav[ing],” and “read[ing]” the ranking to “decide[.]” whether to analyze the image data. *Id.*, 330-31. The Examiner incorporated as-amended claims 4 and 17 into independent claims 1, 13, and 22 through a March 23, 2020, Examiner’s amendment and allowed the application without providing any statement on reasons for allowance. *Id.*, 346-53. The ’321 patent issued on June 30, 2020. ’321 *patent*, Cover.

44. I note that prior art references Laganieri ’363, Laganieri ’817, Laganieri ’416, and Bedros were not cited in the ’321 file history. Nor was the cited

¹ As-filed claims 4 and 17 were substantially incorporated as limitations [1.5.1] and [13.5.1] into the as-issued independent claims 1 and 13.

art related to those references.

VI. TECHNOLOGY BACKGROUND

A. Early digital video capture systems

45. The VuSystem was developed at MIT, in the research group I was a part of, in the early-to-mid 1990's. One of the members of our group designed and built a digital video capture board that connected directly into our networking equipment and provided video to computers on the network making a request for video. *See, e.g., Adam*, p. 1. The VuSystem is a system in which “the computer not only manipulates media,” such as video content, “but also digests it and performs independent actions based on the media content.” *Lindblad*, p 3.² The VuSystem was user-programmable using a visual flow graph that was used to observe and control the pattern of media processing implemented by applications. *Wetherall*, p. 13. In the VuSystem and “[i]ntelligent video programs will be capable of making decisions based on video content. For example, a computer scribe may automatically capture and distill a set of notes from a camera pointed at a blackboard during a lecture [36].

² I have personal knowledge of the teachings of Linbald, Wetherall, and Staisor.

Each of these dissertations and thesis were supervised by my advisor (Prof.

Tennenhouse) and were defended in 1994, 1994, and 1997 while I was a graduate student in Dr. Tennenhouse's lab (1991-1998).

A programmable television agent was capable of reorganizing a video broadcast to meet a user's preferences. Alternatively, a program could distinguish between video-conference participants and their background so that the participants can enhance their images with different background scenes.” *Wetherall*, p. 15. Information within video stream images “[was] identified and transformed into properties.” *Stasior*, p. 17. “The video analysis employed in this thesis focuses on two complementary areas from computer vision: matching and motion. Matching is concerned with determining whether two inputs are the same while motion is concerned with detecting and measuring change. Motion is used primarily to focus the attention of the system, telling it when and where to look. Matching is used to compare input data against models and templates so that a property may represent the degree to which an input matches a particular model. Matching is implemented in a modular fashion whereby representational transformations that expose the relevant information are performed prior to matching.” *Stasior*, p. 17.

B. Cloud Computing

46. Amazon S3, Simple Storage Service, was announced in 2006. “Amazon S3 is storage for the Internet. It is designed to make web-scale computing easier for developers. Amazon S3 provides a simple web services interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the web. It gives any developer access to the same highly scalable, reliable, fast,

inexpensive data storage infrastructure that Amazon uses to run its own global network of web sites.”³ Amazon’s Elastic Compute Cloud (EC2) beta was also announced in 2006. “Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud. It is designed to make web-scale computing easier for developers. Just as Amazon Simple Storage Service (Amazon S3) enables storage in the cloud, Amazon EC2 enables “compute” in the cloud.”⁴

C. Network-based cameras

47. While our group’s network-connected camera was a very early (or even the first ever) video camera connected directly to a data network, commercial network cameras began to be introduced in the late 1990’s. For example, the AXIS NetEye 200 was available by early 1997, and could be “connected directly to

³ <https://web.archive.org/web/20090804042006/http://aws.amazon.com/about-aws/whats-new/2006/03/13/announcing-amazon-s3---simple-storage-service/>

⁴ <https://web.archive.org/web/20090804041927/http://aws.amazon.com/about-aws/whats-new/2006/08/24/announcing-amazon-elastic-compute-cloud-amazon-ec2---beta/>

Ethernet and was managed using a standard browser to take and view snapshots.”⁵ By the mid-to-late 2000’s, cameras with intelligent in-camera processing were available commercially. The Intellio Distributed Intelligence Camera System was available starting in the mid-2000’s, and “enables a large network to be set up thanks to the On-Board intelligence built into the camera itself.”⁶ The Intellio cameras could detect certain events, such as a “Running Person,” “Abandoned Object,” “Stolen Object,” and “Crowd Detection,” among others.⁷ By late 2014, Ring started shipping a doorbell camera that saved video to the cloud.⁸

D. Video compression

48. MPEG-1 is a video compression standard developed and introduced in early 1990’s, followed by MPEG-2. Both MPEG-1 and MPEG-2 leveraged the well-known concept of identifying redundant data in the video so that data would not have

⁵https://web.archive.org/web/19970112064347/http://www.axis.com/products/axis_neteye_200/

⁶https://web.archive.org/web/20091231083345/http://www.intellio.eu/tecnologia_beepitettintelligencia.php

⁷https://web.archive.org/web/20091231083345/http://www.intellio.eu/tecnologia_beepitettintelligencia.php

⁸ <https://web.archive.org/web/20141223144209/https://ring.com/>

to be repeatedly transmitted. Several concepts within the MPEG coders embodied this principle. MPEG uses different types of encoded frames, known as I-frames, P-frames, and B-frames. I-frames, or “Intra-frames” are independently coded frames of video, similar to the way a single photo is encoded using the JPEG image format. I-frames can be decoded into a single frame of video using only the data available in the I-frame itself, just as a photo can be decoded from a single JPEG photo file. In coding and decoding a video frame, the frame of video is broken down into a series of two-dimensional blocks of video, commonly a series of 8x8 pixel blocks that represent the whole video frame. Each of the data blocks in an I-frame is encoded using a Discrete Cosine Transform, that represents the original data block components (luminance and three color components) using frequency-based component values stored together as a “macroblock.”

49. The next type of frame, P-frame, are “Predictive-frames.” These frames are predicted based on blocks from a past frame. This leverages the temporal redundancy in a video, where a frames in the future are likely to have content similar to the current frame. For example, when panning a video camera to the right, only a portion of the frame at the right edge of the frame represents “new” content; most of the next frame is the same as the current frame only shifted to the left slightly. As another example, when a person walks through the field of view of a camera, the background is fixed, and the person moves over different parts of the background as

they walk, while most the background remains the same. Predictive-frames reference information somewhere in a previous frame but encode a “motion vector” that is applied to a previous frame to simply show how far and in what direction a previous frame’s block has moved. This is known as “motion compensation.” Finally, a B-frame is a “Bidirectional-frame” and incorporates predictions based not only on previous frames, but also on future frames. B-frames are used between P-frames or between P-frames and I-frames. A series of frames consisting of one I-frame and a number of B-frames and P-frames is known as a “group of pictures” and one example would be: I B B B P B B B P B B B P B B B. As mathematical errors accumulate throughout the predictive coding process, a new group of pictures with a new independently-encoded I-frame stops the accumulation of predictive errors.

E. State of object recognition/detection

50. Object recognition in images was an active area of research throughout the 2010’s. The “ImageNet” image database was introduced in 2009, to help facilitate the progress of object recognition within images. *See Deng*. The ImageNet data set consisted of over 3,000,000 images that were human-labelled. In the following years from 2010 through 2017, an annual competition “to allow researchers to compare progress in detection across a wider variety of objects” and “measure the progress of computer vision for large scale image indexing for retrieval and annotation” was held using the ImageNet data set. *See* <https://www.image->

net.org/challenges/LSVRC/ (documenting details of each competition held between 2010 and 2017). Various neural network and artificial intelligence-based models emerged during this time, pushed forward by researchers possessing high levels of skill and specialized training. By 2017, a large neural-network based model using 650,000 neurons and 60 million parameters produced a model with a 37.5% error rate based on the top prediction and a 17.0% error rate based on the top five predictions over the ImageNet image database. *See Krizhevsky*, pp. 1, 7.

VII. CLAIM CONSTRUCTION

51. It is my understanding that in order to properly evaluate the '321 patent, the terms of the claims must first be interpreted. It is my understanding that for the purposes of this inter partes review, the claims are to be construed under the so-called *Phillips* standard, under which claim terms are given their ordinary and customary meaning as would have been understood by one of ordinary skill in the art in light of the specification and prosecution history, unless the inventor has set forth a special meaning for a term. For the purposes of this declaration, I have given all the claim terms of the Challenged Claims their plain and ordinary meaning, as would have been understood by a person of ordinary skill in the art, at the time of the alleged invention, which I understand is February 27, 2015 for all claims.

52. Additionally, I have noted that several claim limitations appear to be expressed in means-plus-function form. For example, the “system” limitations in the

claims do not appear to recite sufficient structure in the claims. I have identified means-plus-function limitations and have considered the function and corresponding disclosed structure in forming my opinions where applicable. The corresponding functions are listed in, and evident from, the respective claims. The corresponding structure of each function that I considered are as follows:

- “object detection system”

<u>Claims:</u>	1.2, 13.2.1; 1.5.1, 13.5.1; 1.5.2, 13.5.2; 2, 14; 3, 15
<u>Structure:</u>	Processor with memory in a local analytics system, local camera, or local gateway, <i>e.g.</i> , '321Pat., Figs. 2, 4, 6, 7, 1:64-2:2, 2:50-53, 3:1-4, 3:32-34, 4:45-49, 6:9-11, 6:19-23, 6:29-36, 6:41-43, 7:65-67, 9:3-15, 11:1-8, but lacking disclosure of sufficient algorithms to perform the claimed functions.

- “video analytics system”

<u>Claims:</u>	1.3.1, 13.3; 1.3.2, 13.3; 1.4.1, 13.4; 4, 16; 6
<u>Structure:</u>	Processor with memory in a remote analytics system, remote gateway, or cloud, <i>e.g.</i> , '321Pat., Figs. 2, 4, 6, 7, 1:64-2:2, 2:21-36, 2:59-63, 3:14-22, 3:36-38, 4:50-52, 6:52-7:3, 7:19-41, 9:6-31, 10:24-45, 11:14-30, but lacking disclosure of sufficient algorithms to perform the claimed functions.

- “object analysis system”

<u>Claims:</u>	1.5.3, 13.5.3; 4
<u>Structure:</u>	Processor with memory in a remote analytics system 122, remote gateway, or cloud, <i>e.g.</i> , '321Pat., Figs. 2, 4, 6, 7, 1:64-2:2, 2:21-36, 3:20-22, 6:38-40, 6:52-7:3, 7:19-41, 9:16-37, 10:24-45, 11:14-35, but lacking disclosure of sufficient algorithms to perform the claimed functions.

VIII. CLAIMS 1-21 ARE UNPATENTABLE

53. I have been asked to provide my opinion as to whether the Challenged

Claims of the '321 patent would have been obvious in view of the prior art. The discussion below provides a detailed analysis of how the prior art references identified below teach the limitations of the Challenged Claims of the '321 patent.

54. As part of my analysis, I have considered the scope and content of the prior art and any differences between the alleged invention and the prior art. I describe in detail below the scope and content of the prior art, as well as any differences between the alleged invention and the prior art, on an element-by-element basis for each Challenged Claims of the '321 patent. As described in detail below, the alleged invention of the Challenged Claims would have been obvious in view of the teachings of the identified prior art references as well as the knowledge of a POSITA.

Grounds	Claim(s)	Basis
1A	1-5, 7-21	Laganiere '363, Laganiere '817, Laganiere '461
1B	4, 6, 16	Laganiere '363, Laganiere '817, Laganiere '461, Chen
2	1-21	Bedros

A. Ground 1A: Laganiere '363 Renders Claims 1-5, 7-21 Obvious in view of Laganiere '817 and Laganiere '461

1. Summary of Laganiere '363

55. In my opinion, Laganiere '363 (EX1005) is analogous art to the '321 patent. For example, Laganiere '363 and the '321 patent are in the same field of endeavor. Both are directed to video surveillance systems having video and image

analytics capabilities. *Compare '321 patent*, Abstract, 4:38-44; *with Laganieri '363*, Abstract, ¶25. In addition, both Laganieri '363 and the '321 patent are pertinent to the problem of increased network traffic in video surveillance systems. *Compare '321 patent*, 1:6-63; *with Laganieri '363*, Abstract, ¶¶5-8.

56. Laganieri '363 recognizes the market trend and need for reducing “strain on a network” of existing video surveillance systems. *Laganieri '363*, Abstract, ¶¶2-7. To resolve this issue, Laganieri '363 developed a video analytics system that pre-processes video data “at the source end to reduce the bandwidth requirement for transmitting the video data.” *Id.*

57. Laganieri '363's system includes multiple video sources 102 and 116 (yellow) connected to server 108 (purple) through optional gateway 104, and WAN 106 (red). *Laganieri '363*, Fig. 2, ¶¶29-33; *see also, id.*, Fig. 1, ¶¶25-28. I have annotated and reproduced Figure 2 of Laganieri '363 below:

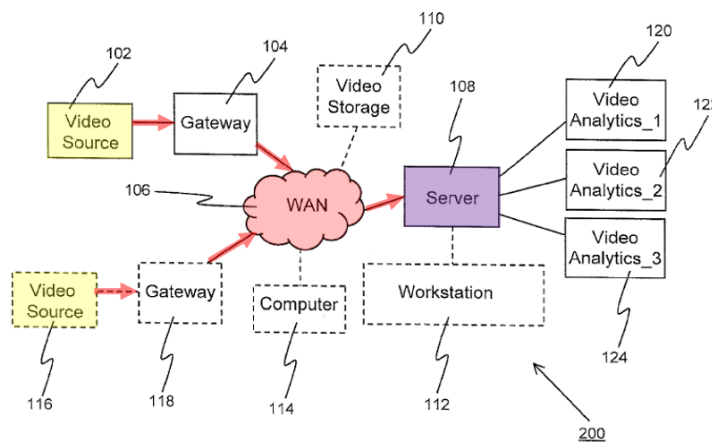


Fig. 2

58. According to Laganieri '363, video data is captured and “pre-

processed... using video analytics” at the video sources to produce a “data stream that requires an amount of bandwidth below a bandwidth limit of WAN 106.” *Laganiere* ’363, ¶34. The pre-processing operation provides several benefits. For example, the pre-processing of images at the source end could “reduc[es] the bandwidth requirement for transmitting the video data.” *Id.*, ¶8. The bandwidth reduction can be achieved through preprocessing data relating to an event of interest, such as license plates or faces of an individual. For example, preprocessing can include “cropping regions of the video data that relate other than the event of interest” such that “an area of each frame outside of a region that contains the identified license plate is cropped” *Id.*, ¶¶21, 30, 41. According to *Laganiere* ’363, a third-party analytics engine can also be used, such as cloud computing and “farm[ing] out” to “third parties if specialized video analytics engines are required.” *Id.*, ¶23.

59. The pre-processing result is also inserted into the video sent for processing: “[o]ptionally, the video source 102 inserts into the transmission of the pre-processed video data at least a pre-processing result, other than the video data itself, relating to a result of pre-processing the video data.” *Id.*, ¶37.

60. Video source 102 transmits the pre-processed data to central server 108 “without exceeding the bandwidth limit of the WAN 106.” *Laganiere* ’363, ¶35. The transmitted “pre-processed data is subject to additional video analytics processing” at central server 108 or at “one of the video analytics engines 120-124.” *Id.*, ¶35.

Figure 5 of Laganier '363 illustrates a “simplified flow diagram” that describes the process of pre-processing data at the video source and sending the pre-processed data to a central location for further processing. Fig. 5 is reproduced below:

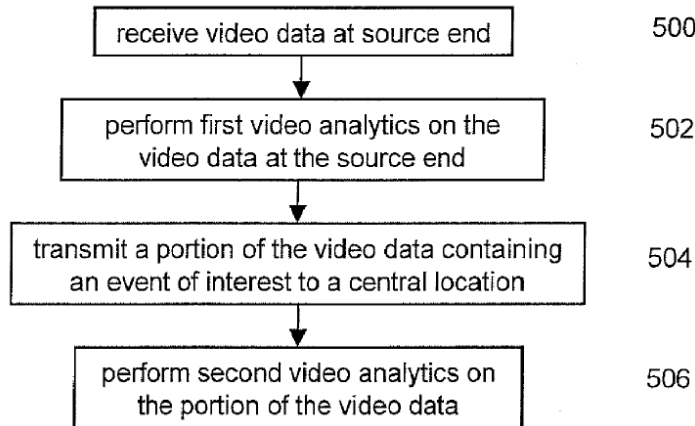


Fig. 5

2. Summary of Laganier '817

61. In my opinion, Laganier '817 is analogous art to the '321 patent. For example, Laganier '817 and the '321 patent are in the same field of endeavor. Both Laganier '817 and the '321 patent are directed to a video surveillance system that can perform facial recognition on the captured images and assess the quality of the captured images. *Compare '321 patent*, 3:8-13, 6:29-36; *with Laganier '817*, Abstract, 2:20-50, 4:20-47.

62. From my review, Laganier '817 shares the same first inventor with Laganier '363. *Compare Laganier '817*, field code (75) (Inventors: Robert

Laganière⁹ (Gatineau (CA) et al.) *with Laganiere '363*, field code (76) (Inventors: Robert Laganiere (Gatineau (CA) et al.).)

63. Laganiere '817 discloses a system, method, and computer-readable medium to create a “face log” from video and/or image data captured by a camera for subsequent facial recognition analysis. *Laganiere '817*, 2:20-50. Reproduced Figure 1 below depicts a system “for face tracking and face log creation.” *Laganiere '817*, 3:55-56.

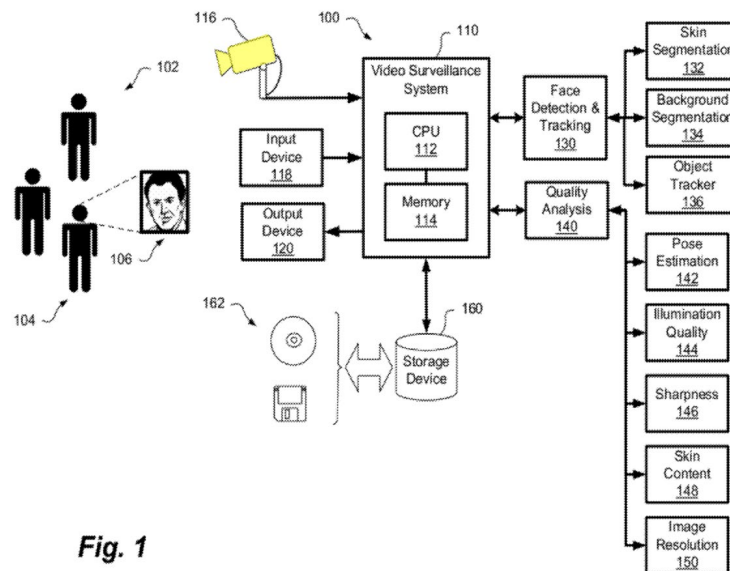


Fig. 1

⁹ The spelling of “Laganière” in Laganiere '817 contains French accent mark (è).

Nonetheless, the first inventors of Laganiere '817 and '363 are the same individual because they share the same first and last names (except for the French accent mark) and the same primary residence (Gatineau, CA). For simplicity, I refer to the US 817 publication as Laganiere '817.

64. Laganier '817 discloses that video cameras 116 (yellow) capture video or still images, and image characteristics are used to select face images, “such as resolution, pose, illumination, skin content and sharpness.” *Id.*, 3:35-38, 55-56. These characteristics are scored and incorporated to create an “overall quality score of an image.” *Id.*, 38-42. The quality scores are then used to create a face log (*id.*, 7:55-10:32), containing only images of an individual of sufficient quality to perform facial recognition (*id.*, 7:48-51, 9:61-63). Laganier '817 also discloses that “potential candidate face image regions must be defined and the region must then be tracked between frames.” *Id.*, 7:57-59. The “highest quality face detected for the individual” is retained. *Id.*, 7:64-66.

3. Summary of Laganier '461

65. In my opinion, Laganier '461 is analogous art to the '321 patent. For example, Laganier '461 and the '321 patent are in the same field of endeavor. Both are directed to video surveillance systems having video and image analytics capabilities. *Compare '321 patent*, Abstract, 4:38-44; *with Laganier '461*, Abstract, ¶¶8-10. In addition, both Laganier '461 and the '321 patent are pertinent to the problem of increased network traffic in video surveillance systems. *Compare '321 patent*, 1:6-63; *with Laganier '363*, Abstract, ¶5.

66. From my review, Laganier '461 shares the same first inventor with Laganier '363 and Laganier '817. *Compare Laganier '461*, field (76) (Inventors

Robert Laganier (Gatineau (CA) et al.) *with Laganier* '817, field code (75) (Inventors: Robert Laganière¹⁰ (Gatineau (CA) et al.) *and with Laganier* '363, field code (76) (Inventors: Robert Laganier (Gatineau (CA) et al.).)

67. I note that the system depicted in Laganier '461 is similar to the system depicted Laganier '363. For example, both Laganier '461 and Laganier '363 include video source 102 connected to server 108 through (optional) gateway 104 and WAN 106. *Compare Laganier* '461, Fig. 2, ¶¶20-22, *with Laganier* '363, Fig. 1, ¶¶25-33.

68. Laganier '461 describes a system “for reducing the amount of space that is required to store video data, based on selective deletion of portions of video data that are stored in a video storage device.” *Laganier* '461, ¶2. Laganier '461 explains that a central server performs the deletion of images. *Id.*, ¶¶24-25.

¹⁰ As discussed, the spelling of “Laganière” in Laganier '817 contains French accent mark (è). Nonetheless, the first inventors of Laganier '817 and '363 are the same individual because they share the same first and last names (except for the French accent mark) and the same primary residence (Gatineau, CA). For simplicity, I refer to the US 817 publication as Laganier '817.

4. Independent Claim 1

a) Claim 1.pre: “A video surveillance system, comprising:”

69. To the extent the preamble is considered limiting, in my opinion, Laganriere '363 discloses or suggests the preamble. Laganriere '363 seeks to overcome limitations with existing video surveillance systems. *Laganriere '363*, ¶¶2-7. Laganriere '363 discloses systems with video sources 102 “for monitoring a known field of view (FOV)” such as “a parking lot, an entry/exit point of a building, and an automated teller machine (ATM).” *Id.*, ¶25; *see also* Figs. 1, 2. The systems also include video analytics engines 120-124 that can detect and read license plates and “perform[] facial recognition.” *Id.*, ¶30. Laganriere '363 suggests these and other video analytics engines are needed by “different monitoring and surveillance applications.” *Id.* Accordingly, a POSITA would have understood Laganriere '363 to disclose video surveillance systems. To that end, Laganriere '363 discloses “systems” that include a “video source 102 [] in communication with a central server 108 via gateway 104 and Wide Area Network (WAN) 106.” *Id.*, ¶25. Figure 1 (below) depicts Laganriere '363's system, provided here:

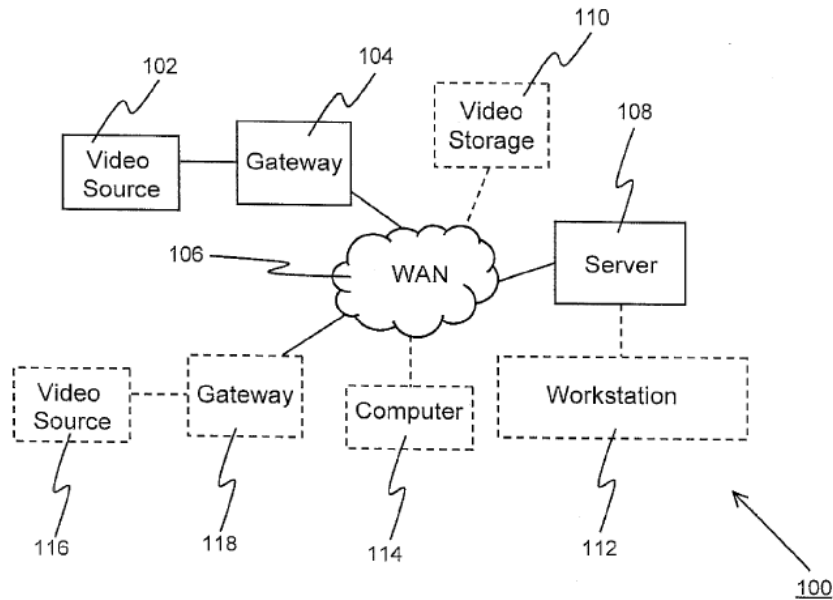


Fig. 1

70. In my opinion, a POSITA would have recognized that Laganieri '363's system is used for "*video surveillance*" because video source 102 "is deployed at a source end for monitoring a known field of view (FOV)" such as "a parking lot, an entry/exit point of a building, and an automated teller machine (ATM)." *Id.* In addition, central server 108 can be in further communication with "video analytics engines 120-124, which are indicated in FIG. 2." *Laganieri '363*, ¶29. Fig. 2 is provided below:

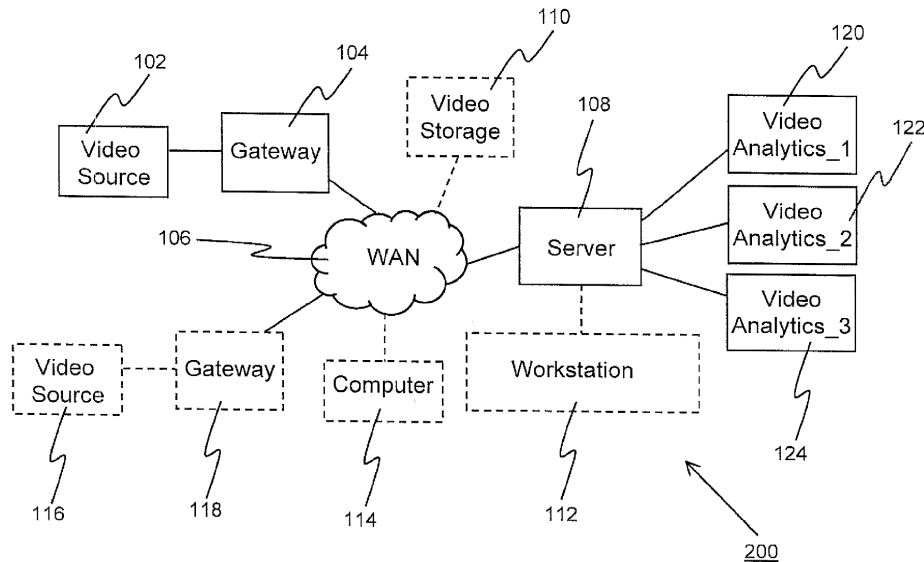


Fig. 2

71. Accordingly, it is my opinion that Laganieri '363 discloses or suggests a system including a video source 102, gateway 104, WAN 106, central server 108, and optionally, video analytics engines 120-124 (“video surveillance system”).

b) Claim 1.1: “one or more surveillance cameras that capture image data”

72. In my opinion, Laganieri '363 discloses or suggests this limitation. As illustrated in Figures 1 and 2 of Laganieri '363, video source 102/116 “deployed at the source end” is a “network IP camera with onboard video analytics capabilities,” or alternatively, “is a ‘dumb’ IP camera or an analogue video camera.” *Laganieri '363*, ¶¶25, 28, 33. A POSITA would have understood that Laganieri '363’s video camera is a “surveillance camera” because it monitors parking lots, building entries, or ATMs and captures “video data relating to an event of interest.” *Id.*, ¶¶25, 34. As shown in Figure 2 of Laganieri '363, multiple cameras

such as video sources 102 and 116 (“one or more surveillance cameras”) are included in the system. *See Laganieri* ’363, ¶¶28, 33, Fig. 2; *see also* Fig. 1. Figure 2 of *Laganieri* ’363 is reproduced below to show video sources 102 and 116 in yellow.

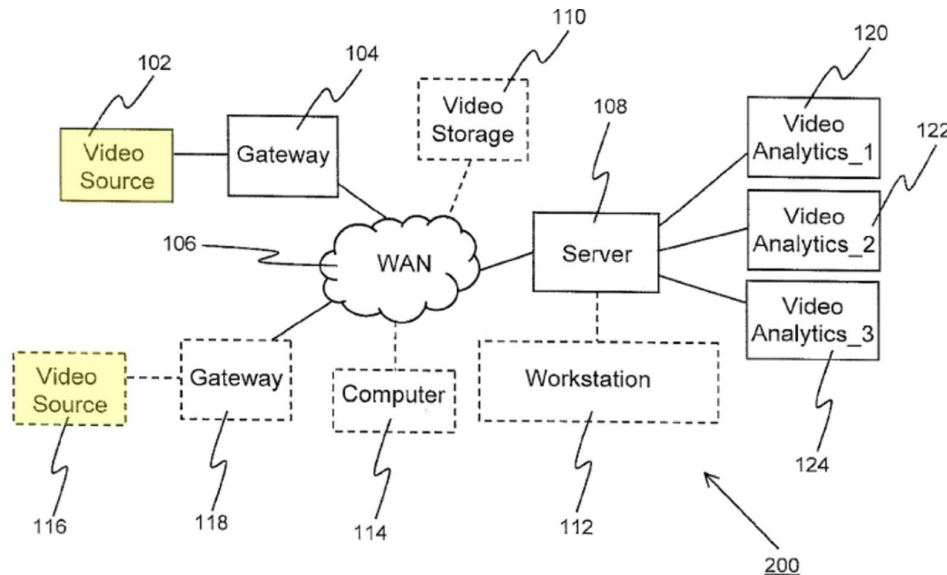


Fig. 2

73. *Laganieri* ’363 describes how a “video camera captures video data at a known frame rate, typically 30 FPS.” *Laganieri*, ’363, ¶34. *Laganieri* describes analyzing the video frames to read license plates, detect faces, etc. *Id.*, ¶¶30, 41 (discussing selection of “a frame with a suitable license plate **image**”). A POSITA would have understood that these video frames captured by the video camera are image data.

74. I note that *Laganieri* ’363 also explains that “video data that is provided from the video source 102.” *Id.*, ¶25. The video data consists of a series of frames. *Laganieri* ’363, ¶41 (“frames of the video data”). A POSITA would have thus

understood the captured video data includes frames that include “*image data*,” as recited in 1.1 because the ’321 patent explains that “images and/or video, [are] also known as image data.” ’321 patent, 5:29-34. Furthermore, Laganieri ’363’s video data includes frames in the same manner as “video data” in the ’321 patent. For example, Laganieri ’363’s “video camera captures video data at a known frame rate, typically 30 FPS.”¹¹ *Laganieri ’363*, ¶34. In the same way, the “frame capture system 106-1 of the surveillance camera 103-1 [in the ’321 patent] captures image data” and “encodes the image data into **frames of video data**.” ’321 patent, 5:60-64, 8:60-64, 10:7-11. Moreover, the ’321 patent acknowledged that “[a] number of components are common to many of these video surveillance systems. **Surveillance cameras capture image data**.” ’321 patent, 1-16-19.

75. Accordingly, it is my opinion that Laganieri ’363 discloses or suggests video source 102 (i.e., cameras) capturing video data that consists of a series of frames (“one or more surveillance cameras that capture image data”).

c) Claim 1.2: “an object detection system that detects objects of interest within the image data and extracts portions of frames that include the detected objects of interest and sends the extracted portions of the frames”

76. In my opinion, Laganieri ’363 discloses or suggests this limitation. Laganieri ’363 discloses an object detection system at the source end.

¹¹ “Frames per second” is abbreviated “FPS.” *Reverse Acronyms* (EX1013), p. 819.

Laganiere '363's Figure 5 shows a method of analyzing video data using the "systems shown in Fig. 1 and Fig. 2." *Laganiere '363*, ¶34. I have annotated and reproduced Figure 5 of Laganiere '363 below.

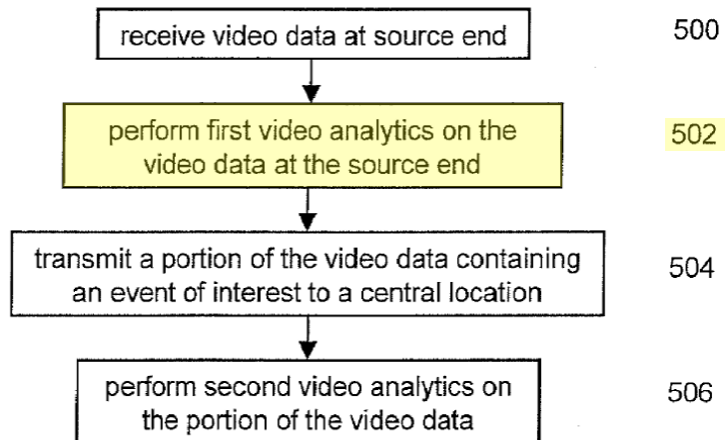


Fig. 5

77. After receiving video data, "[a]t 502 first video analytics is performed on the video data using a first processor disposed at the source end, **the first video analytics for detecting the event of interest in a portion of the video data.**" *Id.*, ¶40. The event of interest can be "a vehicle license plate within video data captured using a video camera that monitors a parking lot entrance." *Id.*, ¶41. Other events can be counting pedestrians, or identifying people or assets. *Id.*, ¶19.

78. Accordingly, it is my opinion that Laganiere '363 discloses or suggests a processor executing software ("an object detection system") disposed at the source end and detects an event of interest ("detects objects of interest within the image data").

79. As an example, “the first video analytics located a vehicle license plate within video data captured using a video camera that monitors a parking lot entrance.” *Id.*, ¶41. The license plate is “an object of interest within the image data” captured by the surveillance camera.

80. Laganier '363 further teaches “extract[ing] portions of frames that include the detected objects of interest” and “send[ing] the extracted portions of the frames.” For example, after a vehicle license plate is detected, “an area of each frame outside of a region that contains the identified license plate is cropped, such that only the region of video data that contains the identified license plate is transmitted to the central location via WAN 106.” *Laganier '363*, ¶41. Cropping regions of the video data is an example of “pre-processing.” *See id.*, ¶21. As Laganier '363 notes, performing pre-processing on video data at the video source (e.g., cameras) provides the benefit of “reduc[ing] the bandwidth requirement for transmitting the video data to below a bandwidth limit of a Wide Area Network (WAN) over which the video data is to be transmitted.” *Id.*, Abstract, ¶25 I also note that Laganier '363 explains that video sources are at the “source end.” *Id.* The transmitting process is performed at operation 504 (blue) of Figure 5, which I annotated and reproduced below. *See Laganier '363*, ¶40.

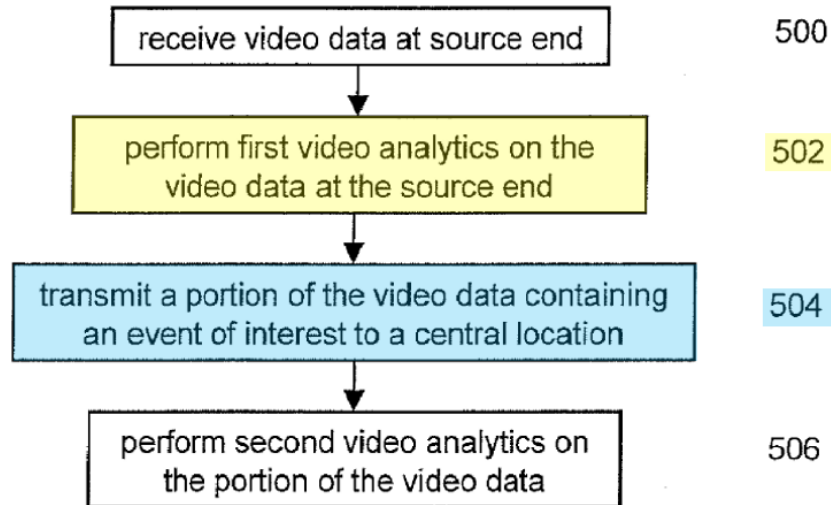


Fig. 5

81. I also note that Laganier '363's cropping and transmitting processes operate in the same manner as the processes of creating and sending "cropped images" described in the '321 patent and for reaping the same benefit. *Compare '321 patent*, 1:47-53, *with Laganier '363*, ¶21. Using the license plate example above, Laganier '363's cropping process removes the area outside the identified license plate, leaving only the region of the frame that contains the identified license plate to be transmitted. *See Laganier '363*, ¶41. This pre-processing (e.g., cropping) reduces the bandwidth requirement of sending image and/or video data from the source end (i.e., video source such as a camera). *Id.*, Abstract, ¶¶8, 34, 38, 40-41. Just as in Laganier '363, the "object detection system 108-1 [of the '321 patent] can create cropped frames" that "include the detected objects of interest." '321 patent, 9:3-5. The '321 patent states that sending "cropped images to the video

analytics system [] further reduce[s] bandwidth.” *Id.*, 3:1-7.

82. Laganieri ’363 discloses or suggests the same or equivalent structure to that described by the ’321 patent for performing the claimed functions. Because Laganieri ’363 discloses that the object detection, extraction, and transmission steps are performed at the source end, it would have been obvious for the functions to be performed (1) by an onboard video analytics in the camera itself or (2) by a “local video analytics engine” which a POSITA would have regarded as equivalent to the ’321 patent’s disclosures of (1) embedding the object detection system (108-1/2) within the camera itself (103-1/2) or (2) using a local video analytics system 112, respectively. A POSITA would have understood that either approach would have required firmware or software programming to perform these functions.

83. Accordingly, it is my opinion that Laganieri ’363 discloses or suggests a processor executing software disposed at a source end (“object detection system”) that detects an event of interest in a portion of the video data (“objects of interest within the image data”) and crops (“extracts portions of frames that include the detected objects of interest”) and transmits the cropped portion of data (“sends the extracted portions of the frames”).

d) Claim 1.3.1: “a video analytics system that receives only the extracted portions of the frames of the image data, transmitted over a network, from the object detection system, the image data received by the video analytics system corresponding to objects of interest detected by the object detection system”

84. In my opinion, Laganieri '363 discloses or suggests this limitation. As I explained in §VIII.A.4.c, Laganieri '363 teaches that in operation 504 **“only the region of video data** that contains the identified license plate **is transmitted to the central location** via WAN 106.” *Laganieri '363*, ¶¶10, 41. The video data is transmitted to the central location from a source end (e.g., video source 102 and/or 116). *Id.*, ¶¶25-29, 40-41.)

85. Laganieri '363 explains that a processor at a central location (“video analytics system”) receives the region of video data containing the identified license plate (“image data... corresponding to objects of interest detected by the object detection system”) from the source end (e.g., video source 102/116) (“object detection system”). *Id.*, ¶¶40-41. The transmission of the region of video data is performed over a Wide Area Network (WAN) 106 (“network”). *Laganieri '363*, ¶¶25, 40-41.

86. The method depicted in Figure 5 is “described with reference to the systems shown in FIG. 1 and in FIG. 2.” *Laganieri '363*, ¶40. The central server 108 depicted in Figures 1-2 receives “pre-processed video data... via WAN” and the region of video data containing the identified license plate is an example of pre-

processed video data. *Id.*, Abstract, ¶¶21, 40-41. Indeed, “[t]he video source 102 is in communication with a central server 108 via gateway 104 and Wide Area Network (WAN) 106.” *Laganiere* ’363, ¶¶25, 28. I annotated and reproduced Figure 1 of *Laganiere* ’363 below to show transmission of video data from video sources 102/116 through WAN 106 (red) to central server 108 (purple).

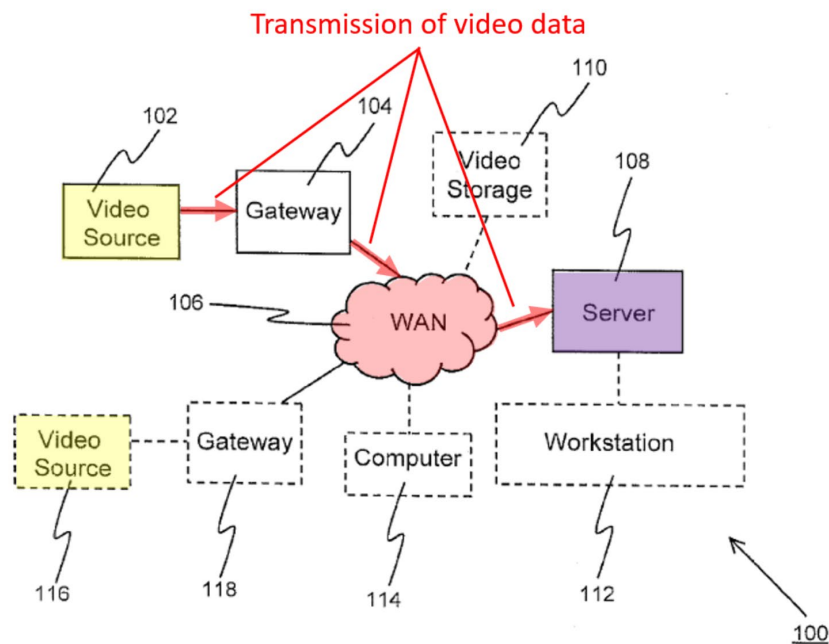


Fig. 1

87. Figure 2 of Laganier '363 shows that “[c]entral server 108 is further in communication with a plurality of video analytics engines 120-124.” *Id.*, ¶¶29-30. The group of video analytics engines receive video data from video source 102 and/or 116 through WAN 106. Specifically, video data is first transmitted from video source 102 and/or 116 through WAN 106 to central server 108, which then “provides the video data to the plurality of video analytics engines 120-124” through various

connection mechanisms such as Local Area Network (LAN), WAN, wireless communication channels, or a direct connection. *Id.*, ¶29-30. I annotated and reproduced Figure 2 of Laganier '363 to show transmission of video data from video sources 102/116 through WAN 106 (red) to central server 108 (purple) and to video analytics engines 120-124 (orange).

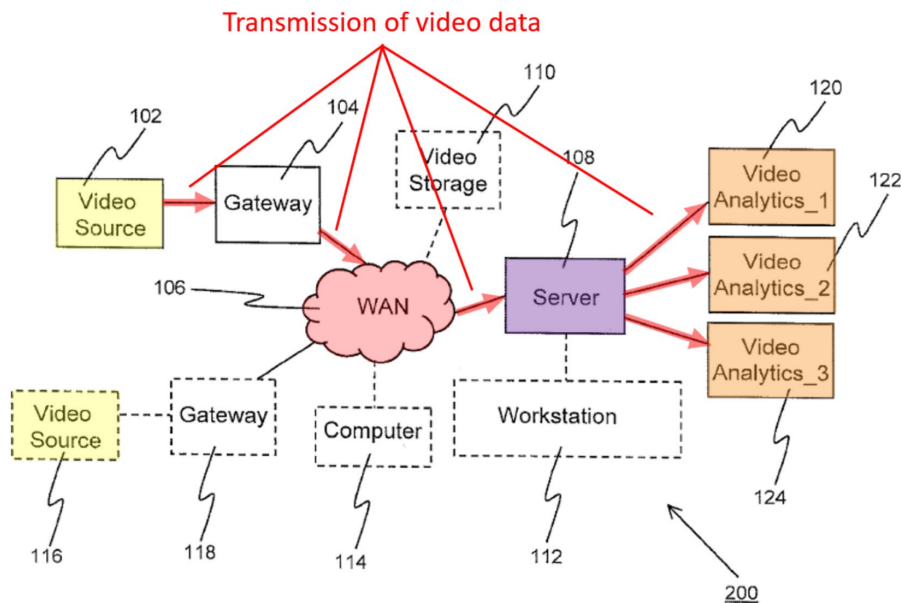


Fig. 2

88. Laganier '363 discloses or suggests the same or equivalent structure to that described by the '321 patent for performing the claimed functions. Laganier '363 teaches that the video analytics system that receives the extracted portions of the image data may be central server 108 which is equivalent to the '321 patent's disclosure that the video analytics system can be a node on an enterprise network. Laganier '363 thus teaches the same or equivalent structure to that described by the '321 patent for performing the claimed functions. Laganier '363

also discloses that “second video analytics” performs receiving and analyzing the image data at a central location (central server 108) alone or with video analytics engines 120-124 using a second processor. For example, Laganieri ’363 discloses “the central server 108 provides the video data to the plurality of video analytics engines 120-124 over a Local Area Network (LAN).” Laganieri ’363, ¶29; *see also*, *id.*, ¶¶10, 28, 30, 40. Alternatively, video analytics engines 120-124 can receive the video data “over one or more of: a WAN; a wireless communication channel such as a cellular telephone network, Bluetooth, a Wi-Fi network, etc.; or a direct connection such as a fiber optic cable or coaxial cable, etc.” *Id.* A POSITA would have understood that these systems would have required associated programming for receiving the image data.

89. Accordingly, it is my opinion that Laganieri ’363 discloses or suggests this limitation.

e) Claim 1.3.2: “the video analytics system analyzing the objects of interest”

90. In my opinion, Laganieri ’363 discloses or suggests this limitation. At operation 506 (below in purple), a “second video analytics is performed on the portion of the video data using the second processor at the central location, the second video analytics for determining information relating to the event of interest.” Laganieri ’363, ¶40.

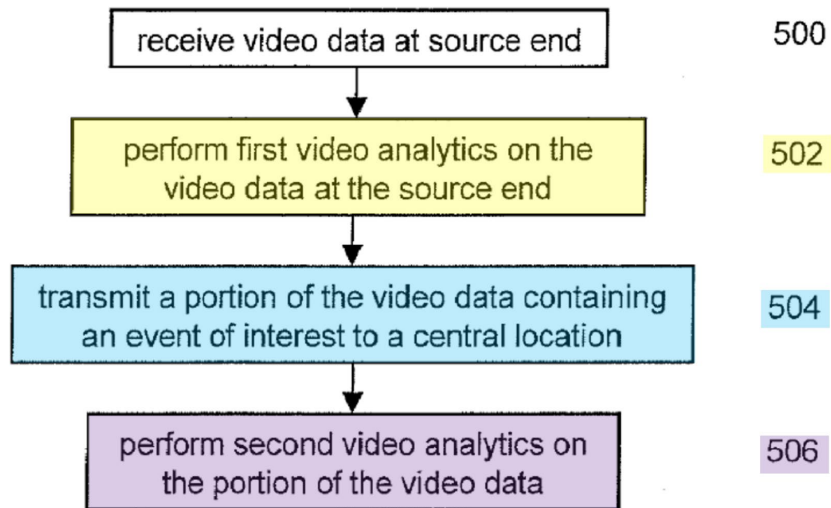


Fig. 5

91. As I explained in §VIII.A.4.d, central server 108 (“video analytics system”) receives from video sources 102 and 116 cropped video data containing an event of interest. *Id.*, ¶¶40-41. Laganier ’363 discloses a central server 108 performs a “**second video analytics** [] for determining the license plate number” after it receives “only the region of video data that contains the identified license plate” (“analyzing the objects of interest”). *Laganier* ’363, ¶¶40-41. Central server 108 may further provide the video data to video analytics engines 120-124, performing various analysis functions such as detecting and reading a license plate and performing facial recognition. *Laganier* ’363, ¶¶29-30. In such case, both central server 108 and video analytics engines 120-124 would form the “video analytics system.”

92. Laganier ’363 discloses or suggests the same or equivalent structure

to that described by the '321 patent for performing the claimed function. Laganieri '363 suggests that the video analytics system that analyzes the objects of interest may be central server 108 which utilizes video analytics engines which is equivalent to the '321 patent's disclosure that the video analytics system can be a node on an enterprise network. A POSITA would have understood that the systems would have required associated programming for analyzing the objects of interest (e.g., reading license plate number, performing facial recognition).

93. Accordingly, it is my opinion that Laganieri '363 discloses or suggests this limitation. Specifically, Laganieri '363 discloses central server 108 alone or with video analytics engines 120-124 ("video analytics system") receiving only the region of video data that contains the identified license plate ("only the extracted portions of the frames of the image data" that "correspond[] to objects of interest detected by the object detection system") over WAN 106 ("transmitted over a network"). It is my opinion that Laganieri '363 further taught that video analytics engines 120-124 ("video analytics system") perform video analytics processes such as reading license plate characters ("analyzing the objects of interest").

f) Claim 1.4.1: "wherein the video analytics system receives the image data from multiple object detection systems after being transmitted over the network"

94. In my opinion, Laganieri '363 discloses or suggests this limitation. Laganieri '363 teaches that video sources 102 and 116 can each be associated with

different clients and both “transmit video data over WAN 106 to a shared central processing facility, e.g. central server 108, which is capable of performing different video analytics processing according to the individual needs of each client.” *Laganiere* ’363, ¶28.

95. Accordingly, it is my opinion that *Laganiere* ’363 taught or suggested central server 108 (“video analytics system”) receiving only the region of video data that contains the identified license plate (“image data”) ¹² over WAN 106 (“transmitted over the network”).

g) Claim 1.4.2: “the video analytics system is operated by a different business entity than the one or more surveillance cameras”

96. In my opinion, *Laganiere* ’363 discloses or suggests this limitation. For example, *Laganiere* ’363 explains that “some video analytics processing may be **‘farmed out’ to third parties** if specialized video analytics engines are required” (*Laganiere* ’363, ¶23) and that “some or all of the video analytics engines 120-124 are fee-per-use-based or subscription based” (*id.*, ¶30). A POSITA would have thus understood that central server 108 and/or video analytics engines 120-124 are “operated by a different business entity” because they are operated by a “third party.”

97. I note that *Laganiere* ’363 further describes various benefits of “moving

¹² I understand that “image data” to refer to “the extracted portions of the frames of the image data.”

the processing infrastructure away from the client's local network and 'into the cloud,'" such as lower costs, more flexibility, and parallel processing. *Laganiere* '363, ¶¶23-24. In fact, *Laganiere* '363's methods of pre-processing video data are particularly useful "when subsequent video analytics processing is performed 'in the cloud.'" *Laganiere* '363, ¶24.

98. It was well-known before the application for the '321 patent was filed that cloud services and local surveillance cameras are operated by different entities. For example, Amazon S3, Simple Storage Service, was announced in 2006. "Amazon S3 is storage for the Internet. It is designed to make web-scale computing easier for developers. Amazon S3 provides a simple web services interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the web. It gives any developer access to the same highly scalable, reliable, fast, inexpensive data storage infrastructure that Amazon uses to run its own global network of web sites."¹³ Amazon's Elastic Compute Cloud (EC2) beta was also announced in 2006. "Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud. It is designed to make web-scale computing easier for developers. Just as Amazon Simple Storage Service

¹³ <https://web.archive.org/web/20090804042006/http://aws.amazon.com/about-aws/whats-new/2006/03/13/announcing-amazon-s3---simple-storage-service/>

(Amazon S3) enables storage in the cloud, Amazon EC2 enables “compute” in the cloud.”¹⁴ It would also have been evident to a POSITA that central server 108 and/or video analytics engines 120-124 are “operated by a different business entity” when payment is required for their video analytics services. *Laganiere* ’363, ¶¶23-24 (cloud service “is sold on demand”); *see also id.*, ¶30 (“the video analytics engines 120-124 are fee-per-use-based or subscription based”). Accordingly, it is my opinion that *Laganiere* ’363 discloses or suggests this limitation.

h) Claim 1.5.1: “wherein the image data corresponding to objects of interest detected by the object detection system are ranked by the object detection system according to factors including close proximity of the objects of interest to the one or more surveillance cameras, picture quality of the objects of interest, and/or contrast of the objects of interest to create a ranking and”

99. In my opinion, the *Laganiere* ’363 in combination with *Laganiere* ’817 discloses or suggests this limitation.

100. As I explained in §VIII.A.4.c, *Laganiere* ’363 discloses performing pre-processing at the source end to reduce bandwidth requirements of sending detected objects to the central server for further analysis. *See Laganiere* ’363, ¶¶34, 38. The pre-processed result can be included in the transmission to central server 108 in

¹⁴ <https://web.archive.org/web/20090804041927/http://aws.amazon.com/about-aws/whats-new/2006/08/24/announcing-amazon-elastic-compute-cloud-amazon-ec2---beta/>

addition to the video data. This result is used by the central server to further “process[]the pre-processed video data based on the at least a pre-processing result.” *Laganiere* ’363, ¶37. The result can also be “in the form of metadata” and the video data is then sent to the server with the metadata. *Id.* Central server 108 reads the metadata and “perform[s] the additional video analytics processing of the pre-processed data.” *Id.*

101. I note that *Laganiere* ’363 provides some examples of pre-processing to reduce bandwidth, such as reducing the resolution of frames, it also does not limit the pre-processing to any particular method. *See Laganiere* ’363, ¶¶38-39. For example, *Laganiere* ’363 also discloses transmitting “only one or a few frames having a suitable quality relative to the other frames” and later discarding some of the sent frames to reduce the bandwidth of transmitted data. *Id.*, ¶41. In the license plate example, *Laganiere* ’363 discloses reducing bandwidth by selecting particular frames “with a suitable license plate image” and transmitting only the “one or a few frames having a suitable quality relative to the other frames.” *Id.* The process for identifying license plates would have been equally applicable to *Laganiere* ’363’s facial recognition aspect by selecting a suitable processing engine from video analytics engines 120-124. For example, “Video Analytics_2 122 detects the vehicle license plate within a data frame” and “Video Analytics_3 124 performs facial recognition.” *Id.*, ¶30.

102. Accordingly, Laganier '363 discloses pre-processing video data at the source end by selecting only high-quality data for subsequent processing at the central computer to reduce the bandwidth of the transmitted data. Laganier '363 further discloses including metadata about the pre-processing along with the video data to be processed to inform the central server how to handle the data. Laganier '363 does not expressly disclose how the data is selected or how the central server uses the metadata. A POSITA would have thus looked to other references to implement Laganier '363's method of reducing bandwidth. In particular, a POSITA would have looked to other disclosure by Laganier, such as Laganier '817, for specifics on implementing the transmission of reduced number of high-quality data.

103. Laganier '817 discloses a system, method, and computer-readable medium to create a "face log" from video and/or image data captured by a camera for subsequent facial recognition analysis. *Laganier '817*, 2:20-50. Laganier '817 explains that concise face logs "ensure adequate images are provided to identify an individual yet not unnecessarily consume processing or storage resources by providing excessive numbers of images." *Id.*, 2:51-55. Moreover, Laganier '817 explains that the "face log need not contain every instance of every face as it only takes one good quality image to identify an individual." *Id.*, 2:55-57. Accordingly, "to reduce the number of faces added to a face log...only high quality image(s) for

each individual [are] stored reducing capacity requirements and operator intervention or judgment.” *Id.*, 2:59-63.

104. I note that Laganieré ’817 also explains that several criteria are used to select face images from the video data, “include image characteristics such as resolution, pose, illumination, skin content and sharpness.” *Id.*, 3:35-38. These characteristics (“factors”) are scored (“ranked”) and incorporated into and used to create an “overall quality score of an image” (“to create a ranking”). *Id.*, 38-42. A POSITA would have understood that resolution, illumination, and sharpness all relate to picture quality of the face image. For example, Chen discloses “[a]s a result, the quality of multiple images of the same person may vary widely. This includes variations in (1) image specific quality, such as resolution, sharpness, contrast, compression artefacts and (2) face specific quality such as face geometry, pose, detectable eye and illumination angles.”) *Chen*, ¶5; *see also*, Laganieré ’817, 6:7-9. Once these scores are calculated, they are combined to create a single quality score. *Id.*, 4:52-7:9 (methods of determining scores S1-6). Laganieré ’817 explains that the individual scores “are combined according to the following weighted sum.” *Id.*, 7:13-16. “In this way, an image’s score primarily reflects the number of criteria satisfied, while the original weighted sum is used to break ties between images satisfying an equal number of criteria.” *Id.*, 7:51-54.

105. Laganieré ’817 explains that the quality score (“ranking”) is used to

create a face log, as detailed in Figures 7-10. *Id.*, 7:55-10:32. The face log contains only images of sufficient quality to perform facial recognition because images below a certain quality threshold are deleted from the log. *Id.*, 7:48-51, 9:61-63.

Figure 11 shows an example of the face log. *Id.*, 10:354-35, Fig. 11.

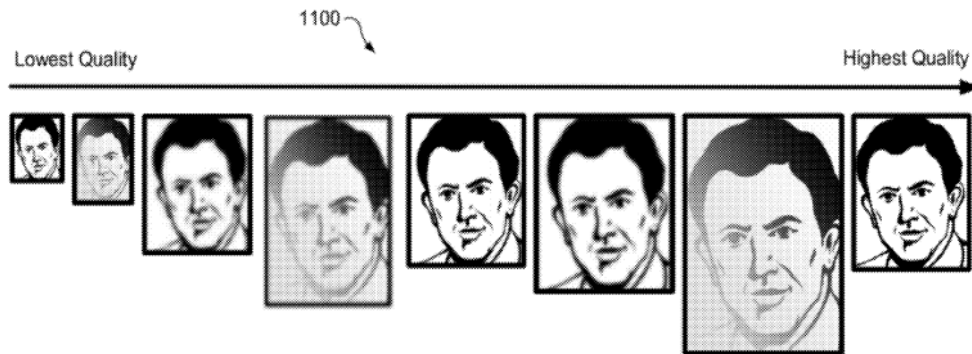


Fig. 11

106. Each face image in the face log is stored with “metadata with the quality score associated with the images.” *Id.*, 8:51-54. Laganieri ’817 further explains that “[o]nce the face log is created and is associated with an individual further purging or reduction of the face log can be performed by removing lower quality images as required therefore only keeping the best quality images.” *Id.*, 10:50-54.

107. As I explained above, Laganieri ’363 discloses detecting objects in image and video data (e.g., faces) and sending only high-quality image data to the central server for further processing. *Laganieri* ’363, ¶41. Furthermore, Laganieri ’363 discloses performing facial recognition on image data sent to the central server which suggests that the objects detected and sent to the central server

could be faces. *Laganiere* '363, ¶30. A POSITA would have understood that the face log in *Laganiere* '817 was a suitable method of accomplishing *Laganiere* '363's goals of reducing bandwidth by transmitting only high-quality object images for further processing.

108. A POSITA would have been motivated to use the *Laganiere* '817's face log in *Laganiere* '363 to reduce the amount of data sent to *Laganiere* '363's central server and to ensure that only high-quality images are sent for subsequent facial recognition. *Laganiere* '363, ¶41. As I explained in §VIII.A.2, *Laganiere* '817 and *Laganiere* '363 share inventors, Applicants, content, and field of use. Moreover, they also both seek to reduce the amount of data transmitted over networks without reducing the ability to perform facial recognition. *Laganiere* '363, ¶¶34, 38; *Laganiere* '817, 2:51-63. Thus, a POSITA would have understood that *Laganiere* '363 and *Laganiere* '817 were combinable and would have been motivated to combine them.

109. A POSITA would have also been motivated to combine them because this implementation would have required nothing more than the application of known, conventional techniques (using image quality to create a group of images to further process) to yield predictable results (reducing bandwidth of transmitted data while ensuring adequate data for subsequent facial recognition) as described in *Laganiere* '363 and *Laganiere* '817. *Laganiere* '363, ¶¶34, 38, 41; *Laganiere* '817,

2:51-63.

110. Accordingly, it is my opinion that the Laganieri '363-Laganieri '817 combination discloses or suggests images containing faces (“the image data¹⁵ corresponding to objects of interest detected by the object detection system”) are scored (“are ranked by the object detection system”) according to image characteristics such as resolution, pose, illumination, skin content and sharpness (“according to factors including close proximity of the objects of interest to the one or more surveillance cameras, picture quality of the objects of interest, and/or contrast of the objects of interest”) to create an overall quality score of an image (“to create a ranking”).

i) Claim 1.5.2: “the ranking is saved in metadata with the image data and”

111. In my opinion, the Laganieri '363-Laganieri '817 combination discloses or suggests 1.5.2. As explained in §VIII.A.4.h, Laganieri '817 discloses that each face image in the face log is stored with “metadata with the quality score associated with the images.” *Laganieri '817*, 8:51-54.

112. Accordingly, it is my opinion that the Laganieri '363-Laganieri '817 combination discloses or suggests that the metadata with quality scores associated

¹⁵ I understand that “image data” to refer to “the extracted portions of the frames of the image data.”

with the images (“the ranking saved in metadata”) is stored with the face images (“image data”)¹⁶.

j) Claim 1.5.3: “an object analysis system of the video analytics system reads the rankings within the metadata and decides whether or not to analyze the image data.”

113. In my opinion, the Laganier ³⁶³-Laganier ⁸¹⁷ combination discloses or suggests 1.5.3. As explained in §§VIII.A.4.h-VIII.A.4.i, the quality scores associated with the images (“the ranking”) is stored with the face images (“image data”).

114. Following transmission of the image data (i.e., face log) Laganier ³⁶³ explains that the central server reads the data and performs operations on it based on the metadata. Laganier ³⁶³, ¶37. Laganier ⁸¹⁷ discloses that higher quality images are better for performing facial recognition. Laganier ⁸¹⁷, 1:24-26; *see also Laganier* ³⁶³, ¶41 (discussing selection of “suitable” license plate images for downstream reading). It would have been obvious to a POSITA combining the teachings of these references for the central server to start with the highest quality images when performing facial recognition. In so doing, the central server would read the rankings within the metadata to decide whether or not to analyze the image

¹⁶ I understand that “image data” to refer to “the extracted portions of the frames of the image data.”

data using facial recognition. Moreover, because the central server performs the functions, Laganieri '363 discloses or suggests the same or equivalent structure to that described by the '321 patent for performing the claimed functions, as I explained above in paragraph 88.

115. Additionally, Laganieri '817 discloses further purging lower quality images from the face log (or license plate). *Laganieri '817*, 10:50-54. By purging lower quality face images from the face log, the central server would be deciding not to analyze those images based (at least in part) on their quality rankings because they would not have been analyzed to perform the recognition. Accordingly, some images in the face log are not ultimately processed at the central server (“decides whether or not to analyze the image data”).

116. To the extent that PO argues that the Laganieri '363-Laganieri '817 combination does not expressly disclose that the central server purges images from the face log, a POSITA would have found it obvious to reduce the number of images in the central server because in a related application filed on the same day as Laganieri '363, Laganieri explains for a similar system “the central server controls a process for selectively erasing portions of video data that are stored in the video storage device, based on a result of video analytics processing of the video data.” *Laganieri '461*, ¶10.

117. As I explained in §VIII.A.3, Laganieri '461 shares inventorship with

both Laganier '363 and Laganier '817. Laganier '461 describes a system similar to Laganier '363 “for reducing the amount of space that is required to store video data, based on selective deletion of portions of video data that are stored in a video storage device.” *Laganier '461*, ¶2; compare *Laganier '461*, Fig. 2 with, *Laganier '363*, Fig. 1. “The selective deletion is performed under the control of a central server based on the results of video analytics processing of the video data.” *Laganier '461*, ¶2.

118. Laganier '461 explains that “under control of the central server 108, video analytics is performed on the portion of the video data, for identifying a subset of the portion of the video data that contains predetermined information relating to the event of interest.” *Id.*, ¶25; see also *id.*, ¶24. Laganier '461 further explains “under control of the central server 108, portions of the retrievably stored portion of the video data, which are other than the identified subset of the portion of the video data, are deleted from the memory storage device 118.” *Id.*, ¶25.

119. A POSITA would have been motivated to use Laganier '461's processing at the central server to further reduce the amount of data the central server must process, thereby speeding up the process of facial recognition. Moreover, each of the Laganier references seeks to reduce the amount of data required for the central server to process. *Laganier '461*, ¶28; *Laganier '363*, ¶¶34, 38; *Laganier '817*, 2:51-63. A POSITA would have also been motivated to combine

Laganiere '461 with Laganiere '363 and Laganiere '817 because this implementation would have required nothing more than the application of known, conventional techniques (reducing data to be processed at a server based on metadata indicating quality) to yield predictable results (processing less data) as described in Laganiere '461, Laganiere '363, and Laganiere '817. *Laganiere '363*, ¶¶34, 38, 41; *Laganiere '817*, 2:51-63.

120. Accordingly, it is my opinion that the Laganiere '363-Laganiere '817 and/or Laganiere '461 combination discloses or suggests central server 108 (“object analysis system of the video analytics system”) that reads the metadata with quality scores associated with the images (“reads the ranking within the metadata”) and purges lower quality images from the face log (“decides whether or not to analyze the image data”)¹⁷.

5. Independent Claim 13

121. The Laganiere combination discloses or suggests Claim 13 for substantially the same reasons as claim 1, as I explained in §VIII.A.4, because the claims are substantively identical, and the Laganiere combination further discloses steps of a method “for analyzing video in a video surveillance system” described

¹⁷ I understand that “image data” to refer to “the extracted portions of the frames of the image data.”

above. *Laganiere* '363, Abstract; *Laganiere* '817, Abstract; *Laganiere* '461, Abstract; *compare* claim limitations 1.pre-1.5.3, with claim limitations 13.pre-13.5.3.

122. Claim 13.pre is substantially similar to 1.pre, except that it recites a “method” instead of a “system.” As I explained in §VIII.A.4, the *Laganiere* combination describes the recited method because it discloses steps to perform the claimed video analysis system of claim 13. §VIII.A.4.

123. The table below correlates the reasons why the *Laganiere* references disclose or suggest the limitations of claim 1 with limitations 13.1-13.5.3.

Claim 13		
Limitation	Substantially Similar Limitation	Reference
13.pre	1.pre	§VIII.A.4.a
A method for analyzing video in a video surveillance system, the method comprising:	A video surveillance system, comprising:	
13.1	1.1	§VIII.A.4.b
capturing image data from one or more surveillance cameras;	one or more surveillance cameras that capture image data;	
13.2.1	1.2	§VIII.A.4.c
detecting objects of interest within the image data via an object detection system; the object detection system extracting portions of frames that include the detected objects of interest and transmitting the extracted	an object detection system that detects objects of interest within the image data and extracts portions of frames that include the detected objects of interest and sends the extracted portions of the frames; and	

portions of the frames of the image data over a network,		
13.2.2	1.3.2	
the image data transmitted by the object detection system corresponding to objects of interest detected by the object detection system; and	the image data received by the video analytics system corresponding to objects of interest detected by the object detection system, the video analytics system analyzing the objects of interest;	§VIII.A.4.e
13.3	1.3.1, 1.4.2	
a video analytics system, which is operated by a different business entity than the one or more surveillance cameras, only receiving the extracted portions of the frames of the image data from the object detection system and analyzing the objects of interest;	a video analytics system that receives only the extracted portions of the frames of the image data, transmitted over a network, from the object detection system, the video analytics system is operated by a different business entity than the one or more surveillance cameras; and	§§VIII.A.4.d-VIII.A.4.g
13.4	1.4.1	
the video analytics system further receiving image data from other object detection systems;	wherein the video analytics system receives the image data from multiple object detection systems after being transmitted over the network and	§VIII.A.4.f
13.5.1-13.5.3	1.5.1-1.5.3	
the object detection system ranking the image data corresponding to objects of interest detected by the object detection system according to factors including close proximity of the objects of interest to the one or more	wherein the image data corresponding to objects of interest detected by the object detection system are ranked by the object detection system according to factors including close proximity of the objects of interest to the	§VIII.A.4.h-VIII.A.4.j

surveillance cameras, picture quality of the objects of interest, and/or contrast of the objects of interest to create a ranking and the ranking is saved in metadata with the image data; and an object analysis system of the video analytics system reading the rankings within the metadata and deciding whether or not to analyze the image data.	one or more surveillance cameras, picture quality of the objects of interest, and/or contrast of the objects of interest to create a ranking and the ranking is saved in metadata with the image data and an object analysis system of the video analytics system reads the rankings within the metadata and decides whether or not to analyze the image data.	
--	--	--

6. Claims 2 and 14: “wherein the object detection system creates cropped images from the image data that include the detected objects of interest and sends the cropped images to the video analytics system.”

124. By my review, claims 2 and 14 depend from claims 1 and 13, respectively, which as I explained above, Laganier ’363 discloses or suggests. Laganier ’363 further discloses or suggests claims 2 and 14’s additional limitation. *Id.*

125. As I explained in §VIII.A.4.c, Laganier ’363 discloses cropping image data to the object of interest. *Laganier* ’363, ¶41. Cropping regions of the video data is an example of “pre-processing.” *See id.*, ¶21.

126. Accordingly, it is my opinion that Laganier ’363 discloses or suggests video sources 102 (“object detection system”) creating the region of frame that

contains the event of interest (“creates cropped image from the image data that include the detected objects of interest”) and transmitting the region of video data that contains the event of interest to the central location (“video analytics system”).

127. Claim 14 is substantially similar to claim 2, except that it recites a “method” instead of a “system.” For the reasons discussed above for claim 2, it is my opinion that Laganier ’363 discloses or suggests claim 14. *Id.*

7. Claims 3 and 15: “wherein the image data include video data and the object detection system provides frames from the video data that include the detected objects of interest and sends the frames to the video analytics system as the image data.”

128. By my review, claims 3 and 15 depend from claims 1 and 13, respectively, which as I explained above, Laganier ’363 discloses or suggests. It is my opinion that Laganier ’363 further discloses or suggests claims 3’s additional limitation. *Id.*

129. A POSITA would have understood “the image data include video data” as recited to mean that the image data is included in video data as frames of video data. As I explained in §VIII.A.4.c, Laganier discloses or suggests video sources 102 (“object detection system”) extracting and transmitting the cropped frame of data (“portions of the frames from the video data”) that contains the event of interest (“detected objects of interest”) to the central location (“video analytics system”) for further processing.

130. A POSITA reading claim 3 in light of claim 1, from which it depends,

would have understood the recited “provides **frames from the video data** that include the detected objects of interest” as referring back to claim 1’s recitation of “extracted portions of the frames.” Accordingly, it is my opinion that Laganieri ’363 discloses or suggests claim 3.

131. To the extent that “*frames*” is interpreted as full frames, claim 3 is indefinite. Claim 1 of the ’321 patent, from which claim 3 depends, requires that the “video analytics system that receives only the extracted portions of the frames of the image data.” A POSITA would have thus understood that if **only** portions of frames can be received, then it would be impossible for full frames to be received by the video analytics system. Regardless, Laganieri ’363 sends full frames of data to the central server for further processing. For example, Laganieri ’363 discloses “providing portions of the video data that relate to other than the event of interest at a lower frame rate than is used for portions of the video data that relate to the event of interest.” *Laganieri ’363*, ¶21; *see also, id.*, ¶38. Therefore, it is my opinion that Laganieri ’363 teaches claim 3’s additional limitation, even under this interpretation.

132. By my review, claim 15 is substantially similar to claim 3, except that it recites a “method” instead of a “system.” For the reasons I explained discussed above for claim 3, Laganieri ’363 discloses or suggests claim 15. *Id.*

8. Claims 4 and 16: “The system of claim 1, wherein the video analytics system includes an object analysis system that analyzes the image data corresponding to objects of interest detected by the object detection system by comparing the image data to reference images to identify the objects of interest, by recognizing patterns in the image data, and/or by applying image processing functions, neural network functions, deep learning functions, and/or object classification functions to the image data.”

133. By my review, claims 4 and 16 depend from claims 1 and 13, respectively, which as I explained above, Laganier '363 discloses or suggests. In my opinion, the Laganier combination further discloses or suggests claim 4's additional limitation. *Id.*

134. As I explained in §VIII.A.4.f, it is my opinion that the Laganier combination discloses a central server 108 (“video analytics system”) including analytics engines (“object analysis systems”) that analyzes portions of video data that contain the event of interest detected by the video source (“image data corresponding to objects of interest detected by the object detection system”)¹⁸.

135. By my review, claim 4 contains three parallel clauses beginning with “by” (i.e., “by comparing,” “by recognizing,” and “by applying”) that are connected with an “and/or” conjunction. It is my opinion that Laganier '363 teaches at least one of these three parallel clauses. For example, Laganier '363 discloses that each

¹⁸ I understand that “image data” to refer to “the extracted portions of the frames of the image data.”

of the analytics engines 120-124 (“object analysis systems”) “performs different video analytics function” (“applying image processing functions”) on the data frames. Laganieri ’363, ¶30. A POSITA would have regarded the video analytics functions such as facial recognition and reading license plate numbers as image processing functions because both functions relate to the analysis of image data.

136. Claim 16 is substantially similar to claim 4, except that it recites a “method” instead of a “system.” Claim 16 appears to have typographical errors, in that it recites “by recognizing” and “by applying” in its recited methods. A POSITA would have realized that “by” was an artifact of transposing claim 4 into a method claim format because claim 16 also recites three parallel clauses that are substantially similar to those recited in claim 4.

137. Accordingly, it is my opinion that Laganieri ’363 discloses or suggests claim 4.

9. Claims 5 and 18: “wherein the object detection system is integrated within the one or more surveillance cameras.”

138. By my review, claims 5 and 18 depend from claims 1 and 13, respectively, which as I explained above, Laganieri ’363 discloses or suggests. It is my opinion that Laganieri ’363 further discloses or suggests claims 5’s additional limitation. *Id.*

139. As I explained in §VIII.A.4.c, it is my opinion that Laganieri ’363 teaches multiple source ends including video sources 102 (“one or more surveillance

cameras”) that contains a processor (“object detection system”) for “detecting the event of interest in a portion of the video data.” Laganieri ’363, ¶40.

140. Accordingly, it is my opinion that Laganieri ’363 discloses or suggests a processor (“object detection system”) integrated in video source 102 (“one or more surveillance cameras”).

141. By my review, claim 18 is substantially similar to claim 5, except that it recites a “method” instead of a “system.” For the reasons discussed above for claim 5, Laganieri ’363 discloses or suggests claim 18’s additional limitation. *Id.*

10. Claims 7 and 17: “further comprising a database that receives metadata associated with the identified objects of interest from the video analytics system.”

142. By my review, claims 7 and 17 depend from claims 1 and 13, respectively, which as I explained above, Laganieri ’363 discloses or suggests. It is my opinion that Laganieri ’363 further discloses or suggests claims 7’s additional limitation. *Id.*

143. As explained in §VIII.A.4.c, Laganieri ’363 discloses that the event of interest (e.g., a vehicle license plate within video data) teaches “identified objects of interest.” Laganieri ’363 further discloses that the “second video analytics is performed for determining the **license plate number**” and “[u]pon completion some or all of the frames are stored in **non-volatile memory**.” Laganieri ’363, ¶41. A POSITA would have recognized a “license plate number” is metadata because it is

data describing the license plate image. “[T]he necessary frames and **the result are stored** in association one with another as evidence and conclusion.” *Laganiere* ’363, ¶41. It would have been obvious to a POSITA to utilize a database to store the image data and result because databases were widely used for persistent storage of application data. *Laganiere* ’363 discloses that central server 108 accesses a database. *Laganiere* ’363, ¶¶35, 38. A POSITA would have known that such a database could be used to associate the image data and result (e.g., license plate number) as *Laganiere* ’363 describes using tables and/or relations between database tables. A POSITA would also have recognized that a memory that stores data with associated other data is a database. *Id.*

144. Accordingly, it is my opinion that *Laganiere* ’363 discloses or suggests a database that receives the license plate number (“receives metadata”) associated with a vehicle license plate (“identified objects of interest”) from the central server 108/308 (“video analytics system”).

145. By my review, claim 17 is substantially similar to claim 7, except that it recites a “method” instead of a “system.” For the reasons I explained discussed above for claim 7, it is my opinion that *Laganiere* ’363 discloses or suggests claim 17’s additional limitation. *Id.*

11. Claims 8 and 19: “further comprising a gateway system that provides access to the video analytics system over a public network.”

146. By my review, claims 8 and 19 depend from claims 1 and 13, respectively, which as I explained above, Laganier ’363 discloses or suggests. It is my opinion that Laganier ’363 further discloses or suggests claims 8’s additional limitation. *Id.*

147. As I explained in §VIII.A.4.d, Laganier ’363’s system includes a “video source 102 [] in communication with **a central server 108 via gateway 104 and Wide Area Network (WAN) 106.**” *Id.*, ¶25. Laganier ’363 explains that “[s]ome non-limiting examples of a WAN include: a computer network such as the **Internet of the World Wide Web.**” *Id.*, ¶21. A POSITA would have understood that the Internet is an example of a public network because it is open to all data traffic. Annotated Figure 2 Laganier ’363 below shows gateway 104 (blue) in communication with central server 108 (purple) through WAN 106 (red).

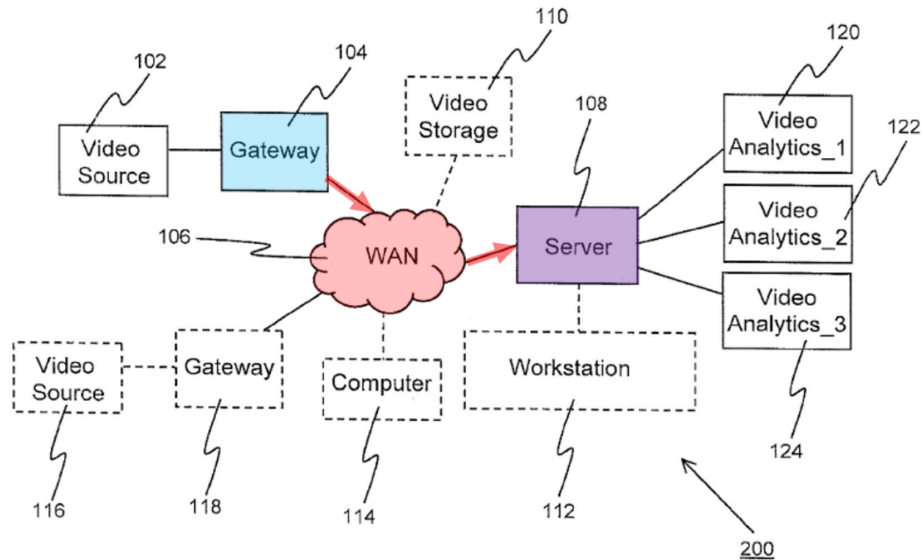


Fig. 2

148. Accordingly, it is my opinion that Laganieri '363 discloses or suggests gateway 104 (“gateway system”) that is in communication with central server 108 (“video analytics system”) over the Internet (“public network”).

149. By my review, claim 19 is substantially similar to claim 8, except that it recites a “method” instead of a “system.” For the reasons I explained above for claim 8, Laganieri '363 discloses or suggests claim 19’s additional limitation. *Id.*

12. Claims 9 and 20: “wherein the object detection system is integrated within the gateway system.”

150. By my review, cclaims 9 and 20 depend from claims 8 and 19, respectively, which as I explained discussed above, it is my opinion that Laganieri '363 discloses or suggests. In my opinion, Laganieri '363 further discloses or suggests claims 9’s additional limitation. *Id.*

151. Laganieri '363 discloses that “[a]lternatively, the video source 102 is a

‘dumb’ IP camera or an analogue video camera, coupled with a not illustrated video encoder and/or a local video analytics engine.” *Laganiere* ’363, ¶25, 29. A POSITA would have recognized that under such configuration, the processor (“object detection system”) disposed at the source end for detecting the event of interest would be in the “video encoder and/or a local video analytics engine.” *Id.*, ¶25, 29. *Laganiere* ’363 does not limit the processor to be within video source 102, rather, the processor for performing the initial pre-processing of video data is “disposed at the source end.” *Laganiere* ’363, ¶¶10, 40. A POSITA would have been motivated to integrate the processor (“object detection system”) into gateway 104 because gateway 104 is disposed at source end. A POSITA would have been motivated to integrate the source end processing into gateway 104 because both are computers and such consolidation would save costs and eliminate unnecessary hardware. A POSITA would have been further motivated to integrate the processor in gateway 104 to take advantage of the known video encoding capabilities of gateway systems. For example, Wang discloses a gateway that executes “algorithms such as face detection/recognition and license plate recognition” on “recorded video and still

pictures” to “extract information.” *Wang*, ¶39; *see also, id.*, ¶38, 40-42.¹⁹

152. By my review, claim 20 is substantially similar to claim 9, except that it recites a “method” instead of a “system.” For the reasons I explained discussed above for claim 9, it is my opinion that Laganieri ’363 discloses or suggests claim 20. *Id.*

13. Claims 10 and 21: “wherein the one or more surveillance cameras each include a gateway system that provides access to the video analytics system over a public network.”

153. By my review, claims 10 and 21 depend from claims 1 and 13, respectively, which as I explained above, Laganieri ’363 discloses or suggests. In my opinion, Laganieri ’363 further discloses or suggests claims 9’s additional limitation. *Id.*

154. Laganieri ’363 discloses that “[o]ptionally, the video source 102 connects to WAN 106 without the gateway 104.” *Laganieri ’363*, ¶¶28, 33. This teaching that the video source 102 could access the WAN without relying on the gateway 104 suggests that the video source 102 incorporates a gateway for accessing the network. Indeed, a POSITA would have understood that a gateway to mean “any

¹⁹ Deever ’150, Deever ’965, Gali, Yang, and Wang are cited in this declaration to demonstrate POSITA’s knowledge and/or as evidence to support motivation to combine. These exhibits are not part of the unpatentability ground.

device that connects two or more dissimilar networks.” *See* Tanenbaum, p. 397. Laganieri discloses that “the video source 102 and/or 116 comprises a plurality of separate video sources disposed at the source end and connected to gateway 104 or 118 **via a not illustrated router.**” Laganieri ’363 also discloses that “video source 102 is a network IP camera with onboard video analytics capabilities, such as for instance **an AXIS 211M Network Camera.**” *Laganieri* ’363, ¶¶25, 29. By the time of the alleged invention, network-connectable cameras were known in the art. Indeed, the AXIS 211M Network Camera could directly connect to a LAN and to the internet. *See also*, 1017 (AXIS 211M manual) at pp. 8, 25.

155. Accordingly, it is my opinion that Laganieri ’363 discloses or suggests video source 102 (“one or more surveillance cameras”) includes a gateway (“gateway system”) that provide access to the Internet (“public network”).

156. By my review, claim 21 is substantially similar to claim 10, except that it recites a “method” instead of a “system.” For the reasons discussed above for claim 10, Laganieri ’363 discloses or suggests claim 21. *Id.*

14. Claim 11: “wherein the video analytics system receives the image data from the object detection system after being transmitted over an enterprise network.”

157. By my review, claim 11 depends from claim 1, which as I explained above, Laganieri ’363 discloses or suggests this limitation. In my opinion, Laganieri ’363 further discloses or suggests claim 11’s additional limitation. *Id.*

158. As I explained in §VIII.A.4.f, central server 108 receives image data from the source end over WAN 106. Laganieri '363 further discloses that “non-limiting examples of a WAN include” “a Wi-Fi network.” *Laganieri* '363, ¶21. A POSITA would have known that a Wi-Fi network is an example of wireless LAN because it connects a group of computers devices that typically share common resources within a limited geographical area. *See, e.g.*, Tanenbaum at p. 9 (“Local area networks, generally called LANs, are privately-owned networks within a single building or campus of up to a few kilometers in size. They are widely used to connect personal computers and workstations in company offices and factories to share resources (e.g., printers) and exchange information.”). A POSITA would also have understood that for small to medium sized business, an enterprise network is an example of LAN that consists of such geographically limited group of computers. A POSITA would also have known that an enterprise network for a larger enterprise could be a WAN with multiple interconnected LANs.

159. Accordingly, it is my opinion that Laganieri '363 discloses or suggests video analytics engines 120-124 (“video analytics system”) receives video data (“image data”) from processor of video source 102 (“object detection system”) through a WAN (“enterprise network”).

15. Claim 12: “wherein the video analytics system receives the image data from the object detection system after being transmitted over a cloud network.”

160. By my review, claim 12 depends from claim 1, which as I explained above, Laganieri '363 discloses or suggests this limitation. In my opinion, Laganieri '363 further discloses or suggests claim 12's additional limitation. *Id.*

161. The '321 patent discloses that “network cloud 60 is preferably a public/service provider network such as the Internet.” '321 patent, 5:13-14. A POSITA would have understood that the phrase “network cloud” to also refer to a “cloud network” because both are directed to an infrastructure layer that can accommodate networking functions. Likewise, Laganieri '363 teaches that “[c]loud computing is a general term for anything that involves delivering hosted services over the Internet.” Laganieri '363, ¶23. Laganieri '363 further discloses that “[a]ccording to at least some of the described embodiments, the pre-processed video data is enhanced to be more useful when subsequent **video analytics processing is performed ‘in the cloud.’**” *Id.*, ¶24. Therefore, Laganieri '363 discloses that central server 108 and/or video analytics engines 120-124 (“*video analytics system*”) perform video analytics “in the cloud” and a POSITA would have been understood that they receive the image data through a cloud network. As I explained in §VIII.A.14, the image data is received from the processor of video source 102 (“object detection system”).

162. Accordingly, in my opinion, Laganieri '363 discloses or suggests this limitation.

B. Ground 1B: Laganieri '363 Renders Claims 4, 6, 16, Obvious in View of Laganieri '817, and Chen, with or without Laganieri '461

163. As I explained above, Laganieri '363 in combination with Laganieri '817 with or without Laganieri '461 renders obvious the claims challenged in Ground 1A. I further show here in Ground 1B that dependent claims 6, 16, and independent claim 22, would have been obvious over the combination of Laganieri '363, Laganieri '817, Laganieri '461, and Chen. To the extent Patent Owner argues, or the Board finds, the combination of Laganieri '363, Laganieri '817, and Laganieri '461 do not disclose or render obvious claim 4, the combination of Laganieri '363, Laganieri '817, Laganieri '461, and Chen renders these limitations obvious.

1. Summary of Chen

164. In my opinion, Chen is analogous art to the '321 patent. For example, Chen and the '321 patent are in the same field of endeavor. Both are directed to video surveillance systems having video and image analytics capabilities. *Compare '321 patent*, Abstract, 4:38-44; *with Chen*, ¶¶4, 53, 67, 90.

165. Chen is directed to “a computer implemented method, software, and computer for assessing the quality of an image.” Chen, Abstract. The method is particularly useful for assessing the quality of “the image of a face” for use “in face

recognition.” *Id.* Chen’s system includes surveillance camera 502, computer 500, and external storage 506, as depicted in Figure 5 of Chen below.

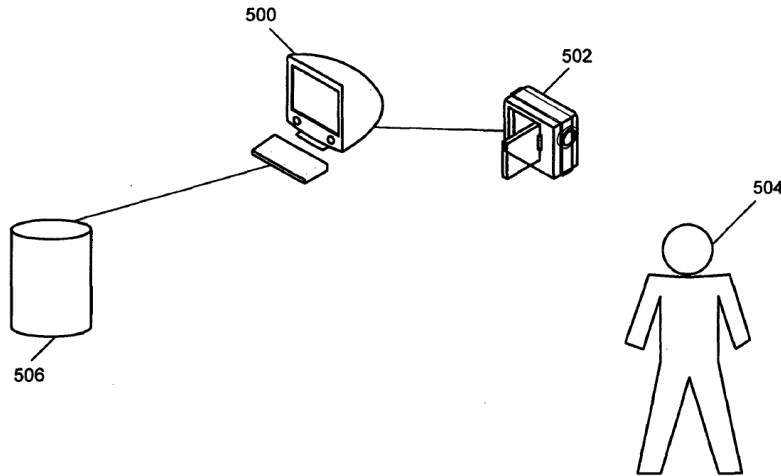


Fig. 5

166. Chen discloses a pre-processing step where results can be fed to more analysis: “[t]he method may be used as a pre-processing step for the image, and further processing may be based on the determined quality of the image. It is an advantage that the determination of quality can be used to select the better quality images for further processing which in turn improves the accuracy of the further processing.” *Chen*, ¶13.

167. Figure 6 of Chen describes “face recognition system” to determine “the identity of the face.” *Chen*, ¶¶90-92. The workflow begins with assessing a set of captured face images for image quality at 604, and “[b]ased on the result of the quality assessment the face images are selected 606 resulting in a subset of face images 608 to be used for face recognition.” *Id.* At operation 610, “a most suitable

face matching method is selected” based on “how well aligned the face images are” with gallery images 612. *Chen*, ¶91. Known face matching methods such as PCA, LBP, and MRH variant are selected based on the alignment. *Id.* “The result of face matching method is provided” to subsequent post processing at 618 and “finally the identity of the face is determined [at] 620.” *Id.*, ¶¶91-92. I have reproduced Figure 6 of *Chen* below.

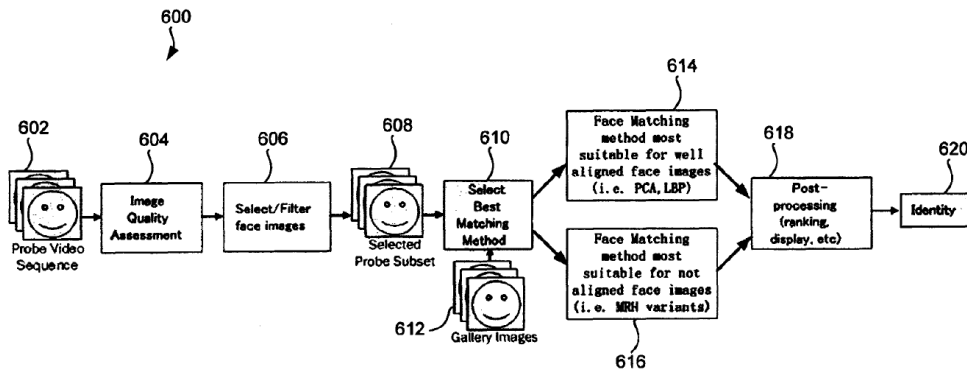


Fig. 6

2. Motivation to Combine

168. As I explained in §VIII.A.4.c, *Laganiere* '363 discloses performing pre-processing at the source end and sending detected objects and the pre-processing result to the central server for further analysis. *See Laganiere* '363, ¶¶34, 38. Central server 108 reads the metadata and “perform[] the additional video analytics processing of the pre-processed data.” *Id.* The additional video analytics can be performed by multiple different video analytics engines 120-124, each of which

“performs a different video analytics function.” *Laganiere* ’363, ¶30. Specifically, “Video_Analytics_3 124 performs facial recognition.” *Id.* In addition, *Laganiere* ’363 discloses generating predetermined actions based on the result of the additional video analytics, such as controlling a doorbell “in dependence upon recognizing a person standing at the door.” *Id.*, ¶36. However, *Laganiere* ’363 does not expressly disclose in detail a facial recognition process or how a person is identified through the facial recognition process. A POSITA would have thus looked to other references for details of facial recognition processes.

169. A POSITA would have looked to *Chen*, which is in the same field of endeavor as *Laganiere* ’363 and is directed specifically to performing a facial recognition process on a set of facial images. For example, *Chen* discloses reliably performing a facial recognition process by performing image quality assessment on an initial set of images and selecting high-quality images to form a sub-set of images for subsequent facial recognition. *Chen*, ¶¶90-92; *Laganiere* ’363, ¶¶34, 41. Similarly, *Laganiere* ’363 discloses pre-processing video data at the source end by selecting only high-quality data for subsequent processing at the central computer. *Chen* further discloses comparing received images with gallery images to select the most suitable face matching method. *Chen*, ¶91.

170. A POSITA would have been motivated to use *Chen*’s algorithms of a facial recognition (e.g., workflow 600 of a face recognition system) in

Laganiere '363's video analytics engine 124 to reliably perform a facial recognition process and to successfully determine the identity of the person.

171. A POSITA would have also been motivated to combine Chen with Laganiere '363's video analytics engines because it would have required nothing more than the application of known, conventional techniques (using, for example, facial recognition processes) to yield predictable results (to reliably and efficiently determine identity of person) as described in Laganiere '363 and Chen. *Laganiere '363*, ¶¶30, 36; *Chen*, ¶¶90-92.

172. A POSITA would have further been motivated to combine Chen with Laganiere '363's video analytics engine 124 because it would have been applying a known technique (comparing received images to gallery images to select the most suitable facial matching method) to a known device of Laganiere '363 (video analytics engine) ready for improvement to yield predictable results (reliably and efficiently performing a facial recognition process) as described in Laganiere '363 and Chen. *Laganiere '363*, ¶¶30, 36; *Chen*, ¶¶90-92.

3. Claims 4 and 16: “The system of claim 1, wherein the video analytics system includes an object analysis system that analyzes the image data corresponding to objects of interest detected by the object detection system by comparing the image data to reference images to identify the objects of interest, by recognizing patterns in the image data, and/or by applying image processing functions, neural network functions, deep learning functions, and/or object classification functions to the image data.”

173. By my review, claims 4 and 16 depend from claims 1 and 13,

respectively, which as I explained discussed above, Laganier '363 discloses or suggests. As I explained in §VIII.A.8, Laganier '363 discloses that each of the analytics engines 120-124 (“object analysis systems”) “performs different video analytics function” (“applying image processing functions”) on the data frames. Laganier '363, ¶30.

174. I note that Chen discloses comparing received subset of face images 608 to gallery images 612 to determine how well the faces align across the two sets of images. Chen, ¶91. Based on that determination, a most suitable face matching process is selected to determine the identity of the person at operation 620. *Id.*, ¶¶91-92. Therefore, Chen discloses comparing a subset of face images 608 (“comparing the image data”)²⁰ to gallery images (“reference images”) to determine the identity of the face (“identify the objects of interest”).

175. Accordingly, it is my opinion that the combination of Laganier '363, Laganier '817, Laganier '461, and Chen discloses or suggests claims 4 and 16.

4. Claim 6: “The system of claim 1, wherein the image data corresponding to objects of interest detected by the object detection system are compared by the video analytics system to reference images to identify the objects of interest.”

176. By my review, claim 6 depends from claim 1, which as discussed above,

²⁰ I understand that “image data” to refer to “the extracted portions of the frames of the image data.”

Laganiere '363 discloses or suggests. As I explained in §VIII.B.3, Chen discloses comparing a subset of face images 608 (“comparing the image data”)²¹ to gallery images (“reference images”) to determine the identity of the face (“identify the objects of interest”).

177. Accordingly, it is my opinion that the combination of Laganiere '363, Laganiere '817, Laganiere '461, and Chen discloses or suggests claim 6.

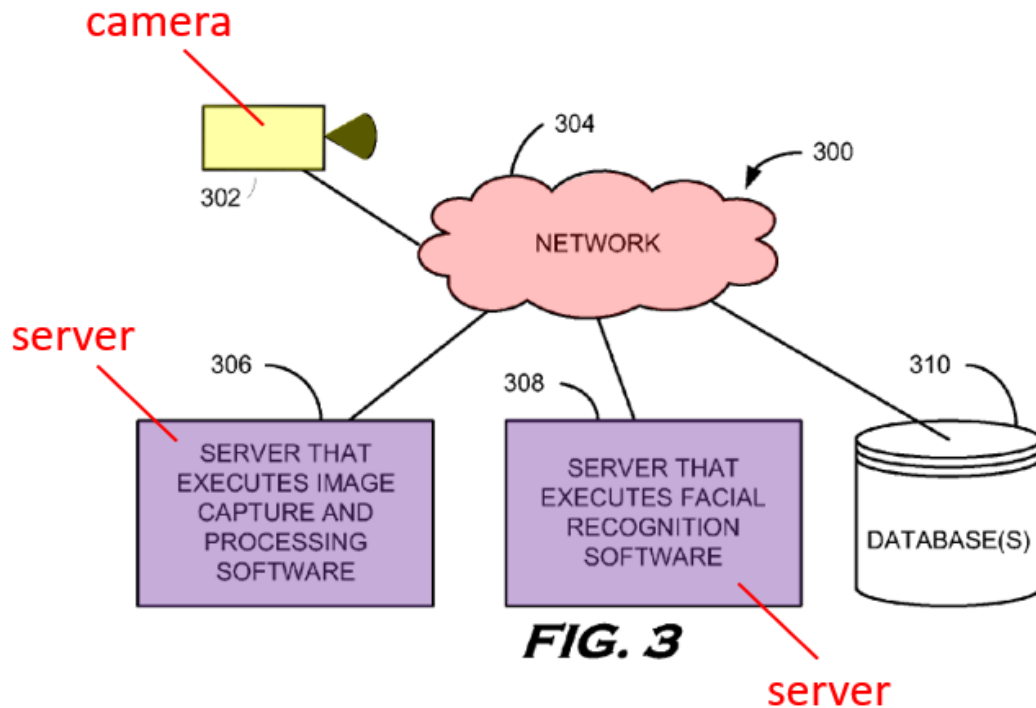
C. Ground 2: Bedros Renders Claims 1-21 Obvious

1. Summary of Bedros

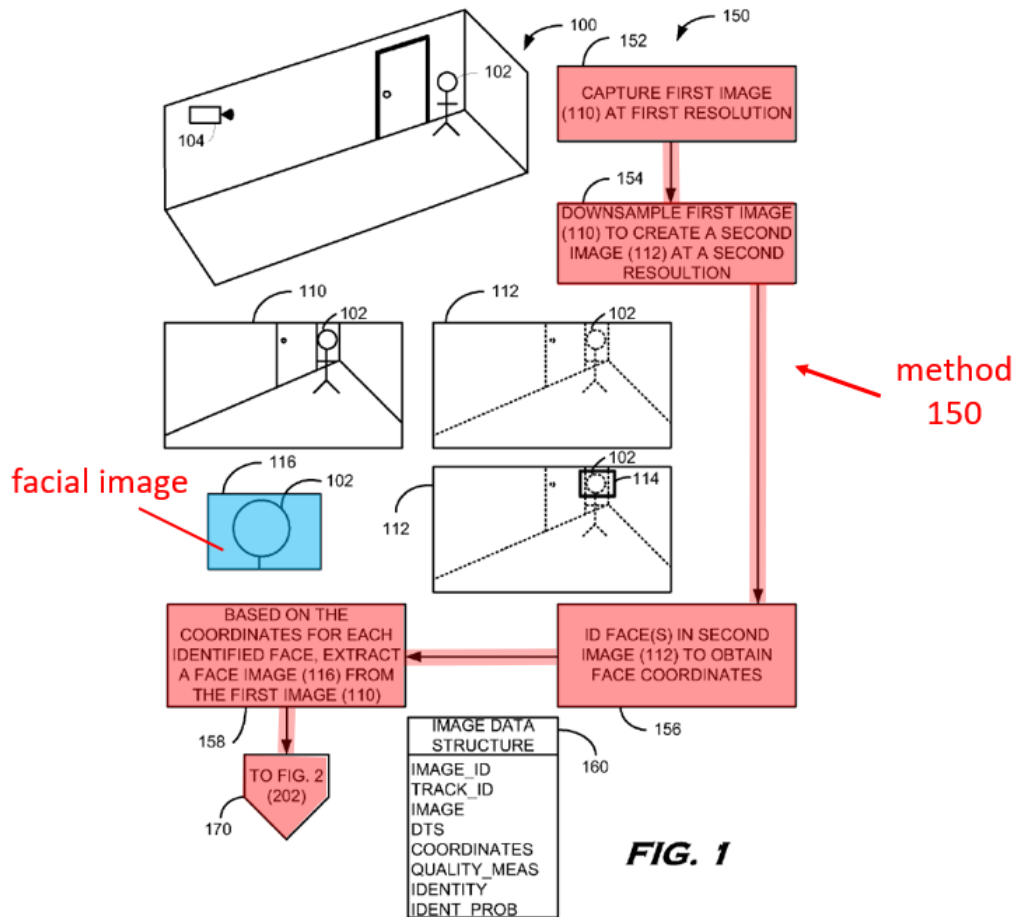
178. In my opinion, Bedros (EX1008) is analogous art to the '321 patent.

179. Bedros is directed to “systems, methods, and software” that “capture images, locate individuals, and provide images” to a “[surveillance-based high-resolution] facial recognition process.” *Bedros*, Title, Abstract, 1:6-26, 1:43-57. Bedros’s system includes “a camera 302 that is connected to a network 304” and “one or more computing devices that operate server process and database 310 systems.” *Id.*, 8:17-20. Computing devices “execute[] image capturing and processing software” (server 306) and “execute facial recognition software” (server 308). *Id.*, 8:13-49. Annotated Figure 3 below shows camera 302 (yellow) connected to servers 306 and 308 (purple) through network 304 (red). *Id.*, Fig.3, 8:13-49.

²¹ See footnote 19.



180. Bedros discloses method 150 and additional method 200 implemented by system 300 of Figure 3. *Id.*, 8:14-16. Method 150 describes acquiring and processing images to “locate, isolate, and extract a facial image 116 that can be submitted to a facial recognition process.” *Id.*, 2:40-43; *see also id.*, 2:61-4:29. The extracted facial image 116 contains the face of a person and excludes the rest of the image. *See id.*, Fig. 1, 3:42-46. Annotated Figure 1 below shows method 150 (red) and facial image 116 (blue). *Id.*, Fig.1, 2:36-43.



181. Method 200 is optionally performed following method 150 and describes “filtering and prioritizing facial images 116 for submission to the facial recognition program” by an “optional processing layer 220.” *Id.*, 4:23-29, 6:23-26. “[A] quality measure is applied [at] 210 to the facial image 116” by evaluating various features of facial image 116, and a score is assigned to each factor or “a composite score may be assigned” to facial image 116. *Id.*, 6:4-16. Shown in annotated Figure 2 below, three processing elements 222, 224, and 226 (blue) of optional processing layer 220 prioritizes facial images 116 based on quality measure

values and determines which facial images 116 to submit first into image queue at operation 230. *Id.*, 6:39-7:18. Based on the prioritization results, prioritization determines which facial images 116 are submitted first into image queue at operation 230. *See id.*, 7:19-25.

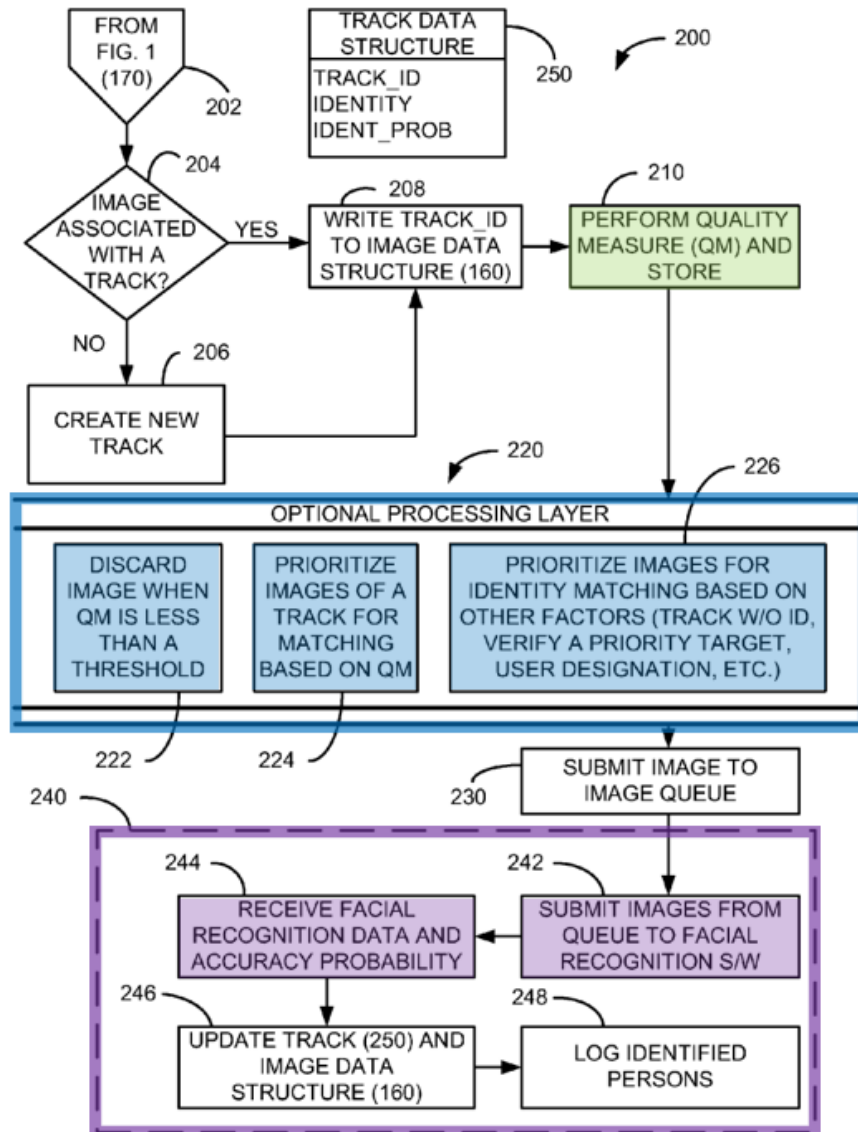


FIG. 2

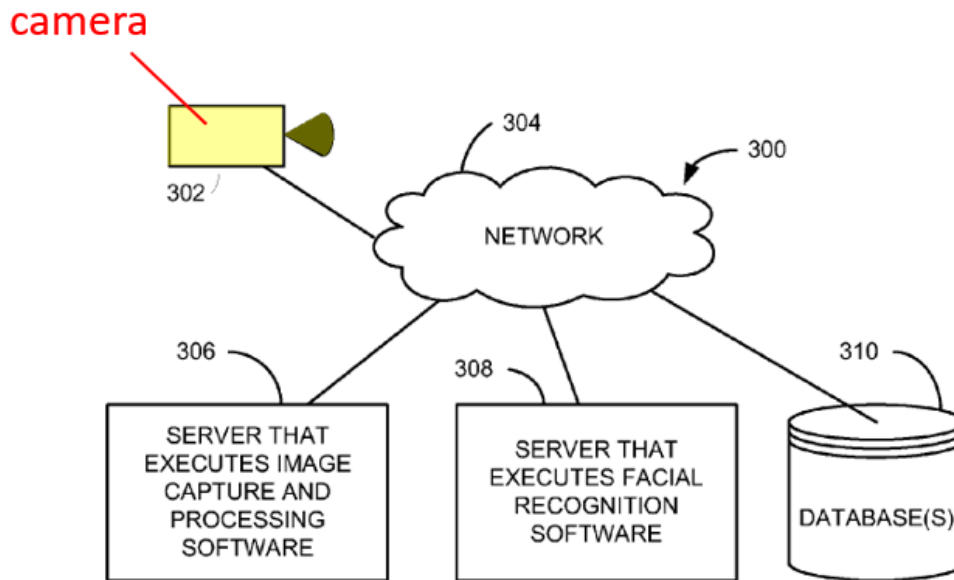
182. Bedros explains that “submitting 230 the facial image 116 to the queue

may include removing and reordering facial images 116...based on processing outcomes within the optional processing layer 220.” *Id.*, 7:21-25. Queue processing 240 “submit[s] 242 images from the queue to the facial recognition program and receiv[s] 244 facial recognition data in response.” *Id.*, 7:26-30.

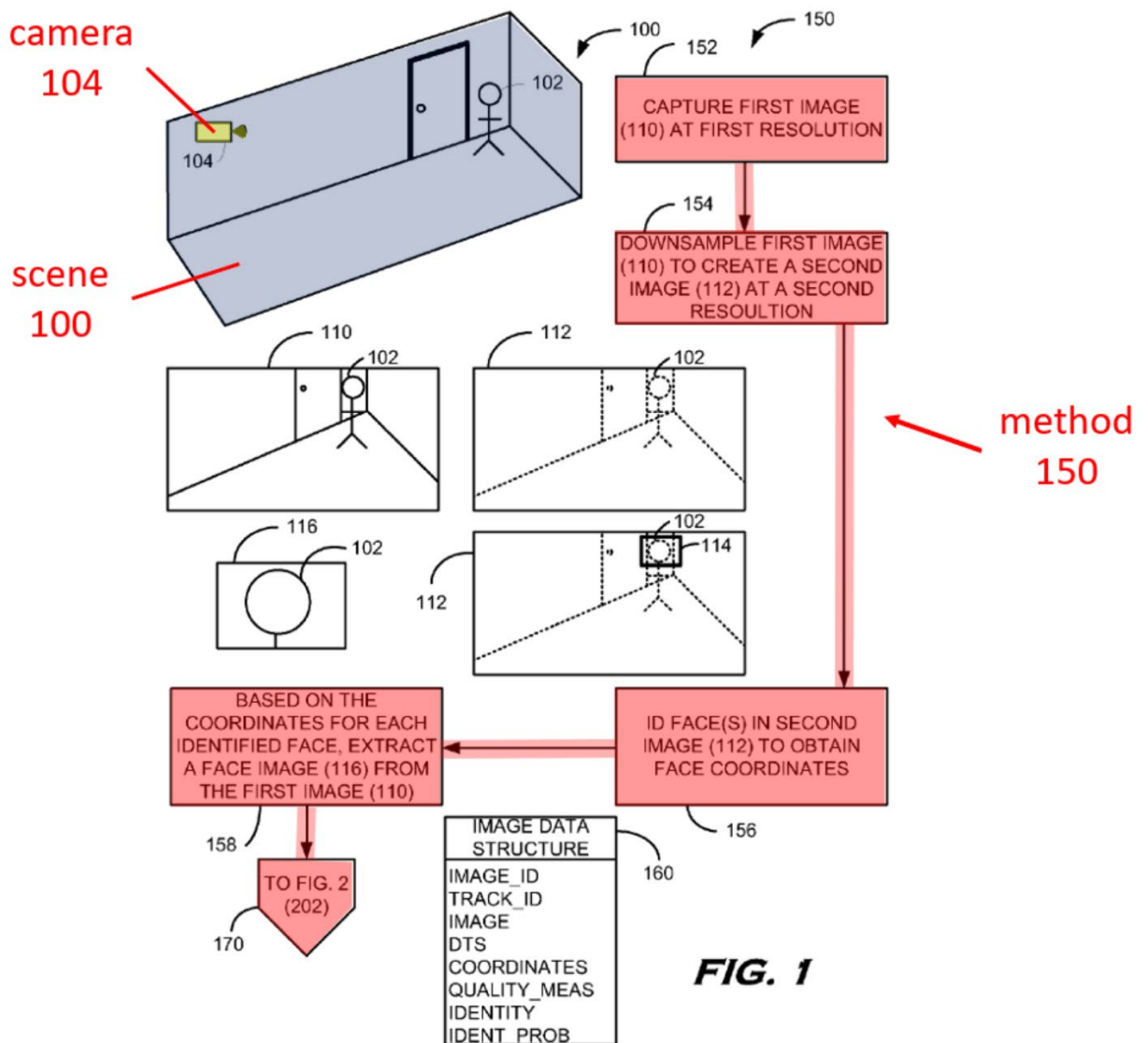
2. Independent Claim 1

a) Claim 1.pre: “A video surveillance system”

183. If limiting, it is my opinion that Bedros discloses or suggests 1.pre. Bedros recognizes limitations with existing video surveillance systems and seeks to overcome those limitations by developing “systems, methods, and software” that capture and provide images to a “[surveillance-based high-resolution] facial recognition process.” *Bedros*, Title, Abstract, 1:6-26, 1:43-57. Bedros discloses “[v]arious...systems, methods, and software that utilize[] a single high-resolution camera to capture images, locate individuals, and provide images to a facial recognition process.” *Id.*, 1:58-62. Bedros discloses the “system” includes “a camera 302 that is connected to a network 304” (*id.*, 8:17-18) and “one or more computing devices that operate server process and database 310 systems” (*id.*, 8:18-20). Figure 3 (annotated below) depicts Bedros’ system. *Id.*, Fig.3, 8:13-49.

**FIG. 3**

184. Bedros' system disclosed in Fig. 3 is deployed, for example, in scene 100 (Figure 1, below) that "includes the presence of a person 102" standing next to a building entrance. *Id.*, 2:46-47, 2:36-43, 8:13-16, Fig. 1. Camera 104 (e.g., "a digital **video camera** or digital **still image camera**") "monitors [] scene 100." *Id.*, 2:44-50, Fig. 1. Scene 100 could include "a thoroughfare, such as a hallway where people move through" or "a work area where people move about the entire scene." *See Bedros*, 5:4-11, 5:26-33.



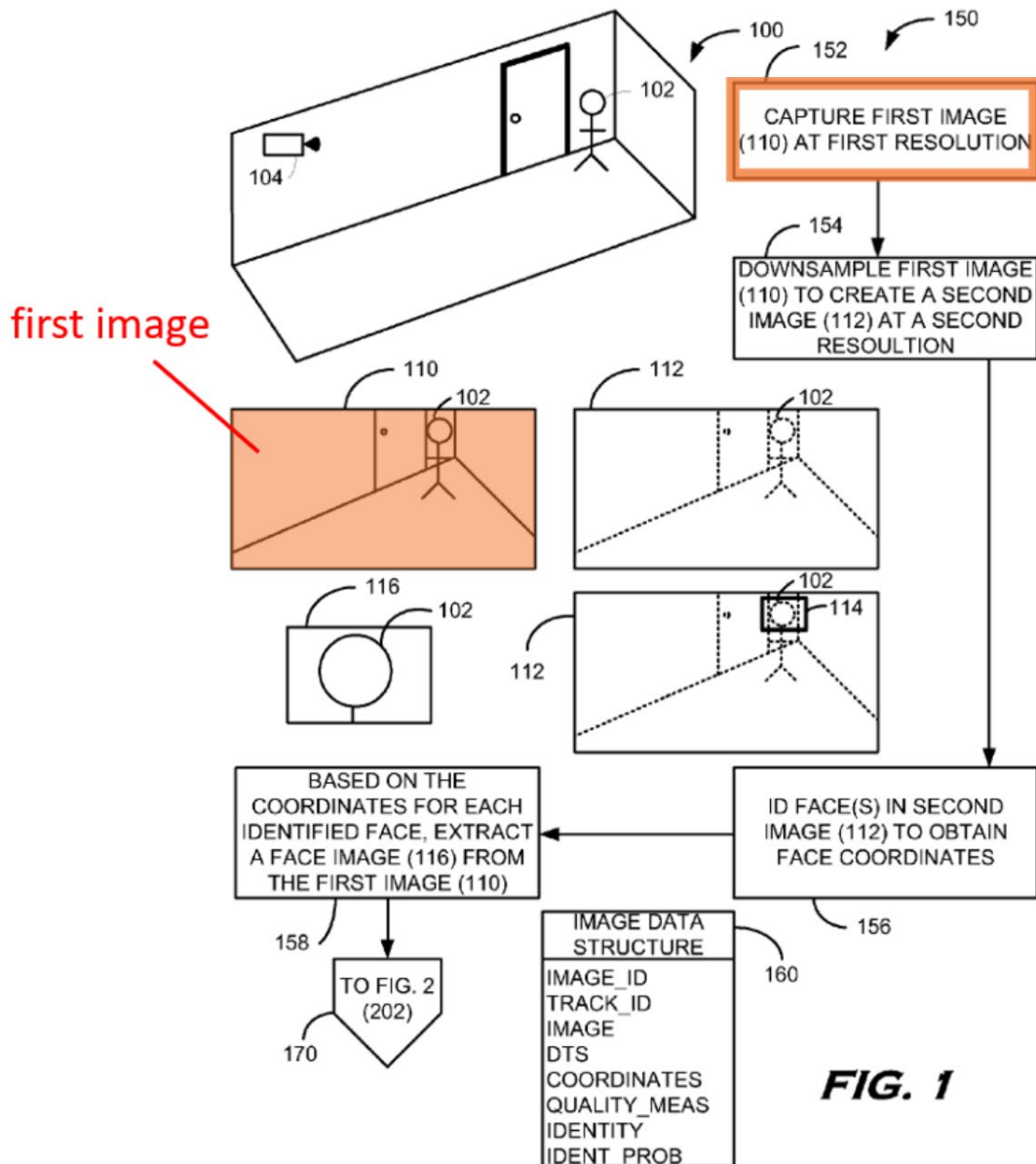
185. According to method 150 which is described with reference to scene 100, camera 104 acquires and processes images to “locate, isolate, and extract a facial image 116 that can be submitted to a facial recognition process.” *Id.*, 2:40-43, *see also id.*, 2:61-4:29. Bedros’s system is used for “*video surveillance*” because its facial recognition process is a “[s]urveillance-based high-resolution facial recognition” process implemented using a digital video camera and servers. *Id.*, Title,

2:44-60, 8:13-31, Fig. 3.

186. Accordingly, it is my opinion that Bedros discloses or suggests a system including camera 302 or 104, network 304, servers 306 and 308, and database 310 (“video surveillance system”).

b) Claim 1.1: “one or more surveillance cameras that capture image data”

187. In my opinion, Bedros discloses or suggests 1.1. Camera 104 that monitors scene 100 “may be a digital video camera or digital still **image camera** with suitable optics to **capture images of scene** as needed.” *Bedros*, 2:44-50. Specifically, method 150 depicted in Figure 1 begins with “capturing 152 a first image 110 acquired by a camera 104 that monitors a scene 100.” *Id.*, 2:44-46. A POSITA would have understood that captured images of the scene include “*image data*,” as recited in 1.1. Indeed, the ’321 patent explains that “images and/or video, [are] also known as image data.” ’321 *patent*, 5:29-34. Moreover, Bedros explains that the captured images are stored in an “image data structure”: “method 150 is a method of acquiring an image 110, processing that image to locate, isolate, and extract a facial image 116 that can be submitted to a facial recognition process” and “once the facial image 116 is extracted 158, the facial image 116 may be stored to an **image data structure** 160.” *Bedros*, 2:40-43, 3:56-58. Annotated Figure 1 below shows first image 110 (orange shade) and capture operation 152 (orange box).



188. A POSITA would have understood that the captured images consist of data because they are processed and stored in an **image data** structure. Specifically, method 150 includes processing the image to “locate, isolate, and extract a facial image 116,” which is then “stored to an **image data** structure 160.” *Bedros*, 2:40-43, 4:56-58. *Bedros* provides several examples of data stored in image data structure

160. *Bedros*, 3:61-67, Fig. 1.

189. Camera 104 captures video data in the same manner as “video data” in the ’321 patent. A POSITA would have understood that Bedros’s video stream consists of a series of images because the images are captured “from an image or **video stream** received from the camera 140.” *Bedros*, 2:61-63. A POSITA would also have understood the video data is captured at varying frame rates. For example, Bedros explains that “as a frame rate of the camera increases, the amount of **image data** to be processed increases,” and that “the more rapid the frame rate from the camera 104, the more data processing that is required.” *Bedros*, 4:8-11, 4:20-22. In the same way, the “frame capture system 106-1 of the surveillance camera 103-1 [in the ’321 patent] captures image data” and “encodes the image data into **frames of video data**.” ’321 patent, 5:60-64, 8:60-64, 10:7-11. Moreover, the ’321 patent acknowledged that “[a] number of components are common to many of these video surveillance systems. **Surveillance cameras capture image data**.” ’321 patent, 1-16-19 (discussing prior art).

190. A POSITA would have understood Bedros’s camera 104 to be a “surveillance camera” because it monitors scenes of interest, such as “the presence of a person 102” in a room. *Id.*, 2:44-46, Fig. 1.

191. Accordingly, it is my opinion that Bedros discloses or suggests at least one camera 104 capturing images of scene (“one or more surveillance cameras that

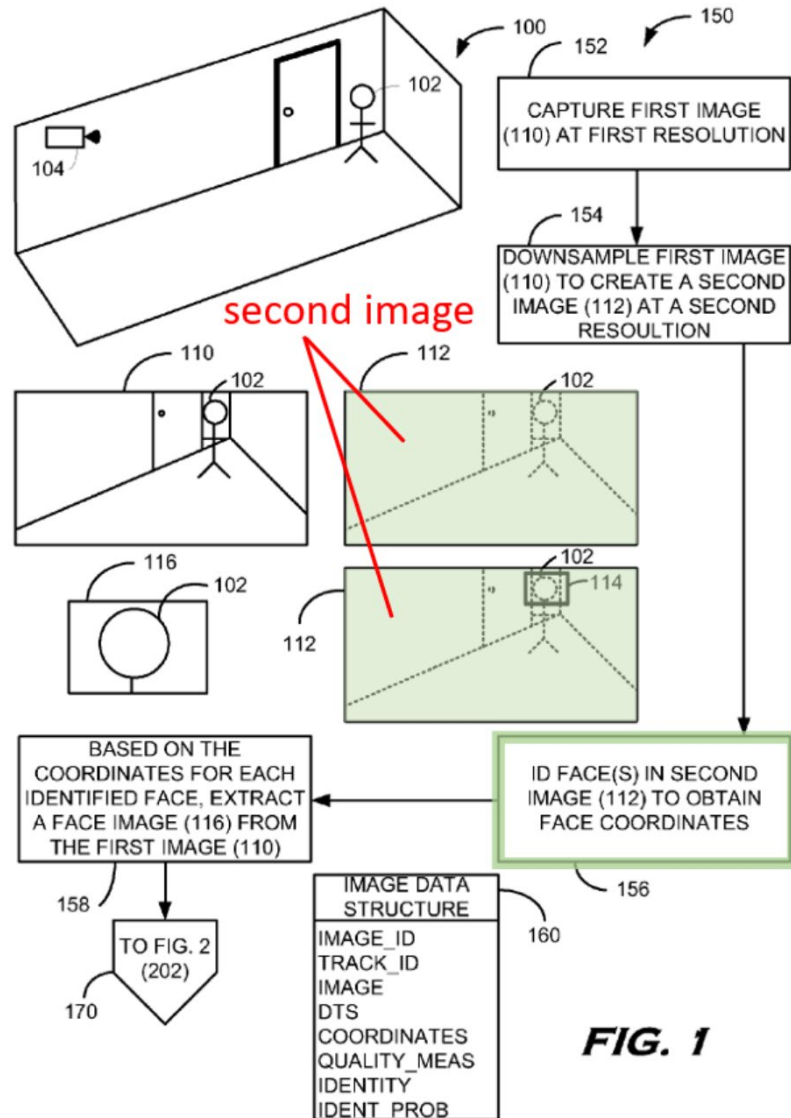
capture image data”).

c) Claim 1.2: “an object detection system that detects objects of interest within the image data and extracts portions of frames that include the detected objects of interest and sends the extracted portions of the frames”

192. In my opinion, Bedros discloses or suggests 1.2. Bedros’s method 150 includes detecting an individual’s face within a captured image (“*detects objects of interest within the image data*”), extracting a face image from the captured image (“*extracts portions of frames that include the detected objects of interest*”), and sending the face image to the facial recognition program (“*sends the extracted portions of the frames*”). See *Bedros*, 2:36-43, 3:19-55, 7:19-25.

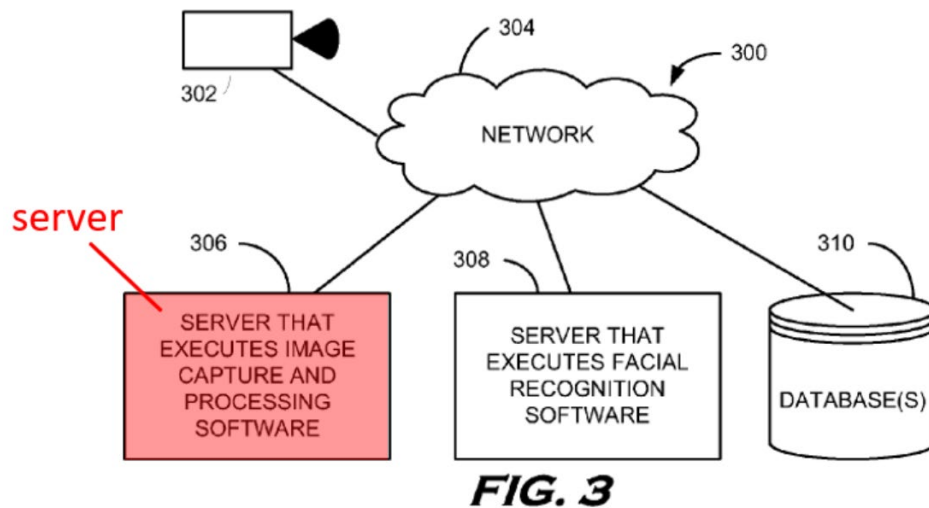
193. In operation 152 of method 150 “a first image 110 [is] acquired by a camera 104 that monitors a scene 100.” *Bedros*, 2:44-46. A POSITA would have understood that image 110 constitutes a frame of image data at least because image 110 is obtained through “an image or video stream received from the camera 104.” *Id.*, 2:61-62, 4:8-11, 4:20-22. Method 150 proceeds to “downsampling 154” where the resolution of first image 110 is downsized to form second image 112. *Bedros*, 3:8-27. Thus, first and second images 110 and 112 contain the same content (e.g., both depict the same image data) but have different resolutions. “[T]he image 112 is [] processed to **locate 156 one or more faces present in the image 112**, if any, and to obtain face coordinates within the image 112.” *Id.*, 3:28-31. Accordingly, Bedros discloses locating a face in an image (“*detects objects of interest within the image*

data”). Figure 1 below shows second image 112 (green shade) and capture operation 156 (green box).

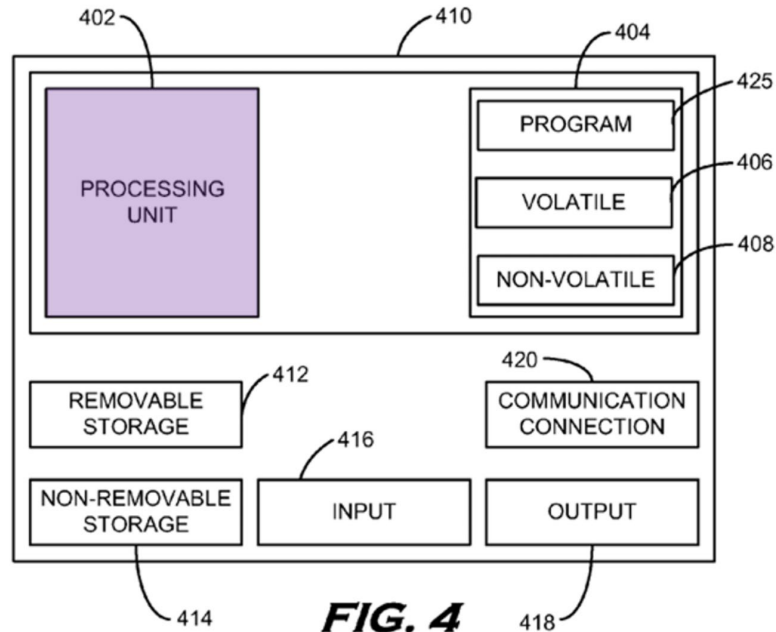


194. Method 150 is implemented by system 300 illustrated in Figure 3. *Bedros*, 8:13-16. In particular, server 306 of system 300 “executes image capturing and processing software, such as software that implements one or both of the methods 150 and 200.” *Id.*, 8:20-22. Specifically, “server 306 receives the images or

video stream in a first resolution, including a first image,” “downsample[s] an entirety of the first image to create a second image at a second resolution,” and “search[es] the second image to identify faces present in the second image.” *Id.*, 8:32-42, Fig. 3 (annotated below).



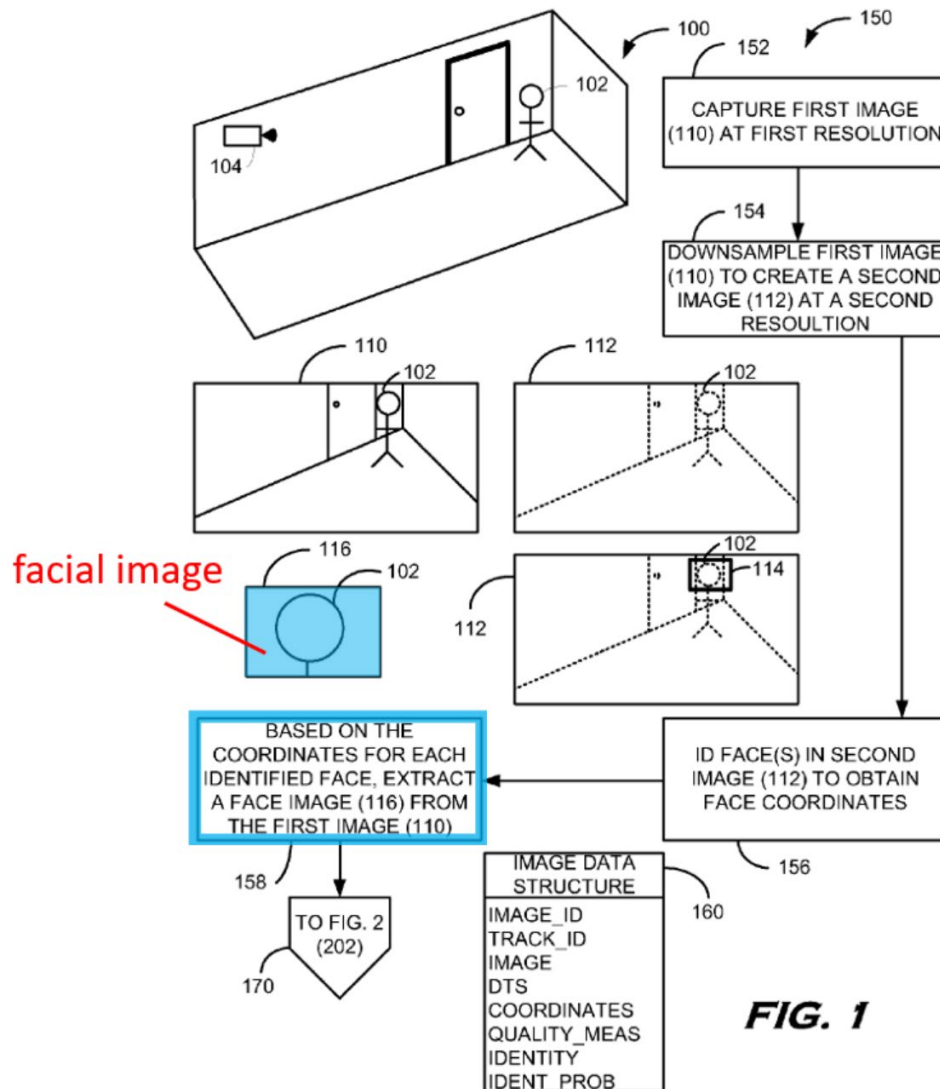
195. Furthermore, Bedros explains that server 306 may be a computing device as depicted in Figure 4, annotated below. *Id.*, 8:50-56. The computing device includes a processor (purple) and memory computer-readable medium for executing instructions. *Id.*, 8:57-9:2. The processor executes instructions stored on the computer-readable medium to perform the functions of methods 150 and/or 200. *Id.*, 9:26-34. Accordingly, it is my opinion that Bedros teaches or suggests server 306 (“*object detection system*”) locating one or more faces (“*detects objects of interest*”) present in the image 112 (“*within the image data*”).



196. Bedros further discloses extracting a face of the person 102 from first image 110. To the extent that PO argues that identifying the face in image 112 does not “detect[] objects of interest within the image data,” because image 110 is the image data, a POSITA would have recognized that extracting the face image 116 from image 110 would have regardless detected the face in the image data.

197. After faces are identified in operation 156, “method 150 continues by **extracting 158 a face image 116** from the first image 110” such that “face image 116 includes the face of the person 102 include[d] in the scene 100 and the captured image 110.” *Bedros*, 3:42-46. Facial image 116 is a portion of first image 110 because only box 114 containing the person’s face is kept in facial image 116 while the rest of second image 112 is excluded. *Id.*, Fig. 1. Annotated Figure 1 below shows facial image 116 (blue shade) and extracting operation 158 (blue box).

Method 150 is implemented using server 306 depicted in Figure 3. *See Bedros*, 8:13-42.



198. Accordingly, it is my opinion that Bedros discloses or suggests server 306 (“object detection system”) that extracts a face image 116 from the first image 110 (“extracts portions of frames”) that includes the face of person 102 (“that include the detected objects of interest”).

199. Bedros discloses sending facial image 116 to a facial recognition program. *See Bedros*, 4:23-29, 7:26-39. Specifically, Bedros explains that method 200 **“includes submitting 230 the facial image 116 to the facial image queue.”** *Bedros*, 7:19-21. Method 200 is a continuation of method 150 and is implemented using the system depicted in Figure 3. *Id.*, 8:46-49 (“server 306 [of Figure 3] software may cause the **facial images 116 to be submitting each facial image 116** to the server 308 that executes the facial recognition software”); *see also, id.*, 4:24-26, 8:13-20.

200. Accordingly, it is my opinion that Bedros discloses or suggests submitting facial image 116 to the facial recognition software (“sends the extracted portions of the frames”).

201. Bedros’s extraction processes operate in the same manner as the processes of creating and sending “cropped images” described in the ’321 patent. Bedros’s extraction process removes the portion of image 110 that is outside of box 114, leaving only a region of the frame that contains the identified face of person 102. *See Bedros*, 3:42-46, Fig. 1. In the same way, the “object detection system 108-1 [of the ’321 patent] can create cropped frames” that “include the detected objects of interest.” *’321 patent*, 9:3-5.

202. By my review, Bedros discloses or suggests the same or equivalent structure to that described by the ’321 patent for performing the claimed functions.

Bedros' server 306 contains a processor and a memory and is equivalent to the local video analytics system 112 of the '321 patent. *Bedros*, 8:50-9:11. Bedros further discloses that a computer program capable of performing the disclosed methods “may be encoded as a computer processor executable instruction set stored on a computer-readable medium.” *Bedros*, 9:30-34. A POSITA would have understood that these computers would have required the same type of software programming to perform the recited detection, extracting, and sending functions.

203. Accordingly, it is my opinion that Bedros discloses or suggests server 306 (“object detection system”) locating one or more faces (“detects objects of interest”) present in the image 112 (“within the image data”), extracting a face image 116 from the first image 110 (“extracts portions of frames”) that includes the face of person 102 (“that include the detected objects of interest”), and submitting facial image 116 to the facial recognition software (“sends the extracted portions of the frames”).

d) Claim 1.3.1: “a video analytics system that receives only the extracted portions of the frames of the image data, transmitted over a network, from the object detection system, the image data received by the video analytics system corresponding to objects of interest detected by the object detection system”

204. In my opinion, Bedros discloses or suggests 1.3.1. As explained in §VIII.C.2.c, facial images 116 (“extracted portions of the frames of the image data”) contain the face of person 102 (“objects of interest detected by the object detection

system”), and “server 306 software may cause the facial images 116 to be submit[ed]...to the server 308 that executes the facial recognition software.” *Bedros*, 8:46-49; *see also, id.*, 3:19-41. A POSITA would have understood that facial image 116 is submitted from server 306 (“object detection system”) to server 308 (“video analytics system”) through network 304 (“network”) because servers 306 and 308 are connected through network 304 and “facial images are submitted to...the facial recognition program **over a network.**” *Bedros*, Fig. 3, 11:1-3.

205. Bedros further explains that server 308 may be a computing device as depicted in Figure 4, annotated above. *Id.*, 8:50-56. The computing device includes a processor and memory or computer-readable medium for executing instructions. *Id.*, 8:50-9:2. The processor executes instructions stored on the computer-readable medium to perform the functions of methods 150 and/or 200. *Id.*, 9:26-34.

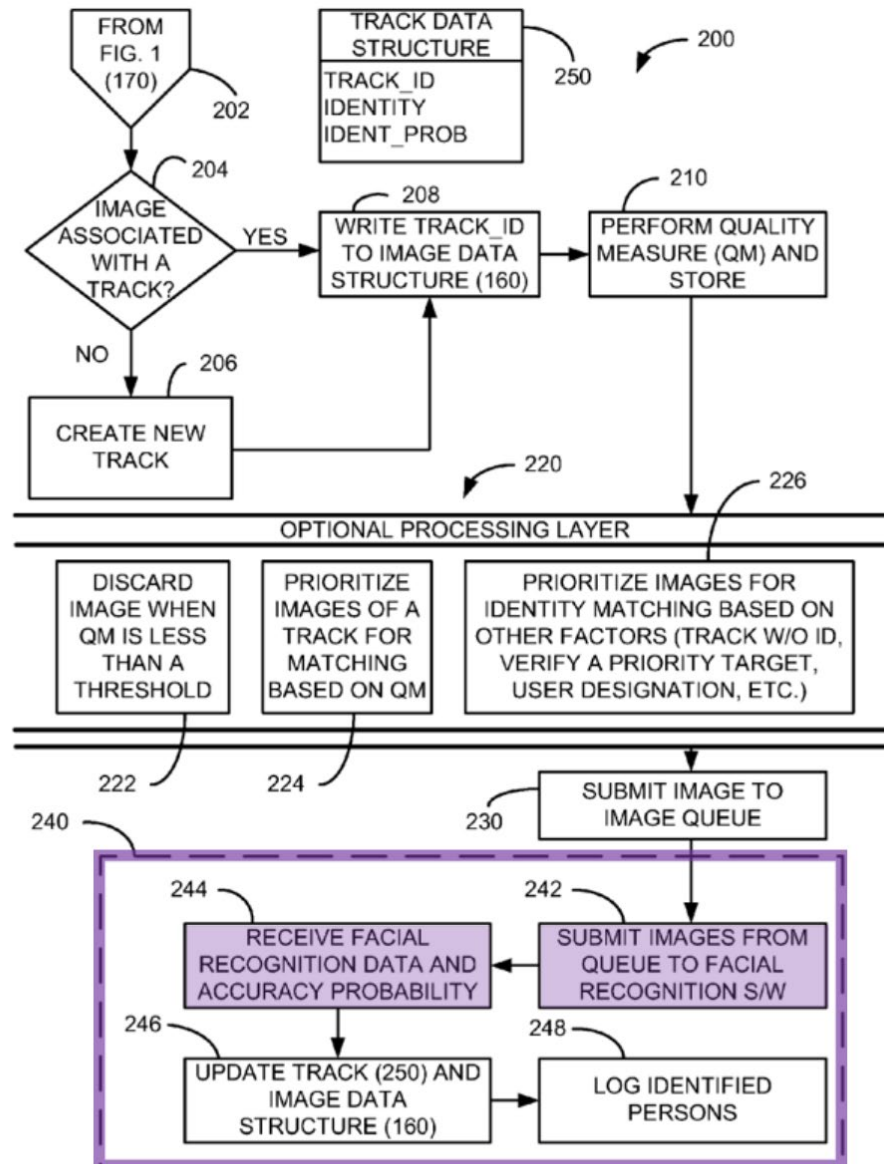
206. Bedros discloses or suggests the same or equivalent structure to that described by the ’321 patent for performing the claimed functions. Bedros’s server 308 includes a processor and memory and is at least equivalent to the ’321 patent’s disclosure that the remote analytics system 122, which can be a node on a network. A POSITA would have understood that these systems would have required associated programming for receiving the image data because Bedros discloses that a computer program capable of performing the disclosed methods “may be encoded as a computer processor executable instruction set stored on a computer-readable

medium.” *Bedros*, 9:30-34.

207. Accordingly, it is my opinion that Bedros discloses or suggests server 308 (“a video analytics system”) receiving facial image 116 (“the extracted portions of the frames of the image data”), transmitted over network 304 (“network”), from server 306 (“object detection system”), the facial image 116 containing face of a person 102 (“the image data received by the video analytics system corresponding to objects of interest detected by the object detection system”).

e) Claim 1.3.2: “the video analytics system analyzing the objects of interest”

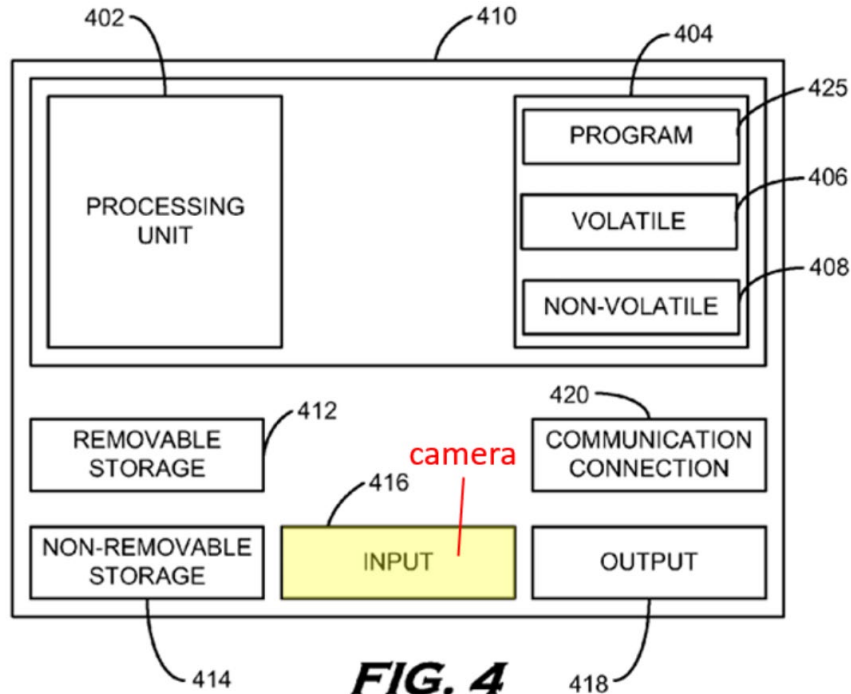
208. In my opinion, Bedros discloses or suggests 1.3.2. As discussed in §VIII.C.2.c, “server 306 software may cause the facial images 116 to be submit[ed to] each facial image 116 to **the server 308 that executes the facial recognition software.**” *Bedros*, 8:46-49; *see also, id.*, 3:19-41. Specifically, facial image 116 is submitted “to the facial image queue” and subsequently submitted to “queue processing 240,” which includes “submitting 242 images from the queue to the facial recognition program and **receiving 244 facial recognition data** in response.” *Id.*, 7:19:21, 7:27-29, Fig. 2 (annotated below showing operations 240 (purple box), and 242 and 244 (purple shaded)). Bedros discloses that server 308 executes the facial recognition software to find “**an identity of a person** of the facial image 116” and provide “data indicative of [the] identity” as facial recognition data in operation 244. *See id.*, 7:30-35, 8:46-49, Fig. 3.

**FIG. 2**

209. Accordingly, it is my opinion that Bedros discloses or suggests server 308 (“video analytics system”) executing the facial recognition software on facial image 116 to find the identity of person 102 (“analyzing the objects of interest”).

f) Claim 1.4.1: “wherein the video analytics system receives the image data from multiple object detection systems after being transmitted over the network”

210. In my opinion, Bedros discloses or suggests 1.4.1. As explained above in §VIII.C.2.d, Bedros discloses server 308 (“video analytics system”) receives facial image 116 (“image data”) over a network from server 306 (“object detection system”). With reference to Figure 4 (annotated below), Bedros further discloses that the computing device depicted in the figure “may be...server[] 306.” *Bedros*, 8:55-56. Each of the computing devices can be a “computer 410,” which includes “an input 416 [that] may include a camera that connects directly to the computer 410.” *Id.*, 9:9-17, Fig. 4 (input 416 annotated in yellow).



211. To the extent that PO argues, or the Board finds, that Bedros does not

disclose or render obvious this limitation, a POSITA would have found it obvious to utilize multiple computers 410 with connected cameras to send image data to server 308, implementing a facial recognition system that monitors multiple locations. For example, Bedros explains that prior art systems were used to identify “individuals and their locations,” but that data processing was prohibitive because they used both high- and low-resolution cameras to implement facial recognition. *Bedros*, 1:6-23, 1:43-57. However, a POSITA would have been motivated to utilize Bedros’ multiple computers with cameras to simultaneously monitor several scenes of interest (e.g., hallways and work areas) that are associated with different levels of human movement (e.g., remain or do not remain in the same space for extended periods), thus maintaining the ability to monitor more than one location. *See Bedros*, 5:3-11, 5:26-33. By monitoring multiple locations, a POSITA would have recognized that the usefulness of the surveillance system is increased because it meets the need of simultaneously monitoring multiple indoor/outdoor areas, such as rooms, hallways, porches, patios, yards, etc.

212. A POSITA would have reasonably expected success in implementing multiple computers (i.e., servers 306) because it represents nothing more than the combination of known elements used according to their ordinary and expected operations. This combination was well within the skill of a POSITA and would have required nothing more than the application of known, conventional techniques

(utilizing more than one computer to send image data to a centralized analysis system). Indeed, Bedros explains that server 308 may be a “service” utilized by the facial recognition system over the Internet. *Bedros*, 8:27-31. A POSITA would have thus been motivated to send multiple different image data from multiple locations to server 308 to cost-effectively utilize the service.

213. Accordingly, it is my opinion that Bedros discloses or suggests server 308 (“video analytics system”) receiving facial image 116 (“image data”) from multiple servers 306 (“object detection system”) over network 304 (“transmitted over the network”).

g) Claim 1.4.2: “the video analytics system is operated by a different business entity than the one or more surveillance cameras”

214. In my opinion, Bedros discloses or suggests 1.4.2. As explained in §VIII.C.2.d, Bedros’s server 308 is a “video analytics system.” Bedros further discloses “the facial recognition software that executes on the one or more servers 308 may be **accessible as a service over a private network, such as a virtual private network (VPN), or public network, such as the Internet.**” *Bedros*, 8:27-31. Given that disclosure, it would have been obvious to a POSITA that Bedros’ system could be used where the surveillance cameras are owned and operated by one business entity (*e.g.*, a store or restaurant) that relies on a facial recognition service operated by a different business entity. With this arrangement, the business using the

security cameras could likely avoid having to create and maintain its own database with reference images of individuals to enable facial recognition. By the time of the alleged invention, a POSITA would have been well aware of the concept of software as a service (SaaS) and how it could be applied here to outsource facial recognition.

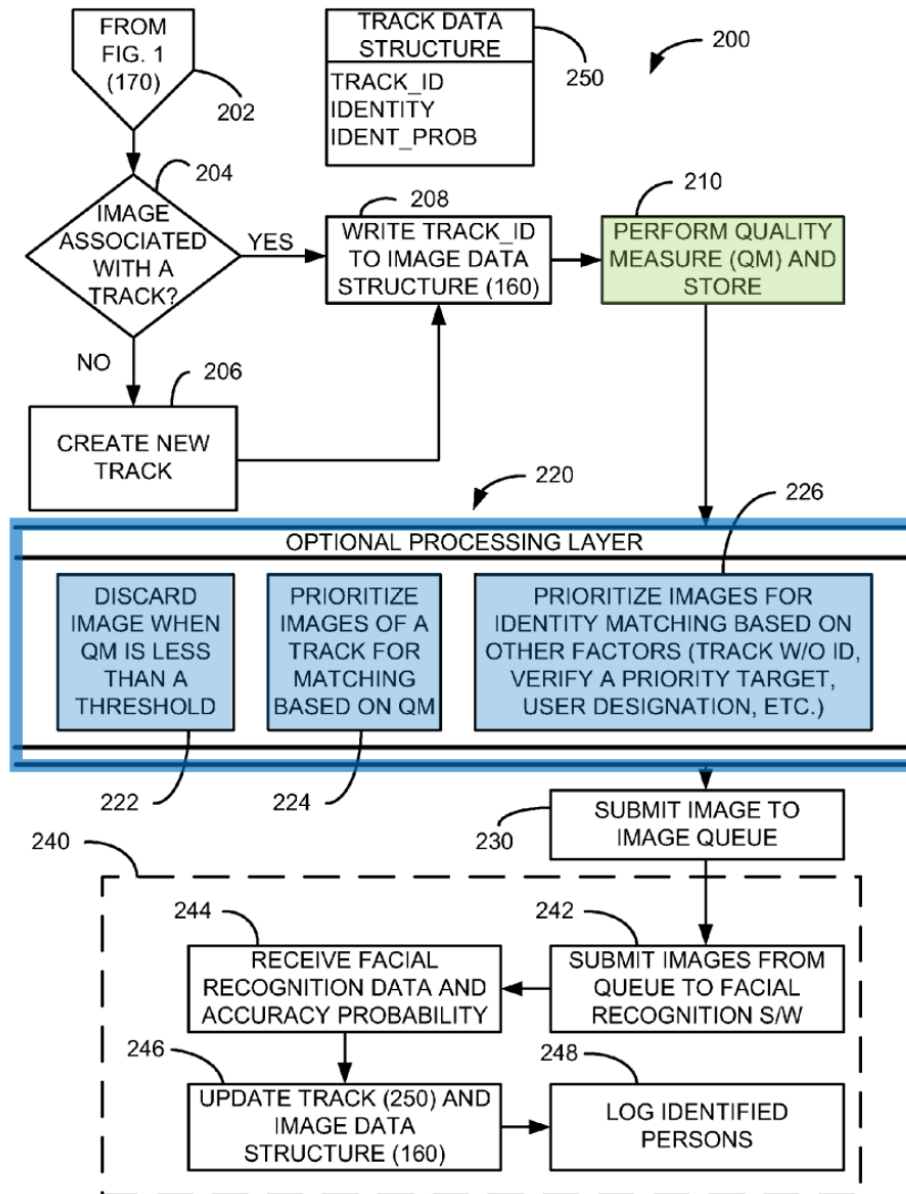
215. Accordingly, it is my opinion that Bedros discloses or suggests this limitation.

h) Claim 1.5.1: “wherein the image data corresponding to objects of interest detected by the object detection system are ranked by the object detection system according to factors including close proximity of the objects of interest to the one or more surveillance cameras, picture quality of the objects of interest, and/or contrast of the objects of interest to create a ranking and”

216. In my opinion, Bedros discloses or suggests 1.5.1. Method 200 includes an “optional processing layer 220” that can “filter or prioritize facial images 116 for submission to a facial recognition program.” *Bedros*, 6:23-26. Optional processing layer 220 filters and/or prioritizes facial images 116 (“image data corresponding to objects of interest detected by the object detection system”) **before** they are being submitted to an image queue at operation 230 and **before** being processed for facial recognition at operation 240. *See Bedros*, 6:59-64, 7:19-30, Fig. 2. Therefore, the prioritization process on facial image 116 is performed in server 306 (“object detection system”) before the images are submitted to server 308 (“video analytics system”) for facial recognition.

217. Optional processing layer 220 (blue box) includes at least three processing elements 222, 224, and 226 (blue shaded) for prioritizing facial images. *See Bedros*, 6:37-7:18, Fig. 2 (annotated below). Before optional processing layer 220 receives facial images 116, “a quality measure is applied [at] 210 to the facial image 116.” *Bedros*, 6:4-5. “The quality measure evaluates various features of the facial image 116 such as **image sharpness, an amount of present texture in the image**, and image brightness.” *Id.*, 6:6-9. A POSITA would have understood that image sharpness and the amount of present texture in the image are “factors” referring to the picture quality of the image.

218. *Bedros* further discloses that “a score maybe assigned to each factor” and/or “a composite score may be assigned” to facial image 116. *Bedros*, 6:13-16. Claims 3, 11, and 17, which recite features including sharpness, texture, and brightness factors. *Id.*, 10:37-45, 11:36-44, 12:47-55. *Bedros* explains that “[a] best quality measure may be a high value or a low value depending on the particular embodiment.” *Id.*, 6:20-22. Therefore, a POSITA would have understood *Bedros* to disclose or suggest facial images 116 are “ranked” based on the factors because the images are assigned different quality measure scores based on various factors, including image sharpness and picture quality.

**FIG. 2**

219. Bedros further discloses that processing element 222 discards facial images 116 that have quality measure “less than a threshold value” (*Bedros*, 6:39-55) and that processing element 224 prioritizes facial images 116 (“to create a ranking”) “based on the quality measure” and “facial images 116 placed in the queue may have a priority assigned to them” (*id.*, 6:56-7:3). The priority determines which

facial images 116 are submitted first into image queue at operation 230. *See id.*

220. Accordingly, it is my opinion that Bedros discloses or suggests that facial images 116 (“the image data corresponding to objects of interest detected by the object detection system”) are given quality measure scores (“are ranked by the object detection system”) based on factors including image sharpness and an amount of present texture in the image (“according to factors including close proximity of the objects of interest to the one or more surveillance cameras, picture quality of the objects of interest, and/or contrast of the objects of interest to create a ranking”) and/or prioritized based on the quality measure of the images by optional processing layer 220 of server 306 (“to create a ranking”).

i) Claim 1.5.2: “the ranking is saved in metadata with the image data and”

221. In my opinion, Bedros discloses or suggests 1.5.2. As explained in §VIII.C.2.h, facial images 116 are ranked and prioritized by assigning quality measure scores based on the quality measure scores. Bedros explains that “the quality measure score or scores...**may then be stored, such as to a record of the image data structure 160 of the facial image 116.**” *Bedros*, 6:15-19. “The quality measure scores may be numeric, alphabetical, or some other data type.” *Bedros*, 6:19-22. A POSITA would have understood that the image data structure 160, including facial image 116, includes metadata of facial image 116 (e.g., quality measure scores “QUALITY_MEAS”) because the quality measure scores describe

the properties of the image data. Metadata is well-known to a POSITA as “data that describes other data.” *See* EX1010, 688; EX1011, 779; EX1012, B14. Moreover, it was common practice at the time to include metadata with image data to provide context for the image. For example, Deever ’150 (EX1014)²² discloses it was well-known to store metadata in an image file describing features of the image. *See Deever ’150* at 6:10-15 (“[t]he focus distance...can be stored as metadata in the image file”), 8:48-56 (“computed global and local motion information can be stored as metadata”), 13:37-56 (“key video frames can be indicated by storing metadata associated with [a] digital video sequence” such as “a list of video frames...identified to be key video frame” and “thumbnail images”).

222. Accordingly, it is my opinion that Bedros discloses or suggests that the quality measures of facial images 116 (“ranking”) are stored to image data structure 160 with the facial image 116 (“metadata with the image data”).

j) Claim 1.5.3: “an object analysis system of the video analytics system reads the rankings within the metadata and decides whether or not to analyze the image data”

223. In my opinion, Bedros discloses or suggests 1.5.3. As explained in §Claim 1.5.2: “the ranking is saved in metadata with the image data and”VIII.C.2.i, optional processing layer 220 in server 306 applies quality measure scores and/or

²² *See* footnote 18.

prioritization results (“ranked...to create a ranking”) to facial images 116 and stores the quality measure scores and prioritization results in a record of the image data structure 160 (“ranking is saved in metadata with the image data”).

224. As explained in §VIII.C.2.f, Bedros explains servers 306 submit the images to server 308 that “executes the facial recognition software.” *Bedros*, 8:46-49. Specifically, after processing in optional processing layer 220 is complete, facial images 116 are submitted to the facial image queue. *Id.*, 7:19-21. Bedros further explains that queue processing “includes submitting 242 images from the queue to the facial recognition program and receiving 244 facial recognition data in response.” *Id.*, 7:26-30.

225. Bedros discloses that “submitting 230 the facial image 116 to the queue may include **removing** and reordering facial images 116 within the queue awaiting submission to the facial recognition program.” *Bedros*, 7:21-25. Thus, certain facial images 116 may be removed from the queue of images that await processing by the facial recognition software as subsequent images arrive from any of the servers 306. A POSITA would have understood that this sort of queue processing—reordering the queue as new images arrive from different places—would have taken place at server 308 (“video analytics system”), where all quality scored images would be present together for processing. A POSITA would have thus understood that the facial image queue resides and is managed on server 308, where facial recognition

take place because ranked images are received from multiple servers 306. Moreover, because the server 308 performs the functions, Bedros discloses or suggests the same or equivalent structure to that described by the '321 patent for performing the claimed functions, as I explained above in paragraph 206.

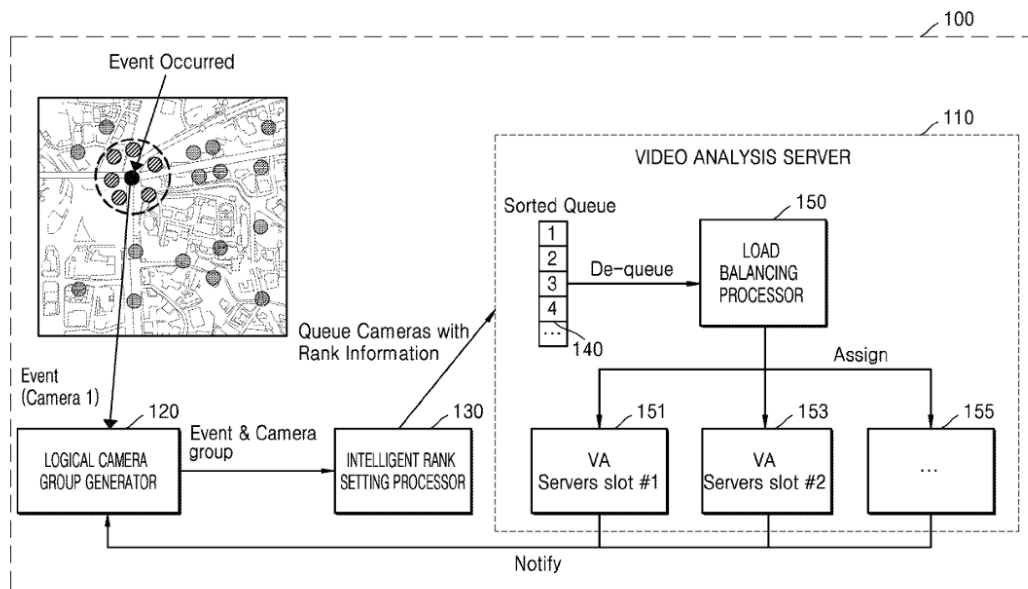
226. To the extent that PO argues or the Board finds, that Bedros does not teach processing the facial image queue at server 308, a POSITA would have been motivated to do so at least because (1) all facial images are ultimately processed at server 308 and (2) servers 306 would not know what ranking other servers 306 assign to the images they send. Processing the queue at server 308 would be more efficient and reduce bandwidth by relieving servers 306 from reading the queue to reorder or remove facial images from it. Performing queue processing on images data at the server receiving information was well-known in the art, as I explain below.

227. For example, Laganier '461 discloses that the “first video data, either with or without preprocessing, is also provided via gateway 104 and WAN 106 to central server 108,” and “under the control of the central server 108, portions of the retrievably stored portion of the video data, which are other than the identified subset of the portion of the video data, are deleted from the memory storage device 118.”

Laganier '461, ¶¶24, 25.

228. As another example, Gali²³ discloses that an “intelligent rank setting processor 130 provides a queue 140 included in the intelligent video analysis server 110” with “information about the ranks of the surveillance cameras.” *Gali*, ¶¶53-54; *see also id.*, 76-77, 87-88, 94-95, Fig. 1. The load balancing processor 150 that is within video analysis server 110 “receives data transmitted from the queue 140” and assigns slots “for processing data transmitted from the surveillance camera.” *Id.* Figure 1 of Gali is reproduced below.

FIG. 1



229. As a further example, Deever '965²⁴ discloses a “video playback system” for forming and displaying video summaries that are “stored as metadata.”

²³ See footnote 18.

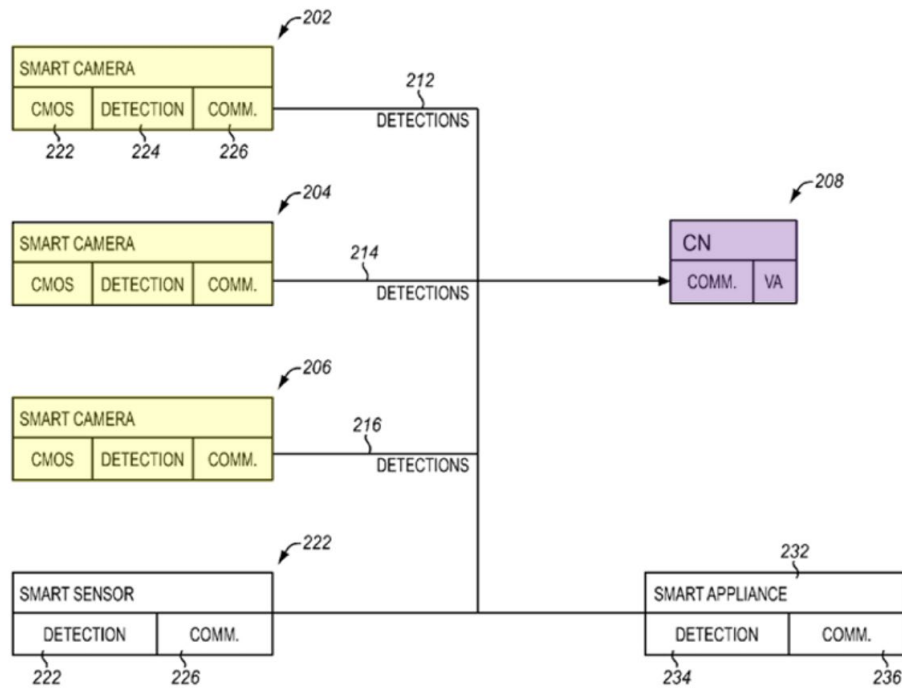
²⁴ See footnote 18.

Deever '965, 14:45-15:35; *see also, id.*, 16:47-17:33. Specifically, *Deever* '965 discloses a processor for assembling multiple key video snippets to form a video summary. *Id.*

230. A POSITA would have been further motivated to separate and delineate tasks performed at servers 306 and servers 308 to streamline operation and to reduce transmission bandwidth—a technique well known before the application for the '321 patent was filed. For example, Yang²⁵ discloses a wireless smart camera network that utilizes an “end-to-end video analytics framework” in which smart cameras “identify and send out potential matches” and “the central node or the cloud can perform” further “real time analytics.” *Id.*, ¶¶20-21. Specifically, Yang’s smart cameras “preprocess the data into **predetermined semantic forms** and send only sparse image data and metadata to the central node.” *Id.*, ¶14. “The semantic form can be a **face** in surveillance, and **cars or license plates** in transportation.” *Id.*, ¶19. Fig. 2 illustrates Yang’s wireless smart camera network. *Id.*, ¶31; Fig. 2.

²⁵ *See* footnote 18.

FIG. 2



231. The smart cameras “detect[s] objects of interest,” “compar[es] these objects to the background,” and “remov[es] the background leaving only that portion of each image of the video sequence that shows the object.” *Id.*, ¶26. The part of image that includes the object “is transmitted [to the central node] and the rest of the image is removed.” *Id.*

232. Subsequently, “the central node or the cloud can perform state of the art high precision algorithms for detection, tracking or recognition to deliver real time analytics without overloading the network and the computation capability.” *Yang*, ¶20.

233. A POSITA would have had a reasonable expectation of success of

performing queue processing at server 308 because such processing represents nothing more than the combination of known elements used according to their ordinary and expected operations. This modification was well within the skill of a POSITA and would have required nothing more than the application of known, conventional techniques (utilizing more than one computer to send image data to a centralized analysis system).

234. Furthermore, to reorder or remove images from the queue, a POSITA would have recognized that queue processing must first read the quality scores of the images in the queue. For example, Deever '965 (EX1015) states "metadata indicates which video frames are included in the video summary." *Deever '965*, 14:48-60. A POSITA would have also understood that the removed facial images 116 would not be further processed using the facial recognition program ("*whether or not to analyze the image data*"). This selection of facial images 116 for removal is "based on processing outcomes within the optional processing layer 220." *Bedros*, 6:49-55; 7:24-25. Therefore, analysis of facial images 116 ("the image data") is determined by the processing outcomes (e.g., quality measure score and/or prioritization processing) ("ranking"). A POSITA would have further understood that the methods executed on server 308 to perform reading the metadata and reordering or deleting image data was a subset of the overall methods performed by server 308. As such, a POSITA would have recognized the subset of methods as the

claimed “object analysis system.”

235. Accordingly, it is my opinion that Bedros discloses or suggests queue processing and/or recursion module (“object analysis system”) of server 308 (“video analytics system”) prioritizing facial images within the queue of images (“decides whether or not to analyze the image data”) based on a prioritization rule (“reads the ranking within the metadata”).

3. Independent Claim 13

236. In my opinion, Bedros discloses or suggests Claim 13 for substantially the same reasons as claim 1, as discussed in §VIII.C.2 because my review reveals the claims are substantively identical, and Bedros further discloses steps of a method “for analyzing video in a video surveillance system” described above. *Bedros*, Abstract; *compare* claim limitations 1.pre-1.5.3, *with* claim limitations 13.pre-13.5.3.

237. By my review, claim 13.pre is substantially similar to 1.pre, except that it recites a “method” instead of a “system.” As discussed in §VIII.C.2, Bedros describes the recited method because it discloses steps to perform the claimed video analysis of claim 13. *Bedros*, Figs. 1-2.

238. The table below correlates the reasons why Bedros discloses or suggests the limitations of claim 1 with limitations 13.1-13.5.3.

Claim 13		
Limitation	Substantially Similar Limitation(s)	Reference
13.1	1.1	VIII.C.2.b
capturing image data from one or more surveillance cameras;	one or more surveillance cameras that capture image data;	
13.2.1	1.2	§VIII.C.2.c
detecting objects of interest within the image data via an object detection system; the object detection system extracting portions of frames that include the detected objects of interest and transmitting the extracted portions of the frames of the image data over a network,	an object detection system that detects objects of interest within the image data and extracts portions of frames that include the detected objects of interest and sends the extracted portions of the frames; and	
13.2.2	1.3.2	§VIII.C.2.e
the image data transmitted by the object detection system corresponding to objects of interest detected by the object detection system; and	the image data received by the video analytics system corresponding to objects of interest detected by the object detection system, the video analytics system analyzing the objects of interest;	
13.3	1.3.1, 1.4.2	§§VIII.C.2.d, VIII.C.2.g
a video analytics system, which is operated by a different business entity than the one or more surveillance cameras, only receiving the extracted portions of the frames of the image data from the object detection system and analyzing the objects of interest;	a video analytics system that receives only the extracted portions of the frames of the image data, transmitted over a network, from the object detection system, the video analytics system is operated by a different business entity than the one or more surveillance cameras; and	
13.4	1.4.1	§VIII.C.2.f

the video analytics system further receiving image data from other object detection systems;	wherein the video analytics system receives the image data from multiple object detection systems after being transmitted over the network and	
13.5.1	1.5.1	
the object detection system ranking the image data corresponding to objects of interest detected by the object detection system according to factors including close proximity of the objects of interest to the one or more surveillance cameras, picture quality of the objects of interest, and/or contrast of the objects of interest to create a ranking and	wherein the image data corresponding to objects of interest detected by the object detection system are ranked by the object detection system according to factors including close proximity of the objects of interest to the one or more surveillance cameras, picture quality of the objects of interest, and/or contrast of the objects of interest to create a ranking and	§VIII.C.2.h
13.5.2	1.5.2	
the ranking is saved in metadata with the image data; and	the ranking is saved in metadata with the image data and	§VIII.C.2.i
13.5.3	1.5.3	
an object analysis system of the video analytics system reading the rankings within the metadata and deciding whether or not to analyze the image data.	an object analysis system of the video analytics system reads the rankings within the metadata and decides whether or not to analyze the image data.	§VIII.C.2.j

4. Claims 2 and 14: “wherein the object detection system creates cropped images from the image data that include the detected objects of interest and sends the cropped images to the video analytics system.”

239. Claims 2 and 14 depend from claims 1 and 13, respectively, which as I explained above, Bedros discloses or suggests. In my opinion, Bedros further

discloses or suggests claim 2's additional limitations.

240. Bedros discloses extracting facial image 116 ("cropped images") from first image 110 ("image data"). *Bedros*, 3:42-46. As I explained in §VIII.C.2.c, the face of the person 102 is "*the detected objects of interest*" and method 150 extracts "a face image 116 from the first image 110" such that "face image 116 includes the face of the person 102." *Bedros*, 3:42-46. Facial image 116 is a cropped image created from first image 110 because it is created by identifying a box 114 that contains a person's face and altering the boundaries of first image 110 based on the coordinates of box 114 such that areas outside of box 114 is excluded. *Id.*, 3:28-55, Fig. 1; EX1016 (Dictionary of Contemporary Photography), p. 44. As I explained in §VIII.C.2.c, the extraction process is performed by server 306 ("object detection system"), which sends facial images 116 ("cropped images") to server 308 ("video analytics system"). *Bedros*, 8:46-49. Accordingly, Bedros discloses or suggests claim 2's additional limitation.

241. By my review, claim 14 is substantially similar to claim 2, except that it recites a "method" instead of a "system." Thus, for the reasons I explained above for claim 2, it is my opinion that Bedros discloses or suggests claim 14.

5. Claims 3 and 15: "wherein the image data include video data and the object detection system provides frames from the video data that include the detected objects of interest and sends the frames to the video analytics system as the image data."

242. Claims 3 and 15 depend from claims 1 and 13, respectively, which as I

explained above, Bedros discloses or suggests. In my opinion, Bedros further discloses or suggests claim 3's additional limitation.

243. A POSITA would have understood “the image data include video data” as recited to mean that the image data is included in video data as frames of video data. As I explained in §§VIII.C.2.b-VIII.C.2.c, Bedros discloses or suggests server 306 (“object detection system”) extracting cropped portions of frames of video data (“portions of the frames”) containing a person’s face (“detected objects of interest”) and sending those frames to server 308 (“video analytics system”) for subsequent processing. A POSITA reading claim 3 in light of claim 1, from which it depends, would have understood the recited “provides frames from the video data that include the detected objects of interest” as referring back to claim 1’s recitation of “extracted portions of the frames.” Accordingly, to the extent that “frames” is interpreted as “portions of frames” (which it should be), Bedros discloses or suggests claim 3.

244. By my review, claim 15 is substantially similar to claim 3, except that it recites a “method” instead of a “system.” Thus, for the reasons I explained above for claim 2, it is my opinion that Bedros discloses or suggests claim 15.

6. Claims 4²⁶, 6, and 16: “wherein the video analytics system includes an object analysis system that analyzes the image data corresponding to objects of interest detected by the object detection system by comparing the image data to reference images to identify the objects of interest, by recognizing patterns in the image data, and/or by applying image processing functions, neural network functions, deep learning functions, and/or object classification functions to the image data.”

245. By my review, claims 4, 6, and 16 depend from claims 1 and 13, respectively, which as I explained above, Bedros discloses or suggests. In my opinion, Bedros further discloses or suggests claim 4’s additional limitation.

246. As I explained in §VIII.C.2.j, Bedros discloses or suggests a subset of methods (“object analysis system”) in server 308 that operate on image data in the image queue (“objects of interest detected by the object detection system”).

247. Bedros discloses that “facial recognition program is a program that operates to receive a digital image including a face and to respond with an identity of an individual matching the face from the received image.” Bedros, 7:63-66. As discussed in §VIII.C.2.c, facial image 116 contains person 102. The facial recognition program determines the identity of person 102 by “compar[ing] facial features from the received image with a facial database” which “typically includes a database of known facial feature landmarks of a known subject faces.” Id., 7:66-8:2.

²⁶ I see that “*an object analysis system*” appears in both claims 1 and 4, and treat it as the same.

Accordingly, Bedros discloses or suggests or suggests identifying the person in the image (“identifying objects of interest”). Examples of facial feature landmarks include characteristics of components of a face, such as “eyes, nose, cheekbones, jaw, facial scars.” *Id.*, 8:3-5. The comparison process utilizes several image analysis algorithms. *Id.*, 8:5-12. Thus, Bedros discloses or suggests “comparing the image data to reference images” and “applying image processing functions.”

248. By my review, claims 6 and 16 are substantially similar to claim 4, except that claim 16 recites a “method” instead of a “system” and both claims 6 and 16 recite that the “video analytics system” performs the recited function of “comparing the image data to references images.” As discussed above, claim 16 contains a topographical errors because it recites “by recognizing” and “by applying” in its recited methods. To the extent it is found that claim 16 does not have a typo, Bedros teaches it regardless because Bedros teaches each one of the three parallel clauses of “comparing,” “recognizing,” and “applying.” To the extent claim 16 requires performance of additional functions, Bedros further teaches “applying image processing functions.” For example, Bedros discloses performing facial recognition program and receiving facial recognition data in response. *Bedros*, 7:26-39. As I explained above, Bedros’ subset of methods is part of server 308 (“video analytics system”). Accordingly, it is my opinion that a POSITA would have understood that server 308 performed the methods. *Id.* For the reasons discussed

above for claim 4, Bedros discloses or suggests claim 16.

7. Claims 5 and 18: “wherein the object detection system is integrated within the one or more surveillance cameras.”

249. By my review, claims 5 and 18 depend from claims 1 and 13, respectively, which as I explained above, Bedros discloses or suggests. In my opinion, Bedros further discloses or suggests claim 5’s additional limitation.

250. As I explained in §§VIII.C.2.b-VIII.C.2.c, Bedros discloses server 306 (“*object detection system*”) and at least one camera 104 (“*one or more surveillance cameras*”). Camera 104 is equipped with “suitable optics to capture images of scene as needed based on requirements for the particular embodiment.” *Bedros*, 2:47-50. Adjustments to the optics may be made “**based on commands determined by a control system** based on determined environmental changes within the assigned scene 100.” *Id.*, 2:54-60. Server 306 can be a computer 410 that “executes image capturing and processing software.” *Bedros*, 8:20-21, 8:54-55. Computer 410 includes input 416 that is **directly connected to a camera**. *Id.*, 9:14-17. Therefore, computer 410 (i.e., server 306) is a control system that controls the image capture of camera 104 and adjustments to camera optics. A POSITA would have recognized that camera 104 requires a control system and thus with a direct connection between camera 104 and server 306 would have considered them both to be a part of the same camera system.

251. To the extent the Board finds that Bedros does not disclose this claim,

it is my opinion that Bedros renders the claim obvious. A POSITA would have recognized that camera 104 requires a control system and would have been motivated to integrate server 306 into camera 104 because server 306 issues image-capturing commands to camera 104. It was also well-known to a POSITA at the relevant time to integrate camera controlling hardware and software with image processing components into surveillance cameras. For example, IP-based cameras contain built-in video analytics that recognizes significant features within a series of frames and transmits only important events. *AXIS 211M manual*, pp. 25-28; *see also Laganier* '363, ¶ 4. It was also well-known that transmitting high-resolution images lead to increased network traffic, and utilizing built-in video analytics in cameras provide the benefit of reducing network strain. A POSITA would have also been motivated to integrate server 306 into camera 104 because this implementation would have required nothing more than the application of known, conventional techniques. Furthermore, a POSITA would have had a reasonable expectation of success because smart IP cameras were well-known at the time.

252. By my review, claim 18 is substantially similar to claim 5, except that it recites a “method” instead of a “system.” Thus, for the reasons I discussed above for claim 5, it is my opinion that Bedros discloses or suggests claim 18.

8. Claims 7 and 17: “further comprising a database that receives metadata associated with the identified objects of interest from the video analytics system.”

253. By my review, claims 7 and 17 depend from claims 1 and 13, respectively, which as discussed above, Bedros discloses or suggests. In my opinion, Bedros further discloses or suggests claim 7’s additional limitation.

254. As I explained in §VIII.C.6, Bedros discloses identifying person 102 (“*identified objects of interest*”). Bedros explains that the facial recognition program (operating on server 308) receives facial image 116 and “respond[s] with an identity of an individual matching the face from the received image.” Bedros, 7:63-66. That information is stored in “image data structure 160 [that] may be a data structure held in...a database table record” as “IDENTITY.” *Bedros*, 3:58-4:3, 7:45-49. As I explained above in §VIII.C.2.i, data about the image stored in image data structure 160 is metadata.

255. By my review, claim 17 is substantially similar to claim 7, except that it recites a “method” instead of a “system.” Thus, for the reasons I explained above for claim 7, it is my opinion that Bedros discloses or suggests claim 17.

9. Claims 8 and 19: “further comprising a gateway system that provides access to the video analytics system over a public network.”

256. By my review, claims 8 and 19 depend from claims 1 and 13, respectively, which as discussed above, Bedros discloses or suggests. In my opinion,

Bedros further discloses or suggests claim 8's additional limitation.

257. As I explained in §VIII.C.2.d, Bedros's server 308 discloses or suggests "video analytics system." Server 306 is connected to "a computer network" through communication connection 420 of computer 410 and transmits processed image data from camera 302. *Bedros*, 9:9-17. In particular, Bedros explains that server 306 may include a "router...or other common network node." *Bedros*, 9:20-22. Furthermore, images from camera 302 are sent to server 306 before being sent to server 308. *Bedros*, 8:32-49. Indeed, Bedros further explains that computer 410 (e.g., server 306) can be directly connected to the camera and to a network via connection 420. *Bedros*, 9:14-17. Accordingly, a POSITA would have further recognized Bedros's server 306 as a "gateway system" because the '321 patent describes the "gateway" as a "node on the local network." '321 patent, 3:26-27. Routers are commonly recognized to be gateways or nodes in a network. *See Tanenbaum*, p. 405 ("gateway" is an older term for 'router'); EX1010, 731 (defining node as "(broadband local area networks) The point where signals leave one system and enter another"). Indeed, the user manual of AXIS 211M states that "[t]he broadband router has many different names: ... 'Internet Gateway' but the essential purpose of the device is the same." *AXIS 211M manual*, p. 33. Furthermore, "[t]he communication connection may include... **the Internet**, or other networks." *Id.*, 9:22-25. Internet is a type of "public network." *Id.*, 8:30-31. Server 306 ("gateway system") provides access to server 308 ("video

analytics system”) and both are connected to the Internet (“public network”).

258. By my review, claim 19 is substantially similar to claim 8, except that it recites a “method” instead of a “system.” Thus, for the reasons I explained above for claim 8, it is my opinion that Bedros discloses or suggests claim 19.

10. Claims 9 and 20: “wherein the object detection system is integrated within the gateway system.”

259. By my review, claims 9 and 20 depend from claims 8 and 19, respectively, which as I explained above, Bedros discloses or suggests. In my opinion, Bedros further discloses or suggests claim 9’s additional limitation. As I explained in §VIII.C.2.c) VIII.C.9, Bedros discloses or suggests server 306 (“object detection system” and “gateway system”).

260. By my review, claim 20 is substantially similar to claim 9, except that it recites a “method” instead of a “system.” For the reasons I explained above for claim 9, it is my opinion that Bedros discloses or suggests claim 20.

11. Claims 10 and 21: “wherein the one or more surveillance cameras each include a gateway system that provides access to the video analytics system over a public network.”

261. By my review, claims 10 and 21 depend from claims 1 and 13, respectively, which as I explained above, Bedros discloses or suggests. In my opinion, Bedros further discloses or suggests claims 10’s additional limitation.

262. I explained in §VIII.C.9 that server 306 (“gateway system”) provides access to servers 308 (“video analytics system”) and both are connected to the

Internet (“public network”). Furthermore, a POSITA would have recognized that camera 104/302 and server 306 would have considered them both to be a part of the same camera system, or alternatively, a POSITA would have been motivated to integrate server 306 into camera 104 as I explained in §VIII.C.7. Indeed, as I explained in §VI.C, network-connectable cameras were well known in the art by the time of the alleged invention, and it was well within the skill of a POSITA to incorporate object detection into the camera. Accordingly, a POSITA would have understood that camera 104 (“one or more surveillance cameras”) includes server 306 (“gateway system”) that is connected to servers 308 (“video analytics system”) through the Internet (“public network”).

263. By my review, claim 21 is substantially similar to claim 10, except that it recites a “method” instead of a “system.” For the reasons I explained above for claim 10, it is my opinion that Bedros discloses or suggests claim 21.

12. Claim 11: “wherein the video analytics system receives the image data from the object detection system after being transmitted over an enterprise network.”

264. By my review, claim 11 depends from claim 1, which as I explained above, Bedros discloses or suggests. In my opinion, Bedros further discloses or suggests claim 11’s additional limitation.

265. As I explained in §VIII.C.2.f, server 308 (“video analytics system”) receives facial image 116 (“image data”) from multiple servers 306 (“object

detection system”) over network 304 (“transmitted over the network”). Network 304 can be **“an internal network 304 of an entity** implementing the system 300.” *Bedros*, 8:24-27. A POSITA would have understood that the “internal network 304 of an entity” teaches “enterprise network” because it connects devices that are within an entity.

13. Claim 12: “wherein the video analytics system receives the image data from the object detection system after being transmitted over a cloud network.”

266. By my review, claim 12 depends from claim 1, which as I explained above, *Bedros* discloses or suggests. In my opinion, *Bedros* further discloses or suggests claim 12’s additional limitation.

267. The ’321 patent discloses that **“network cloud 60 is preferably a public/service provider network such as the Internet.”** ’321 patent, 5:13-14. I explained in §VIII.A.15 that a POSITA would have understood that “network cloud” also refers to “cloud network.” As I explained in §VIII.C.9, server 308 receives facial image 116 (“*image data*”) from processing units 402 of multiple servers 306 (“object detection system”) over the Internet (“public network”). *Bedros*, 8:27-31. A POSITA would have thus understood that the Internet includes a cloud network. Moreover, by the time of the alleged invention, accessing servers “via the cloud” was well known, especially for specialized software such as a facial recognition software. A POSITA would have understood that given that server 308 provides a

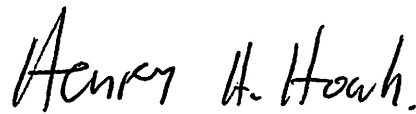
service (facial recognition), it could be implemented as a service via the cloud.

Accordingly, it is my opinion that Bedros discloses or suggests this limitation.

IX. CONCLUSION

268. This declaration and my opinions herein are made to the best of my knowledge and understanding, and based on the material available to me, at the time of signing this declaration. I declare that all statements made herein on my own knowledge are true and that all statements made on information and belief are believed to be true, and further, that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 or Title 18 of the United States Code.

Date: June 3, 2024



Henry Houh