

UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE PATENT TRIAL AND APPEAL BOARD

VUSIONGROUP SA,

Petitioner

v.

HANSHOW TECHNOLOGY CO., LTD.,

Patent Owner

Case No.: IPR2024-00963

Patent No. 10,701,321

**DECLARATION OF DR. MINGSHAO ZHANG IN SUPPORT OF PATENT
OWNER'S RESPONSE FOR *INTER PARTES* REVIEW OF U.S. PATENT
NO. 10,701,321**

I. INTRODUCTION

1. I, Mingshao Zhang, PhD., am currently an Associate Professor at Southern Illinois University Edwardsville. I have been retained on behalf of Hanshow Technology Co., Ltd. (“Hanshow”) to offer technical opinions relating to U.S. Patent No. 10,701,321 (Ex-1001, the “’321 Patent”) and prior art references relating to its subject matter.

2. I have reviewed the ’321 Patent and its prosecution history. I have also reviewed the following documents:

- a. The Petition for IPR2024-00963 filed by Petitioner (“Petition”)
 - b. The declaration of Petitioner’s expert, Dr. Henry Houh (Ex. 1003)
 - c. U.S. Patent App. Pub. No. 2012/0195363 (“Laganiere363”, Ex-1005)
 - d. U.S. Patent No. 8,139,817 (“Laganiere817”, Ex-1006)
 - e. U.S. Patent App. Pub. No. 2011/0113461 (“Laganiere461”, Ex-1007)
 - f. U.S. Patent No. 8,634,601 (“Bedros”, Ex-1008)
 - g. U.S. Patent App. Pub. No. 2014/0044348 (“Chen”, Ex-1009)
3. I have also reviewed various other references and documentation including:
- a. The transcript of the deposition of Dr. Henry Houh (Ex. 2003)
 - b. The Board’s decision instituting trial in this proceeding
 - c. Excerpt of THE AUTHORITATIVE DICTIONARY OF IEEE STANDARDS TERMS, 7th ed. (2000) (EX-1010)

- d. Excerpt of Merriam-Webster's Collegiate Dictionary, 11th ed. (2005)
(EX-1011)
- e. Excerpt of NIST Special Publication 800-53, rev. 4 (Jan. 22, 2015) (Ex-1012)
- f. Excerpt of REVERSE ACRONYMS, INITIALISMS AND ABBREVIATIONS DICTIONARY, 10th ed., vol. 3, part 1 (Oct. 23, 1985) ("Reverse Acronyms", Ex-1013)
- g. U.S. Patent No. 8,619,150 ("Deever '150", Ex-1014)
- h. U.S. Patent No. 8,432,965 ("Deever '965", Ex-1015)
- i. Excerpt of Dictionary of Contemporary Photography (1974, Ex-1016)
- j. Excerpt of AXIS 211M Network Camera User's Manual (Feb. 2010)
(“AXIS 211M manual”, Ex-1017)
- k. U.S. Patent App. Pub. No. 2016/0050396 (“Gali”, Ex-1021)
- l. U.S. Patent App. Pub. No. 2016/0191779 (“Yang”, Ex-1022)
- m. U.S. Patent Application Publication No. 2013/0166711 (“Wang”, EX-1023)
- n. Christopher John Lindblad, A Programming System for the Dynamic Manipulation of Temporally Sensitive Data, (1994) (Ph.D. dissertation, Massachusetts Institute of Technology) (“Lindblad”, EX-1024)

- o. David James Wetherall, An Interactive Programming System for Media Computation (1994) (Master of Science dissertation, Massachusetts Institute of Technology) (“Wetherall”, EX-1025)
 - p. William F. Stasior, An Interactive Approach to the Identification and Extraction of Visual Events (1997) (Ph.D. dissertation, Massachusetts Institute of Technology) (“Stasior”, EX-1026)
 - q. Jia Deng et al., ImageNet: A Large-Scale Hierarchical Image Database, 2009 IEEE Conference on Computer Vision and Pattern Recognition, (2009) (“Deng”, EX-1027)
 - r. 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”, EX-1028)
 - s. Andrew Tanenbaum, Computer Networks (3rd ed. 1996) (“Tanenbaum”, EX-1029)
 - t. 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”, EX-1030)
4. Counsel has informed me that I should consider these materials through the lens of one of ordinary skill in the art related to the ’321 Patent as of February 27, 2015, the priority date of the ’321 Patent, and I have done so.

5. I have no financial interest in the parties or in the outcome of this proceeding. I am being compensated for my work as an expert on an hourly basis. My compensation is not dependent on the outcome of these proceedings or the content of my opinions.

6. In writing this declaration, I have considered the following: my own knowledge and experience, including my work experience in the fields of Image Processing, Computer Vision; and Robotics. my experience in teaching those subjects; and my experience in working with others involved in those fields. In addition, I have analyzed various publications and materials that I cite in my declaration. My opinions, as explained below, are based on my education, experience, and expertise in the fields relating to the '321 Patent. Unless otherwise stated, my testimony below refers to the knowledge of one of ordinary skill in those fields as or before the '321 Patent's priority date of February 27, 2015. Any figures that appear within this document have been prepared with the assistance of Counsel and reflect my understanding of the '321 Patent and the prior art discussed below.

7. This declaration explains the conclusions that I have formed based on my analysis. To summarize those conclusions, based upon my knowledge and experience and my review of the relevant documents, I believe that the challenged claims of the '321 Patent are not rendered obvious by any combination of prior art asserted in the Petition or addressed in Dr. Houh's declaration.

II. EXPERIENCE

8. I hold a Ph.D. in Mechanical Engineering from Stevens Institute of Technology, awarded in 2016. My dissertation, titled "Vision-based Object Tracking using 3-D Scanners with Applications in Educational Laboratories," underscores my expertise in advanced imaging and tracking technologies.

9. I am currently an Associate Professor at Southern Illinois University Edwardsville, a role I have held since August 2022. Prior to this, I served as an Assistant Professor at the same institution from August 2016 to August 2022.

10. I have extensive experience teaching a variety of courses related to Mechatronics and Robotics, with a particular focus on image processing and computer vision. My teaching portfolio includes courses such as Robotic Vision, Design in Mechanical and Robotics, Sensors and Actuators.

11. Many of my academic publications are focused on vision-based control systems, 3D reconstruction, and the integration of virtual and real environments. Notable publications include the development of a vision-based control system for mobile 3D printers and immersive educational systems that combine real and virtual environments. These publications demonstrate my knowledge and innovative approach to image processing and computer vision.

12. Additionally, I co-invented a multifunctional camera system for video-assisted thoracic surgery, which was approved in December 2022.

III. LEGAL STANDARDS

13. I have been informed by Counsel and understand that a claim is unpatentable for obviousness under 35 U.S.C. § 103 “if the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to which the claimed invention pertains.”

14. I am informed by Counsel and understand that an obviousness inquiry requires consideration of the following factors: (1) the scope and content of the prior art; (2) the differences between the claims and the prior art; (3) the level of ordinary skill in the pertinent art; and (4) any objective indicia of non-obviousness cited by the patent owner, such as commercial success, long-felt but unresolved need, failure of others, industry recognition, copying, and unexpected results. I understand that the foregoing factors are sometimes referred to as the “Graham factors.”

15. I have been informed by Counsel and understand that documents and materials that qualify as prior art can render a patent claim unpatentable as obvious. I am informed by Counsel and understand that all prior art references are to be looked at from the viewpoint of a person of ordinary skill in the art at the time of the invention, and that this viewpoint prevents one from using his or her own insight or hindsight in deciding whether a claim is obvious. I understand that the words of a claim are generally given their ordinary and customary meaning as understood by a

POSITA in view of the entire patent specification in which they are present and the prosecution history of the patent.

16. I understand, nevertheless, that a claim's words may be given a meaning different than their plain and ordinary meaning if: (1) a patentee explicitly and unambiguously defined a term in a different manner as compared to its plain meaning; or (2) a patentee "disclaims" the full scope of a claim term's plain meaning, such as, by clearly and unambiguously stating during prosecution that a claim term does or does not encompass certain subject matter that may otherwise be encompassed by the plain words of the claim.

17. I am informed by Counsel and understand that when a patented invention is a combination of known elements, the Patent Trial and Appeal Board (PTAB) must determine whether there was an apparent reason to combine the known elements in the fashion claimed by the patent at issue by considering the teachings of prior art references, the effects of demands known to people working in the field or present in the marketplace, and the background knowledge possessed by a person having ordinary skill in the art.

18. I am informed by Counsel and understand that a patent claim composed of several limitations is not proved obvious merely by demonstrating that each of its limitations was independently known in the prior art. I am informed by Counsel and understand that identifying a reason those elements would be combined can be

important because inventions in many instances rely upon building blocks long since uncovered and claimed discoveries almost of necessity will be combinations of what, in some sense, is already known. I am informed by Counsel and understand that it is improper to use hindsight in an obviousness analysis, and that a patent's claims should not be used as a "roadmap."

19. I have been informed by Counsel and understand that in *inter partes* review proceedings, the Petitioner bears the burden of proof to prove that the claims that were challenged in the Petition and for which the Board instituted review are unpatentable. It is my understanding that the Petitioner also bears the burden of proving a proposition of unpatentability by a preponderance of the evidence. I have been informed that the preponderance of the evidence standard requires that a reasonable factfinder would find a material fact more probable than the nonexistence of that fact. I understand that this standard does not permit speculation concerning particular facts and is rather focused on whether the evidence more likely than not demonstrates the existence or non-existence of particular material facts. I have been informed that "preponderance of the evidence" is a lower standard than "clear and convincing evidence" (which requires a fact to be substantially more likely to be true than untrue), or "beyond a reasonable doubt" (which is an exceedingly high standard that I understand is usually used in criminal matters). Throughout my Declaration, I have applied the "preponderance of the evidence" standard. I have also been

informed that, if the prior art does not demonstrate that an independent claim is unpatentable, it cannot be used to demonstrate that any corresponding dependent claim is also unpatentable.

IV. PERSON OF ORDINARY SKILL IN THE ART

20. I understand that a person of ordinary skill in the art (“POSITA”) is a hypothetical person skilled in the art, not a judge, a layperson, or a genius in the art at hand.

21. I am informed and understand that the following factors may be considered in determining the level of ordinary skill in the art: (1) educational level of the inventor; (2) type of problems encountered in the art; (3) prior art solutions to those problems; (4) rate at which innovations were made in the field; (5) sophistication of the technology; and (6) educational background and experience of those who were actively working in the field during the relevant time period (priority date).

22. In ¶ 19 of his declaration, Dr. Houh opines that a POSITA for the ’321 Patent “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.” He then adds, “[l]ack of work experience can be remedied by additional education, and vice versa.” (Ex. 1003, ¶ 29). I understand this is the same definition of a POSITA applied by the Board in its decision to

institute this proceeding. For the purposes of my analysis in the context of this IPR proceeding, I have likewise applied this definition for a POSITA.

23. For purposes of this declaration, I have been asked to assume the priority date of the '321 Patent is February 27, 2015. When I refer to a POSITA, I am referring to one of ordinary skill in the art as of that date, unless otherwise specified.

24. As of that priority date, I met, and in fact exceeded, the qualifications of a POSITA. Accordingly, I am qualified to opine as to the knowledge and understanding of a POSITA as of the priority date of the '321 Patent, and my analysis, conclusions, and opinions presented herein are considered from such a perspective.

V. OVERVIEW OF THE '321 PATENT

25. Based on my review and analysis of the '321 Patent, I understand it discloses innovative solutions to address significant technical challenges in video surveillance systems.

26. Video surveillance systems typically include several common components: (i) surveillance cameras that capture image data, which is then transmitted over data networks; (ii) network video recorders (NVRs) that store this image data; (iii) video analytics systems that analyze the image data to detect and determine temporal and spatial events; and (iv) control systems that coordinate the operation of these components. *Id.*, 1:6-7, 16-26. In my experience and as acknowledged in the '321

Patent's specification, the methods employed by video analytics systems for detecting and analyzing classes of objects in image data have become increasingly accurate and important, enabling more sophisticated functions such as object detection, which might determine the presence or type of object, and object analysis, which might identify an object. *Id.*, 1:27-36.

27. Prior art video surveillance systems typically operated by having surveillance cameras send their image data over networks to video analytics systems for further processing. *Id.*, 1:54-59. However, technological advances have resulted in cameras able to capture images with significantly higher resolutions, which has correspondingly increased the volume of data being transmitted. *Id.*, 1:59-63.

28. This increased data traffic volume creates several technical problems. First, it limits the available bandwidth of the networks, causing performance problems for both the video surveillance system components and other nodes communicating over the same networks. *Id.*, 1:46-53. Second, because video analysis methods are typically memory- and processor-intensive tasks, the video analytics systems need to be high-performance or specialized computer systems, significantly increasing costs. *Id.*, 1:65-2:2. Third, as analysis methods are constantly improving, these systems require periodic updates to make use of the latest technology, further adding to the system's overall cost. *Id.*, 2:5-7.

29. In my opinion, the '321 Patent innovatively addresses these problems by disclosing a video surveillance system that separates the object detection and object analysis functions into distinct systems. This separation represents a significant technical improvement over prior approaches and is implemented as follows: the object detection system detects objects of interest and classes of objects within the image data, *Id.*, 2:11-12, while the object analysis system analyzes the image data containing the detected objects, to, for example, recognize or identify those objects of interest, *Id.*, 2:12-15.

30. This architectural division provides several technical benefits. For example, it reduces the volume of image data transmitted over networks by positioning the object detection system closer to the cameras and sending only the "interesting" portions of the image data to the more remote object analysis system. *Id.*, 2:15-20.

31. One innovative aspect of the '321 Patent is how the object detection system ranks the objects of interest and saves these rankings in the metadata of the image data. The object analysis system can then read these rankings from the metadata and make its own decision about whether to analyze the image data. This creates a flexible, adaptive system where processing decisions can be made dynamically based on the metadata, rather than following fixed approaches such as those found in the prior art. This flexibility represents a significant technical advancement over prior systems that used such fixed processing approaches.

32. This adaptive capability enables more efficient resource utilization while maintaining effective surveillance capabilities. Specifically, the technical benefits of this architecture include reduced memory and processor requirements for both the object detection system and the video analytics system, thus substantially reducing costs while maintaining or even improving system performance.

33. Figure 5 of the '321 Patent illustrates an embodiment of this innovative video surveillance system, showing the separation of the object detection system and object analytics system.

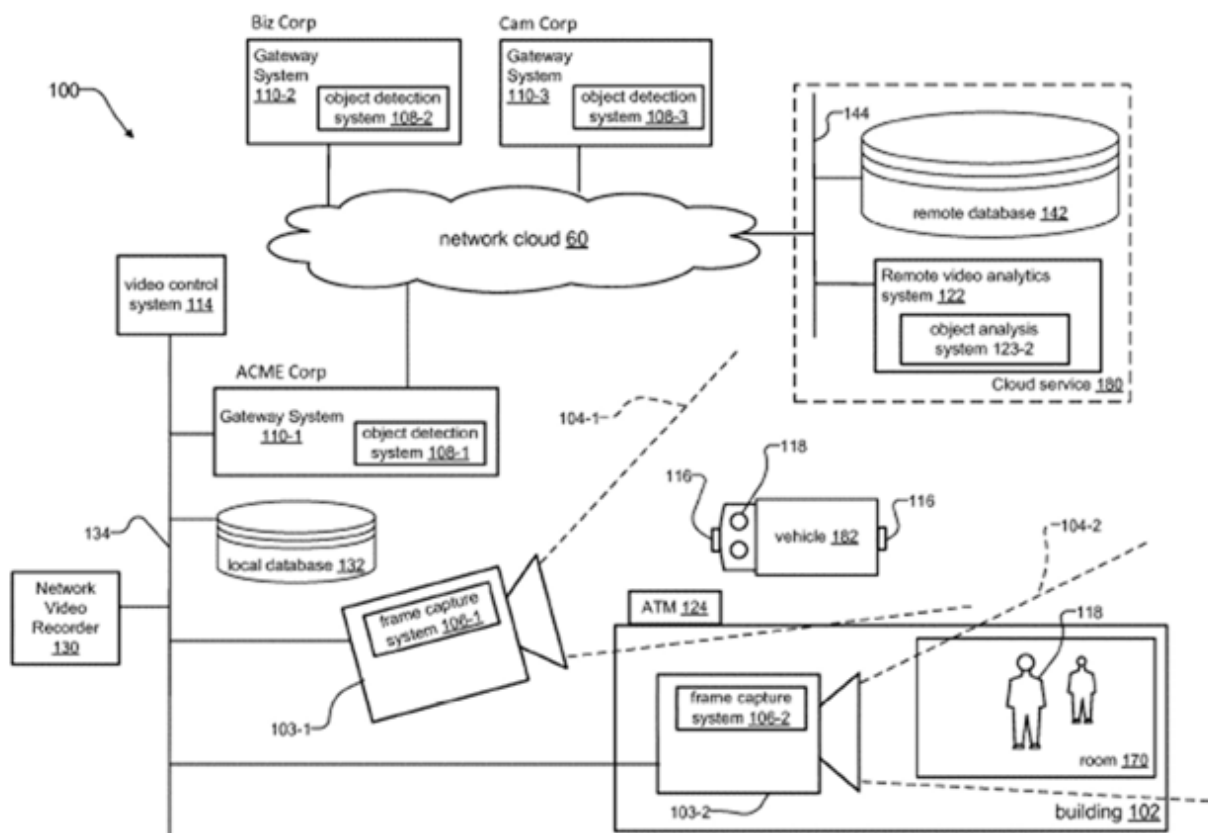


FIG. 5

Ex. 1001, FIG. 5.

34. Unlike prior art systems, the '321 Patent discloses a distributed video surveillance system that separates the object detection system from the object analytics system, achieving significant network bandwidth savings. The system's design allows the object analytics system to selectively process image data based on rankings generated by the object detection system and saved in the metadata of the image data, creating an adaptive and flexible approach to video surveillance that was not present in prior systems.

VI. OVERVIEW OF THE ASSERTED PRIOR ART REFERENCES

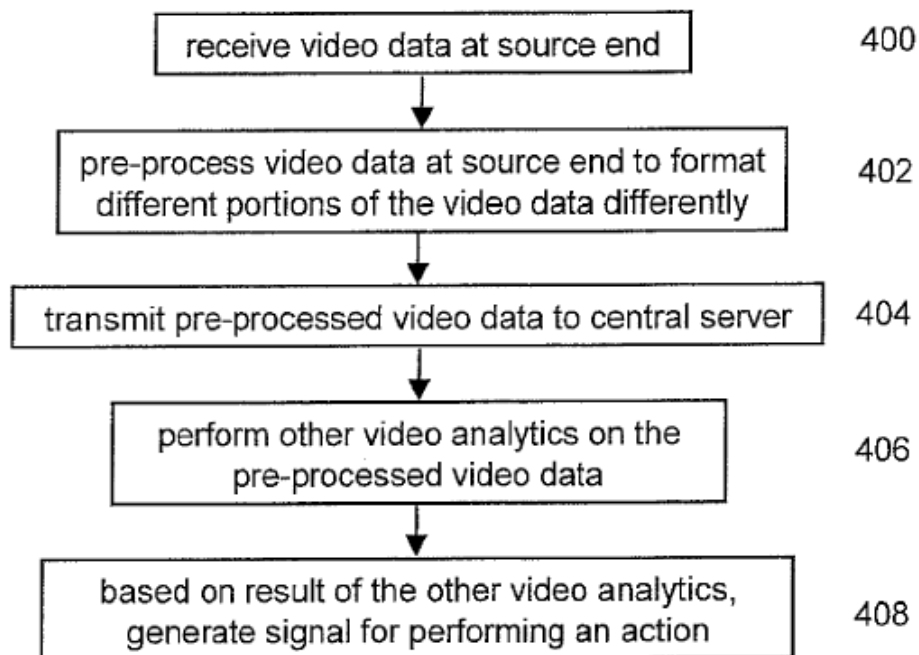
35. A summary of the asserted prior art references relied upon in the Houh Declaration is included below.

A. Laganiere363 (Ex-1005)

36. In my analysis of the prior art, I have examined Laganiere363 (U.S. Patent App. Pub. No. 2012/0195363). Laganiere363 discloses a method for performing video analytics where video data relating to an event of interest received at a source end is pre-processed at the source end using video analytics to reduce 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. Ex-1005, [Abstract]. After this initial pre-processing, the pre-processed video data is transmitted to a central server via the WAN where other video analytics processing is performed. *Id.* Based on a result of the other video analytics processing, a signal

is generated for performing a predetermined action in response to an occurrence of the event of interest at the source end. *Id.* I note that Laganiere363 also teaches that the bandwidth may be optionally reduced by transmitting only one or a few frames having a suitable quality relative to the other frames. *Id.*, [0041].

37. There are several technical aspects of Laganiere363 that are relevant to this matter which include the following. First, examining Figure 4 of Laganiere363, reproduced below, I note the figure specifically says, “based on result of the other video analytics, generate signal for performing an action.” This indicates that all the actions in their later video analytic system are predetermined in nature. This is a significant technical distinction from the approach of the ’321 Patent.



Ex-1005, FIG. 4.

38. Furthermore, after reviewing claims 5-11 of Laganier³⁶³, particularly claims 8 and 9, it is apparent that their main objectives are focused on preserving better copies of “events of interest,” which fundamentally differs from the approach of the ’321 Patent. According to the ’321 Patent’s approach, the system ranks all objects of interest, keeps all copies, and sends them to the server. This represents a technically distinct methodology for handling video data and bandwidth management compared to the approach taught in Laganier³⁶³.

B. Laganier⁸¹⁷ (Ex-1006)

39. I have likewise examined Laganier⁸¹⁷ (U.S. Patent No. 8,139,817).

40. Laganier⁸¹⁷ teaches a system and method for constructing face image logs. Ex-1006, [Abstract]. I note that Laganier⁸¹⁷ addresses a specific problem in facial recognition technology: face logs capture a high rate of face images, which can overwhelm any human operator or automated biometric face recognition systems attempting to process the face image log in real-time. *Id.*, 1:32-37. Additionally, large face logs result in wasteful processing. *Id.*, 1:37-39. Laganier⁸¹⁷ identifies a need for quality appraisal of face images for the purpose of selecting high quality faces from video sequences. *Id.*, 1:53-55.

41. As depicted in Figure 1 of Laganier⁸¹⁷, the system comprises three main components: (1) a face detection module for determining face images of one or more subjects from candidate regions within each frame of the video sequence; (2) a

quality analysis module for receiving face images from the face detection module and determining a quality score based a plurality of criteria of the face images; and (3) a storage means for storing the analyzed face images in a face log with an associated quality score, wherein the face log is periodically purged based upon the score of the acquired images. *Id.*, 2:29-39.

42. This purging process is reflected in at least Figure 7 (reproduced below), Step 722, which states “[p]urge face history based upon quality.” In my opinion, this indicates that the motivation behind Laganier817’s system is to acquire high quality face images for each person. Based on my expertise in video processing systems, I conclude that Laganier817’s system is designed to store only one (or very few) facial images per person. In my opinion, this design is likely intended for security personnel to review these images manually (by person), rather than for processing by a video analytic system.

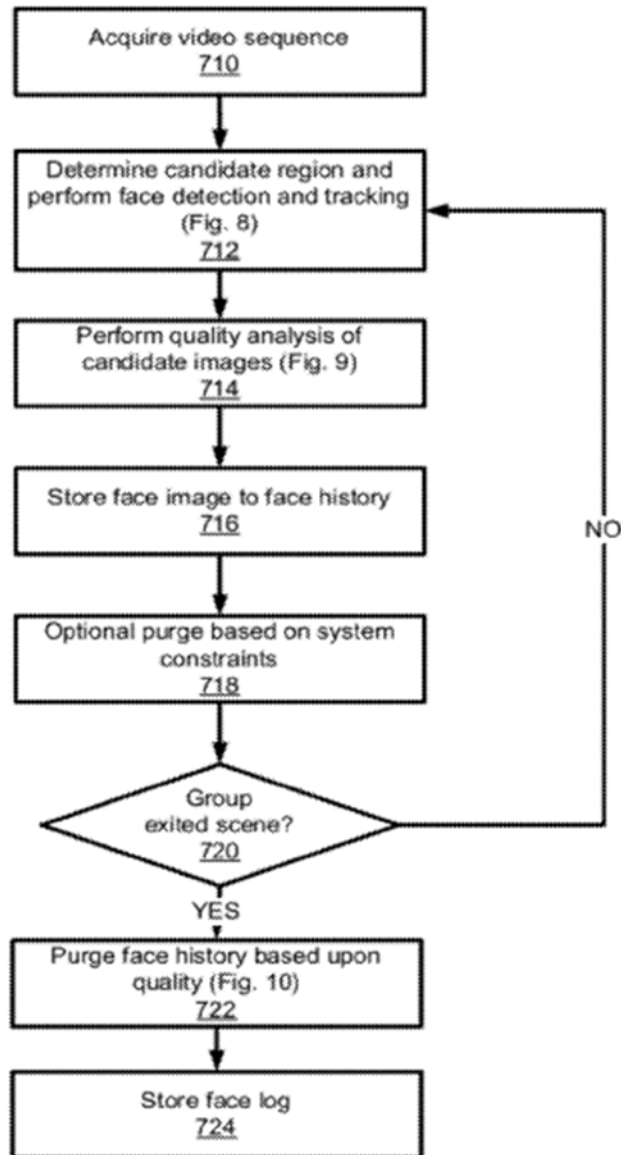


Fig. 7

Ex-1006, FIG. 7.

C. Laganriere 461 (Ex-1007)

43. Laganriere461 (U.S. Patent App. Pub. No. 2011/0113461) teaches a method and system for reducing the amount of space required to store video data using

selective deletion of portions of video data stored in a video storage device. Ex-1007, [0002]. This selective deletion process is performed under the control of a central server based on the results of video analytics processing of the video data. *Id.*

44. The implementation of this method is illustrated in a simplified block diagram in FIG. 1 of Laganieri461.

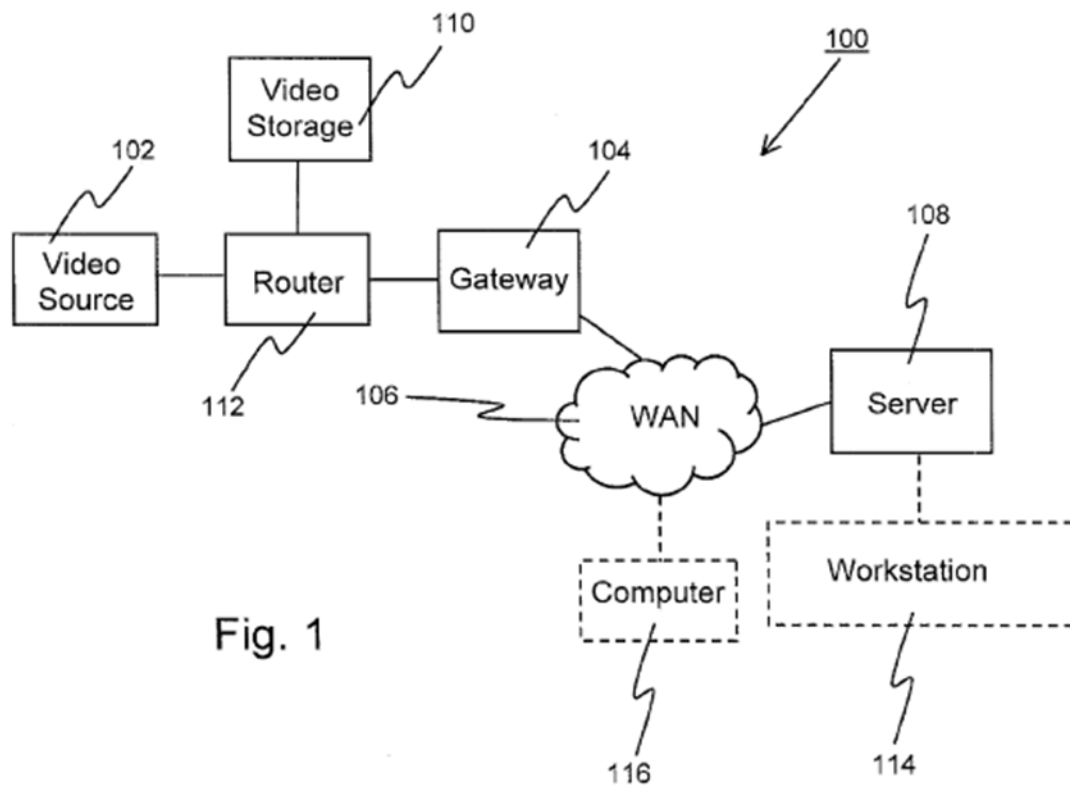


Fig. 1

Ex-1007, FIG. 1.

45. A simplified flow diagram of the system's operation is presented in FIG. 4.

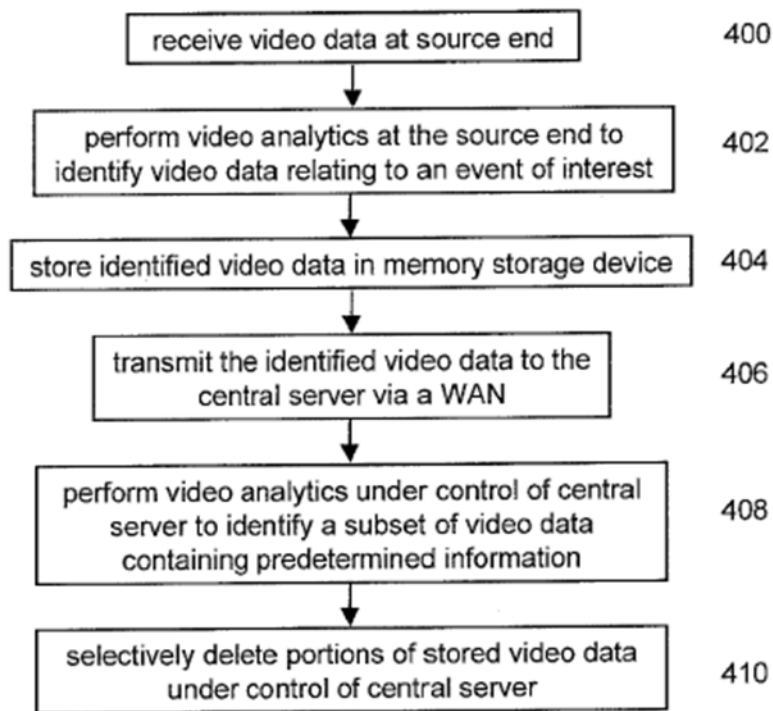


Fig. 4

Ex-1007, FIG. 4.

D. Bedros (Ex-1008)

46. I have also reviewed Bedros (U.S. Patent No. 8,634,601). Bedros teaches systems, methods, and software that utilize a single high-resolution camera to capture images, locate individuals, and provide images to a facial recognition process. Ex-1008, 1:58-62. In examining the Bedros system architecture, I observe that it includes a particular server 306 that receives digital images or a stream of digital video frames from a camera 302 over the network 304, extracts facial images

116, and optionally submits each facial image 116 to one or more servers 308 that execute facial recognition software. *Id.*, 8:20-49.

47. I note that some embodiments of Bedros's system also include prioritization and filtering that choose which captured images from a stream of images to process and when to process them when there are many images to be processed. *Id.*, 1:62-65. A method of filtering and prioritizing facial images for submission to the facial recognition system executed by the one particular server 306 is shown in its FIG. 2. *Id.*, 4:26-29, 8:20-22.

48. Bedros implements a rigid, predetermined sequence of processing steps. As the annotation in the reproduction of FIG. 2 below illustrates, step 222, which is part of an optional processing layer 220, discards an image when the quality measure of the image is less than a threshold. This step occurs several steps before step 242, which submits images from a queue to facial recognition software (and therefore several steps before the facial recognition software receives images).

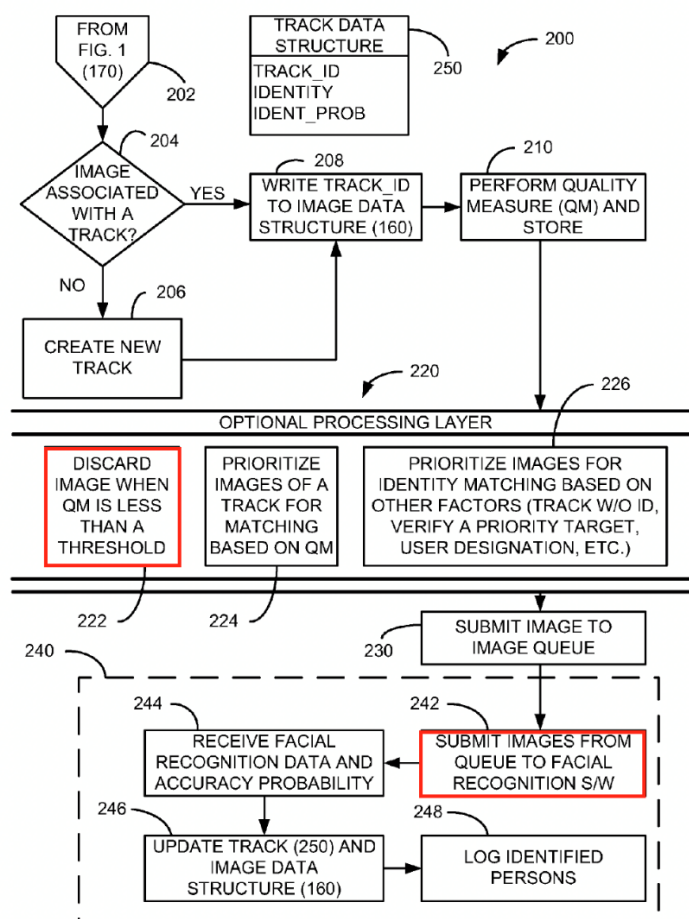


FIG. 2

Ex-1008, FIG. 2 (annotated)

49. In my opinion, this represents a significant distinction from the '321 patent, which discloses a more flexible solution. The Bedros system follows a fixed sequence of predetermined steps as illustrated by Figure 2, with specific decision points that cannot be arbitrarily altered during operation. This rigid approach contrasts with the adaptive methodology presented in the '321 patent, which allows for greater responsiveness to varying conditions and requirements. Based on my expertise in video processing systems, this predetermined pathway in Bedros would

significantly limit its ability to adapt to different operational scenarios compared to the solution disclosed in the '321 patent.

E. Chen (Ex-1009)

50. I have also reviewed Chen (U.S. Patent App. Pub. No. 2014/0044348). Chen teaches a method, software, and computer system for assessing the quality of an image, which might include evaluating the quality of facial images to determine their suitability for use in facial recognition applications. Ex-1009, [0002].

VII. CLAIM CONSTRUCTION

51. I have been informed by Counsel that claim terms are to be given their ordinary and customary meaning as understood by a person of skill in the art at the time of the invention. I have been further informed that, if the meaning of a claim term is disputed, intrinsic evidence, such as the claim language itself, the invention's written description, and the prosecution history of the patent, is the primary means by which the meaning of the term is to be understood.

52. I have reviewed the claims at issue in this proceeding. In my opinion, no claim construction is necessary for the claims at issue, and all claim terms can be understood according to their plain and ordinary meaning as would be understood by a person of ordinary skill in the art.

53. I am aware that Petitioner's expert, Dr. Houh, has proposed construing certain claims as means-plus-function claims. I am informed by Counsel that the standard

for whether a claim should be construed as a means-plus-function claim is whether the words of the claim are understood by a person of ordinary skill in the art to have a sufficiently definite meaning as the name for structure. I have been further informed that, there is a rebuttable presumption that §112(f) does not apply and Petitioner bears the burden of going forward with evidence to rebut this presumption by a preponderance of the evidence. I note that Petitioner provides no evidence to rebut this presumption and presents solely a hypothetical argument “if the Board finds the presumption is overcome for any of the ‘system’ limitations.” Paper 2 (“Petition”), 7; Ex-1003, ¶52. While I understand there may be legal disagreement regarding whether such construction is even appropriate given Petitioner’s initial position that no claim construction was necessary, I can state that, even if the Board were to adopt Dr. Houh’s proposed means-plus-function constructions, my technical conclusions regarding non-obviousness and other substantive issues in this proceeding would remain unchanged. The technical merit of the claimed inventions stands regardless of which claim construction approach is applied.

**VIII. PETITIONER AND ITS EXPERT HAVE NOT ESTABLISHED BY A
PREPONDERANCE OF THE EVIDENCE ANY CLAIM IS
UNPATENTABLE**

A. Ground 1

- a. Laganriere363, Laganriere 817, and/or Laganriere461 do not disclose
or render obvious “ranked by the object detection system” ([1.5.1],**

[13.5.1]) 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” ([1.5.3], [13.5.3])

54. I have reviewed the prior art references Laganriere363, Laganriere817, and Laganriere461. In my opinion, these references, individually or in combination, fail to disclose or render obvious “ranked by the object detection system” ([1.5.1], [13.5.1]) 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” ([1.5.3], [13.5.3]) limitations.

55. In my opinion, independent claims 1 and 13 require two distinct technical components: an “object detection system” and an “object analysis system.” The object detection system performs the technical function of ranking detected image data corresponding to objects of interest according to different factors and then saving this ranking in metadata along with the image data and the object detection results. Separately, the object analysis system performs the distinct technical function of reading these rankings within the metadata and determining whether or not to analyze the image data.

56. These two systems represent separate and distinct elements of the claimed invention. The technical architecture described in the plain language of the claims illustrates this separation, as it specifically requires the object detection system and

the video analytics system (which contains the object analysis system) to be separated by a network and to transmit data over this network. This separation between the two systems represents an important technical distinction that would be recognized by a person of ordinary skill in the art. This separation between the two systems can: (1) save the bandwidth of the intervening networks by sending only the portions of the image data that includes the detected objects of interest; and (2) reduce the memory and processor requirements for both the object detection system and the video analytics systems, thus reducing costs and allowing for more efficient processing while maintaining accurate object recognition and analysis.

57. I understand that Petitioner concedes that Laganriere³⁶³ does not disclose either of these limitations. Petition, 23. In my opinion, Petitioner's attempts to use Laganriere⁸¹⁷ and Laganriere⁴⁶¹ to cure these deficiencies fail for several reasons.

58. I understand that Petitioner maps the “object detection system” of independent claims to a face image quality scoring in Laganriere⁸¹⁷'s face log creation and the “object analysis system” of independent claims to purging lower quality images, again in Laganriere⁸¹⁷'s face log creation. Petition, 23-25, 28. However, in my opinion, even Petitioner's own evidence and arguments demonstrate that the alleged “quality scoring” and “purging lower quality images” are not implemented by separate and distinct elements. Therefore, they cannot teach or suggest the separate

“object detection system” and “object analysis system” required by independent claims.

59. In my analysis of FIG. 1 of Laganiere817, I observe that the system for face tracking and face log creation includes one video surveillance system 100 associated with video camera 116. The video surveillance system 100 comprises “a central processing unit for executing software (CPU) for performing video analysis.” Ex-1006, 3:55-66, 4:3-4. By executing code stored in the memory 114, the face detection and tracking module 130 analyzes the video frame and provides face images to the quality analysis module. *Id.*, 4:21-26. “The quality analysis module 140 analyzes the face image to assign a quality score to the image and to determine if it is of suitable quality for storage in the face log.” *Id.*, 4:28-30. Thus, in my opinion, the same quality analysis module creates a quality score for a face image and determines whether to store it in the face log.

60. I understand Petitioner further argues that “Laganiere817 discloses purging lower quality images from the face log once a face (or license plate) has been recognized...[b]y purging lower quality images from the face log, the central server would also decide 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.” Petition, 28. However, in my review of Laganiere817, I find no disclosure of a central server separate from its video surveillance system, let alone a central server

purging lower quality images from the face log based on their quality rankings (created by the video surveillance system). Moreover, the cited materials fail to teach or suggest that the alleged “quality scoring” and “purging lower quality images” are implemented by separate entities or separate hardware. In fact, Laganiere⁸¹⁷ explicitly discloses that “the purging of the face log may be performed by the quality analysis module 140,” that is, the same quality analysis module within the video surveillance system creates a quality score for a face image and performs the purging of lower quality images based on the quality score. *Id.*, 9:42-45.

61. In addition, in my opinion, figure 7 of Laganiere⁸¹⁷ describes a method that is intended for execution in a single locality. This method starts from “Acquire video sequence 710” and ends with “Store face log 724”, and the main goal is to pick up the highest quality face from analyzing *one* video. *Id.*, 7:55-8:54, FIG. 7. In my opinion, such an essentially lineal process is most suitable for execution in a single locality, and a POSITA would not look to implement this method in two or more distributed servers separated by a network.

62. Modifying Laganiere⁸¹⁷ in the way proposed by Petitioner – saving lower quality images together with quality scores in the face log for future processing – would require the face recognition system outside of Laganiere⁸¹⁷’s system to receive quality scores in addition to images below a certain quality threshold, only for it to read the quality scores and discard the lower quality images. This directly

contradicts Laganier817's explicit teaching that "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.

63. This proposed modification also contradicts the fundamental purpose of Laganier817's face log construction, which is to reduce capacity requirements and operator intervention or judgment and avoid overwhelming face recognition systems. *See id.*, 2:26-63, 3:4-6. A person of ordinary skill in the art would not reconstruct Laganier817's face log creation system in a way that would overwhelm face recognition systems (which are analogous to the video analytics system containing the object analysis system) with quality scores in addition to excessive numbers of lower quality images.

64. Furthermore, using lower quality images would not only overwhelm the system, but it would yield unreliable face recognition results. This is precisely why Laganier817 only keeps high quality images. From my review of Laganier817, a person of ordinary skill in the art would not find any need to use a face recognition system outside of Laganier817's system just to "save lower quality images together with quality scores in the face log for future processing" – in fact, Laganier817 teaches quite the opposite approach.

65. Accordingly, Laganier363 and Laganier817, individually or in combination, do not teach a "central server" that purges images from the face log. I

understand the Petitioner has additionally cited Laganiere461. Petition, 28. But Laganiere461 has nothing to do with quality rankings or face logs. Instead, it describes a system similar to Laganiere363. *Id.*, 29.

66. Laganiere461, like Laganiere363, does not disclose the central server purging images “from the face log based on their quality rankings,” nor either of the limitations [1.5.1]/[13.5.1] and [1.5.3]/[13.5.3], alone or in combination with the other references Petitioner has cited.

67. My review of Laganiere461 reveals that the selective deletion of portions of video data is “based on a result of video analytics processing of the video data.” Ex-1007, [0010]. As shown in FIG. 4, for example, the central server has to perform video analytics on all the transmitted video data before selectively deleting portions of stored video data. *Id.*, [0025], FIG. 4. Laganiere461 further discloses the central server performing video analytics “to identify a single frame of data in a sequence of frames that provides an image of the face of an individual that is of sufficient quality to enable identification of the individual.” *Id.*, [0028]. In my opinion, this is the opposite of the central server “read[ing] the rankings within the metadata” created by a separate object detection system before “decid[ing] whether or not to analyze the image data.”

68. Only Laganiere817 remotely discloses “quality scoring” and “purging lower quality images.” However, in my opinion, there is no clear explanation of how the

alleged procedures are implemented by separate and distinct elements. Laganiere817 does not disclose two separate and distinct “object detection system” and “object analysis system” that respectively implement the limitations [1.5.1]/[13.5.1] and [1.5.3]/[13.5.3] as independent claims require.

69. A person of ordinary skill in the art would not view those alleged steps for face log creation to be separate in Laganiere817’s system. The claims of Laganiere817 refer to “generating... face logs from a video sequence.” In my opinion, since Laganiere817’s video surveillance system includes only one video camera monitoring one scene or location, a POSITA would understand that the alleged steps of Laganiere817 are implemented by a single system. (Ex-1006, 3:56-58). Further, in my opinion, since Laganiere817 aims for creating logs that “contain only *the best* images available for each individual observed” (Ex-1006, [abstract]) (emphasis added), a single-locality system with localized access to all the images, instead of a distributed system across a network, would achieve the purpose. Similarly, I note that Laganiere817 is directed to a method and system for “generating concise and complete face logs from a video sequence.” Ex-1006, 2:20-22, 2:29-31. I therefore conclude that a POSITA would understand that Laganiere817 contemplates an integral end-to-end processing from a video sequence to a video log, not two separate and distinct systems.

70. Because the combination of Laganier references does not disclose two separate and distinct “object detection system” and “object analysis system” that respectively implement the limitations [1.5.1]/[13.5.1] and [1.5.3]/[13.5.3], it is my opinion that Petitioner fails to meet its burden of showing that independent claims 1 and 13, and all dependent claims, are unpatentable by a preponderance of the evidence under Grounds 1A and 1B.

b. There is no motivation to combine Laganier363, Laganier817, and/or Laganier461 in the way proposed by Petitioner

71. The proposed combination of Laganier references likewise fails to render the claimed invention obvious. In my opinion, Petitioner has not articulated a consistent technical reason why a person of ordinary skill in the art would engage in the convoluted disassembly-and-reassembly of disclosures as proposed.

72. First, I understand that Petitioner highlights that Laganier363 “reduces bandwidth by [the first video analytics] selecting particular frames ‘with a suitable license plate image’ and transmitting only the ‘one or few frames having a suitable quality relative to the other frames.’” Petition, 23, citing Ex-1005, [0041]. Dr. Houh also confirmed that this process is performed by the first video analytics on the video source side, before transmission of data to the central server. Ex-2001, 66:22-71:10.

73. Given this understanding, if a POSITA were to follow Petitioner’s rationale that “POSITAs would have understood the face log in Laganier817 was a suitable

method of accomplishing Laganiere363's goals of reducing bandwidth by transmitting only high-quality object images for further processing," Petition, 25, I would expect both Laganiere817's quality scoring and purging of lower quality images to be implemented in Laganiere363's first video analytics component before transmission to the central server, since this will avoid transmitting lower-quality images (and their associated metadata) through the network and thus reduce bandwidth.

74. In my analysis, Laganiere363 discloses a first video analytics component for detecting events of interest in video data and selecting/transmitting frames of suitable quality. Ex-1005, [0040]. This is analogous to Laganiere817's face detection, tracking, quality scoring, and purging of lower quality images during face log creation. Ex-1006, 4:20-30. Separately, Laganiere363 discloses a central server for performing secondary video analytics, which is analogous to subsequent facial recognition operations that are outside Laganiere817's described system. Ex-1005, [0040]; Ex-1006, 3:4-7.

75. Even if Laganiere817's face detection/tracking and quality scoring/purging functionalities were somehow implemented in Laganiere363's first video analytics component, this combination still fails to disclose two separate and distinct systems as in the claimed invention, that is, an "object detection system" that ranks the detected image data corresponding to objects of interest according to different

factors and saves the ranking in metadata with the image data, and separately, an “object analysis system” that reads the rankings within the metadata and decides whether or not to analyze the image data.

76. There would be no reasonable expectation of success in implementing only Laganier817’s quality scoring in the first video analytics while implementing the purging of lower quality images in the central server. This approach artificially divides what is an organic whole in Laganier817’s disclosure, unnecessarily complicating the implementation.

77. Such an implementation would not achieve the stated goal of bandwidth reduction. Petition, 25. Transmitting lower quality images through the network only to discard them later would defeat the purpose of reducing bandwidth. Even though metadata is typically much smaller than image data in terms of file size, implementing the mechanism as the petitioner suggested would at least require the system to transmit all the image data along with additional metadata associate with them all, increasing the bandwidth, not reducing it. Furthermore, in Laganier817, performing face recognition using lower quality images would yield unreliable results. Based on my experience, a POSITA would find no motivation to keep the lower quality images based on the teachings of Laganier817, much less to transmit them through the network.

78. Furthermore, I disagree with the assertion that “POSITAs would have found it obvious to combine these references’ teachings for the central server to start with the highest quality images when performing facial recognition.” *See* Petition, 28.

79. A motivation for the central server to “start with the highest quality images” is not equivalent to a motivation for the central server to discard the lowest quality images. If the central server is to “start with the highest quality images,” then it logically cannot discard the lowest quality images, since it would need to keep them in storage until it finished processing the higher-priority images.

80. Moreover, I find no teaching in any of the references suggesting an advantage or even a possibility for the central server to start with the highest quality images. Laganriere363 explicitly discloses that the second video analytics either analyze only one or a few frames selected by the first video analytics or analyze all the frames stored local to the source and later discard them. Ex-1005, [0041]. This is consistent with Laganriere363’s objective of “speeding real-time security response.” *Id.*, [0004]. In image processing, “real-time” means processing visual data instantly as it is received, enabling quick analysis and decision-making. A server aiming to provide real-time response in a video surveillance system cannot arbitrarily change the processing order to start with the highest quality images. This is because the server would not know when or whether even higher-quality images will arrive and cannot afford to postpone the processing of the current image indefinitely while waiting for

potentially higher-quality images. If the goal is to achieve real-time processing, the server would have to analyze all data immediately upon receipt; it could not selectively process only high-quality images first.

81. Regarding Petitioner's argument about using Laganier⁴⁶¹'s disclosures to teach that "the central server purges image from the face log," there is a fundamental incompatibility in the proposed combination. Petition, 28-29. Laganier⁴⁶¹ discloses selective erasing after video-analytics based on the results of video analytics, which is fundamentally different from Laganier⁸¹⁷'s purging immediately after facial detection. I have reviewed claims 3 and 4 of Laganier⁸¹⁷, and these claims reinforce my opinion because they indicate to me that Laganier⁸¹⁷'s purging largely relies on simple criteria that would be immediately available after facial detection and would not rely on any further video analytics. Thus, the operations disclosed in Laganier⁴⁶¹ and Laganier⁸¹⁷ would necessarily occur at opposite ends of the video analytics process. A POSITA would not disregard these fundamental structural differences to create a system that fails to achieve either reference's purpose of reducing the amount of space required to store video data and enabling quality appraisal of captured images.

82. I also disagree with the assertion that "POSITAs would have been motivated to use Laganier⁴⁶¹'s processing at the central server to further reduce the amount of data the central server must process." *See* Petition, 28-29. Laganier⁴⁶¹ discloses

selectively erasing video data *after* the central server has performed video analytics on the video data. Ex-1007, [0002], [0010], FIGS. 3-4. Thus a POSITA would understand that Laganiere461's purging will not reduce the amount of the data the central server must process.

83. Furthermore, Laganiere461 teaches the central server performing selective deletion after video analytics "for reducing the amount of space that is required to store video data." Ex-1007, [0002]. Specifically, the central server performs video analytics "to identify a single frame of data in sequence of frames that provides an image of the face of an individual that is of sufficient quality," and storage of the single high-quality frame for each individual "reduces the storage requirement significantly and facilitates identification by human users." *Id.*, [0028].

84. Given that metadata has a smaller file size compared to images, transmitting lower quality images would require more storage space and processing power than just storing metadata. A POSITA reviewing Laganiere461's explicit teaching of reducing storage requirements (Ex-1007, [0002], [0028]) would be discouraged from, rather than motivated to, transmit lower quality images to the central server, where they would occupy storage space only to be discarded later.

85. In my opinion, Petitioner has failed to demonstrate how and why a POSITA would combine the prior art references in the manner proposed. Even if Laganiere817's quality scoring and face log purging were to be implemented

together in Laganieri363's first video analytics to reduce bandwidth or together in Laganieri461's central server to reduce storage space requirements, such combination does not render the claimed invention obvious because it fails to teach or suggest an "object detection system" that ranks the detected image data corresponding to objects of interest according to different factors and saves the ranking in metadata with the image data, and separately, an "object analysis system" that reads the rankings within the metadata and decides whether or not to analyze the image data.

86. Petitioner's proposed approach of implementing quality scoring in the first video analytics while implementing purging in the central server (Petition, 24-25, 28) would eliminate both the advantages of reduced bandwidth and reduced storage requirements described in the references. In my opinion, such a combination could only be conceived through hindsight.

B. Ground 2

a. Bedros Fails to Disclose or Render Obvious "the video analytics system receives the image data from multiple object detection systems after being transmitted over the network" ([1.4.1], [13.4])

87. Petitioner's argument that Bedros discloses or renders obvious "the video analytics system receives the image data from multiple object detection systems after being transmitted over the network," is flawed. I understand that Petitioner contends

that “Bedros discloses server 308 (‘video analytics system’) receives facial image 116 (‘image data’) over a network from server 306 (‘object detection system’).” Petition, 61. However, this mapping fails to identify any disclosure of “multiple object detection systems” in Bedros.

88. In my opinion, Bedros clearly discloses “one particular server 306” – as contrasted with “one or more servers 308” – that executes image capturing and processing software. Ex-1008, 8:20-22. Petitioner argues that Bedros discloses the computing device depicted in Figure 4 may be server 306 and the computing device can be a “computer 410.” Petition, 61 (citing Ex-1008, 8:55-56, 9:9-17). However, this is unclear and inadequate to demonstrate “multiple” servers 306, especially considering that FIG. 4 clearly shows only a single computer 410.

89. I understand Petitioner further argues that “POSITAs 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.” Petition, 61-62. I disagree with this assertion because Bedros does not disclose a need for multiple cameras monitoring multiple locations.

90. I notice that Petitioner cites the “individuals and their locations” language as support for such a need. Petition, 62. However, examining the full context reveals that this is about locating individuals in images from a monitored environment. Ex-1008, 1:6-10. In my opinion, the “locations” referred to here are locations of

individuals within an image (from one camera), not multiple physical locations requiring coverage by multiple cameras.

91. In fact, the cited materials show that Bedros disparages the multiple-camera solution because of “deficiencies identified such as the added cost of procuring, installing, and maintaining multiple cameras.” *Id.*, 1:45-46. Bedros further disparages a two-camera solution as presenting problems of latency and missed opportunity. *Id.*, 1:46-57.

92. Bedros’ way of saving bandwidth or computational resources is to down-sample a high-resolution video, not extract objects of interest through recognition. Ex-1008, 8:34-39. The deficiencies of multiple cameras that Bedros discusses specifically relate to using two cameras (one high resolution and one low resolution) to monitor the same scene, not using multiple cameras to monitor different locations.

93. I understand Petitioner further cites different embodiments of Bedros as disclosing a motivation to “simultaneously monitor several scenes of interest (e.g., hallways and work areas) associated with different levels of human movement (e.g., remain or do not remain in the same space for extended periods).” Petition, 62 (citing Ex-1008, 5:3-11, 5:26-33). But by presenting different embodiments, Bedros is discussing why spatial proximity may vary depending on, partially, the (singular) monitored scene 100. Ex-1008, 5:25-26. Bedros discusses hallways and work areas as two examples of scenes that require different spatial proximities. *Id.*, 5:26-33.

94. However, Bedros does not present these two examples as two scenes that are to be monitored simultaneously. In fact, Bedros' disparagement of multi-camera systems shows that the different embodiments represent different uses of Bedros' method and system (implemented with a single camera) in different scenarios by different users.

95. I understand that Petitioner also argues that POSITAs would be motivated "to send multiple different image data from multiple locations to server 308 to cost-effectively utilize the service." Petition, 62-63. I have found no evidence to support this assertion. As discussed above, Bedros does not disclose a usage scenario or a need that involves monitoring multiple locations simultaneously. Therefore, in my opinion, a POSITA would not be motivated to modify Bedros particularly in light of its explicit disclosure of "one particular server 306." *See* Ex-1008, 8:20.

b. Bedros does not disclose or render obvious "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" ([1.5.3], [13.5.3])

96. I understand Petitioner argues that Bedros' "queue processing," Ex-1008, 7:19-30, discloses the limitation in question. Petition, 68-71. I disagree. Petitioner identifies step 230 (submitting the facial image to the facial image queue, which may include removing and reordering facial images within the queue) of method 200 as

performing this “queue processing.” *Id.*, 69. However, Bedros clearly distinguishes “queue processing 240” from “submitting 230.” Ex-1008, 7:19-62. I further note that the “removing and reordering” language describes submitting 230, not queue processing 240. Ex-1008, 7:21-25. Therefore, in my opinion, Petitioner’s description of Bedros’ disclosure is inaccurate.

97. I understand that Petitioner’s argument consists of three assertions. First, Petitioner claims that a queue of images resides and is managed on servers 308. Second, Petitioner claims that this queue processing involves removing “certain facial images 116... from the queue of images.” Finally, Petitioner claims that, to remove images from the queue, “POSITAs would have recognized queue processing must first read the quality scores of the images in the queue.”

98. All three claims are incorrect. I will address each in turn.

**i. The alleged “queue processing” is executed by
server 306**

99. I disagree with Petitioner’s characterization that Bedros teaches processing the facial image queue at servers 308. Petition, 69. Bedros explicitly discloses that the queue resides on and is managed by “[o]ne particular server 306” which executes software that implements methods 200, not servers 308. This is clearly stated in Ex-1008, 8:20-22.

100. Furthermore, the “queue processing” (submitting 230) is performed on “facial images 116 within the queue awaiting submission to the facial recognition program.” Ex-1008, 7:23-24. I note that Dr. Houh, Petitioner’s own expert, confirmed this in his deposition. Ex-2001, 108:10-14. *See also* Ex-1003, ¶169.

101. A step of submitting cannot be performed by the servers 308 executing facial recognition software, as these servers are recipients of the submission. Therefore, the “queue processing” can only take place on server 306, otherwise step 230 would not be able to operate on facial images before submission to the facial recognition program.

102. A person of ordinary skill in the art looking at Bedros’ explicit disclosures would understand that none of the images in a queue awaiting submission to the facial recognition program have been submitted to the facial recognition program, therefore none of the images in the queue have “arrive[d]” at the facial recognition program. I understand that Petitioner characterizes the queue as a queue “awaiting processing by the facial recognition software as subsequent images arrive from any of servers 306,” that is, a queue of images that has already arrived at the facial recognition programs running on servers 308. Petition, 69. This characterization is technically incorrect, as the language “subsequent images arrive from any of the servers 306” would make no sense in this context.

103. Regarding Petitioner's alternative argument that "POSITAs would have been motivated to [process the facial image queue at server 308] 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," I disagree with both premises. *Id.*

104. First, there cannot be "other servers 306," as this contradicts Bedros' explicit disclosure of "one particular server 306" (Ex-1008, 8:20). Even if one were to imagine a setup involving multiple servers 306 (contrary to Bedros' explicit disclosure), I do not see any technical advantage associated with knowing the rankings from other servers 306.

105. Second, Bedros clearly teaches one or more servers 308. *See* Ex-1008, 8:23-31. Thus, Bedros contemplates scenarios in which more than one server 308 is present. In this case, no single server 308 will ultimately process all facial images. Petitioner's proposed motivations are unsupported by anything in Bedros' disclosure and appear to be based on hindsight.

106. I also disagree with Petitioner's assertion that "[p]rocessing the queue at server 308 would have been more efficient and reduced bandwidth by relieving servers 306 from reading the queue to [remove and reorder] facial images from it." Petition, 69. This assertion contradicts established principles of efficient network design. In Petitioner's own mapping of limitation [1.5.1]/[13.5.1], I understand Petitioner

acknowledges that optional layer 220 in server 306 applies quality measure scores and/or prioritization results to facial images 116. This means server 306 would inevitably read and analyze image data to filter and prioritize facial images. Based on my experience with video processing systems, if images are not removed by server 306 (not “servers” 306) and sent to server 308 only to be removed there, this will increase bandwidth usage and result in less efficient operation.

107. Regarding Petitioner's argument that “[p]erforming queue processing on image data at the server receiving information was well-known in the art,” (Petition, 69-70) I note that none of the cited references relate to removing images as part of “queue processing” (submitting the facial image to the facial image queue) on the server:

- Laganier⁴⁶¹ discloses purging data after video analytics, not while they are in a queue waiting for processing, Ex-1007, [0024-0025];
- Gali discloses sorting the queue (that is, removing and reordering and putting back in the correct order, as a POSITA would understand the term “removing and reordering” in Bedros; context), not discarding images from the queue, Ex-1021, [0053-0054], [0094-0095], FIG. 1;
- Deever⁹⁶⁵ has nothing to do with removing or reordering data items in a queue, *see* Ex-1015, 14:45-15:35, 16:47-17:33; Ex-1003, ¶229.

108. These references do not support an argument that POSITAs would have been motivated to modify Bedros by processing the facial image queue at server 308. In my professional experience with video processing systems, Petitioner's proposed modification would, in fact, increase—rather than decrease—transmission bandwidth. This is due to the inherent technical constraints associated with the proposed changes, which would introduce additional data overhead rather than optimizing bandwidth efficiency.

109. Regarding the argument that “POSITAs would have been further motivated to separate and delineate tasks performed at [server] 306 and servers 308 to streamline operation and to reduce transmission bandwidth,” I note that Petitioner's expert cites to Yang. Petition, 70

110. Yang merely discloses smart cameras detecting objects of interest and removing background, while “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.” Ex-1003, ¶¶ 230-232, citing Ex-1022, [0026], [0020]. In my opinion, the separation of work between smart cameras and central node or cloud as discussed in Yang is completely different from the claimed invention. I have reviewed claim 10 of Yang and it reinforces my opinion, because it indicates to me that the purpose of Yang's central node is to generate commands and send the commands back to a local processor for it to adjust parameters for analyzing the

sequence of images. Ex-1022, claim 10. There is no technically sound motivation for a POSITA to separate and delineate Bedros' integrated method 200 to be performed by server 306 and servers 308 in the particular way Petitioner proposes, since the essentially lineal method is, in my opinion, most suitable for execution in one locality, and not through the remote-command-passing protocol of Yang, which will unnecessarily complicate the implementation by back and forth communication and increase possibilities of error.

**ii. Bedros does not disclose limitation [1.5.3]/[13.5.3]
because server 308 does not read the rankings
within the metadata**

111. The argument that servers 308 must be able to read the quality scores of the images to perform queue processing is fundamentally flawed for several reasons. Petition, 70.

112. First, I observe that servers 308 do not perform the “queue processing” from step 230. *See supra* ¶¶99-110.

113. Second, even if one were to assume that servers 308 implement some variation of step 230, Bedros never discloses that step 230 is based on quality scores. Instead, Bedros explicitly discloses that step 230 is based on “processing outcomes within the optional processing layer 220.” Ex-1008, 7:24-25. I note that quality measures are calculated in step 210, before the optional processing layer 220. Ex-1008, 6:4-

22, FIG. 2. Therefore, it is my opinion that quality scores are not outcomes within the optional processing layer 220.

114. Third, Bedros never discloses sending anything other than facial images 116 to servers 308. Bedros describes saving quality scores in image data structure 160 of the facial image 116. Ex-1008, 6:16-19. However, Bedros never discloses sending image data structure 160 to servers 308. Bedros only discloses image 116, not image data structure 160, as being submitted to the queue. Ex-1008, 7:19-25. Thus, the only thing that can be submitted from the queue to facial recognition would be image 116, not image data structure 160. Ex-1008, 7:27-30.

115. Furthermore, I note that image data structure 160 also contains received identity and accuracy probability data from facial recognition (performed by servers 308), and these fields would not exist when image 116 is submitted to facial recognition. Ex-1008, 7:42-44. This represents another technical reason why Bedros could not have disclosed sending image data structure 160 to servers 308. Therefore, I understand that Bedros's server 306 maintains image data structure 160 throughout method 200 and does not transmit it to servers 308. Ex-1008, FIG. 2.

116. In my experience working with image processing systems, I know that metadata (resulting from image processing, like object or facial recognition) can be embedded within the image file itself or stored externally in separate files or databases. In an application like Bedros, though not explicitly mentioned, both

approaches are possible. However, the external storage approach is more convenient for this type of system. The most likely approach would be to store it in server 306, since this approach is useful for managing large amounts of metadata or when the metadata needs to be frequently updated without modifying the original image file.

117. Therefore, the quality scores are not accessed by servers 308, and indeed not needed by servers 308 even if servers 308 were to perform a variation of step 230 (which Bedros explicitly allocated, as part of method 200, to server 306, see Ex-1008, 8:20-22).

iii. Bedros does not disclose limitation [1.5.3]/[13.5.3]

**because server 308 does not determine whether to
analyze the image data**

118. Bedros discloses (as part of optional processing layer 220 in server 306) a processing element that discards 222 images when the quality measure of a facial image 116 falls below a threshold value. Ex-1008, 6:38-40. Specifically, Bedros states that “[d]iscarding a facial image 116 may include deleting the facial image 116, removing a record from the image data structure 160, marking an image data structure 160 record for deletion, or otherwise removing the facial image 116 from processing within the method 200 and preventing the facial image 116 from being submitted to the facial recognition program.” Ex-1008, 6:49-55.

119. Bedros further discloses that “[f]ollowing processing of the optional processing layer 220, the method 200 includes submitting 230 the facial image 116 to the facial image queue. In some embodiments, 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 based on processing outcomes within the optional processing layer 220.” Ex-1008, 7:19-25. Bedros therefore explicitly discloses that the “removing and reordering” of step 230 occurs after the optional processing layer 220. This technical sequence means that step 222 has already taken place when step 230 is executed. When step 230 happens, it is processing facial images 116 which server 306 has already discarded by executing the processing element 222 in the optional processing layer.

120. Step 230 should not be understood as implementing another discarding procedure that closely follows the dedicated discarding processing element 222. Such an implementation would provide no technical advantage and would result in duplicative processing and increased probability of error—outcomes that would be technically undesirable in video processing systems.

121. I conclude that the language “removing and reordering” describes an integral process: removing images from the queue, reordering them based on priority values assigned in step 224 just before 230, and then returning the correctly-ordered images to the queue. This achieves the explicitly stated objective that “the facial images 116

that are the most fit for facial recognition processing will be submitted first.” *See* Ex-1008, 7:1-3.

122. Furthermore, even reading the word “removing” separately from the integral phrase “removing and reordering” and interpreting step 230 as implementing some kind of discarding procedure, I would understand that the “removing and reordering” in step 230 is “based on processing outcomes within the optional processing layer 220.” *Id.*, 7:21-25. Therefore, server 308, when solely implementing step 230, would not independently read the quality measure value, compare it with the quality measure threshold, and determine whether to discard, because this logic has already been executed in the optional processing layer 220.

123. Because Bedros does not disclose claim limitations [1.4.1]/[13.4] and [1.5.3]/[13.5.3], it is my opinion that Petitioner fails to meet its burden of showing that independent claims 1 and 13, and all dependent claims, are unpatentable by a preponderance of the evidence under Ground 2.

IX. DECLARATION

124. I declare that all statements made of my own knowledge are true and that all statements made on information and belief are believed to be true, and that these statements were made with the knowledge that willful false statements and the like are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code.

3/26/2025

(Date)



Mingshao Zhang, PhD.

Department of Mechanical and
Mechatronics Engineering
School of Engineering
Southern Illinois University Edwardsville
Edwardsville, IL 62026-1805
Room: EB 3175