

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,
Petitioner,

v.

138 EAST LCD ADVANCEMENTS LIMITED,
Patent Owner.

Case No. IPR2024-00960
U.S. Patent No. 7,486,807

DECLARATION OF NIKOLAOS PAPANIKOLOPOULOS, PH.D.

TABLE OF CONTENTS

I. ASSIGNMENT	1
II. BACKGROUND AND QUALIFICATIONS	2
III. LEVEL OF SKILL IN THE ART AND BASIS FOR OPINIONS	12
IV. RELEVANT LEGAL PRINCIPLES	13
A. Claim Construction.....	13
B. Anticipation	15
C. Obviousness.....	15
V. THE '807 PATENT	16
A. Summary of the '807 Patent Inventions	16
B. Technical Problems Addressed by the '807 Patent Inventions	17
C. Technical Solutions and Benefits Provided by the '807 Patent Inventions ..	19
VI. CLAIM CONSTRUCTION	26
A. "Similar to the object of the image previously contained in the database" means "sufficiently similar to identify the specific object of the image previously contained in the database."	26
1. The Specification of the '807 Patent Supports the Construction.....	27
2. The Claims of the '807 Patent Support the Construction.	32
VII. ANALYSIS OF THE ASSERTED REFERENCES	34
A. Overview of Stubler.....	34
B. Overview of Wang.....	37
VIII. THE TEACHINGS OF STUBLER DO NOT SATISFY ALL ELEMENTS OF ANY OF CLAIMS 1-14 OF THE '807 PATENT.	38
A. Stubler Fails to Disclose "an image storing section for storing the image which is supposed to be classified and retrieved together with a keyword in a database and an object of the image being previously contained in the database."	

1. “An Object of the Image” Must be Usable to Determine Whether the Image to be Classified Includes an Object that is “Similar to the Object of the Image Previously Contained in the Database,” as a POSITA would Understand that Phrase.	40
2. Stubler Discloses “Region Metadata,” which Merely Consists of High-Level Descriptions of Particular Regions.	44
3. Stubler’s “Region metadata” Is not Usable to Determine Whether the Image to be Classified Includes an Object that is “Similar to the Object of the Image Previously Contained in the Database,” as a POSITA would Understand that Phrase.	48
B. Stubler fails to disclose “an object information inputting section for confirming by the user, adding, and correcting the keyword which is proposed by the keyword proposing section when the inputted object acknowledged by the object acknowledging section is similar to the object of the image previously contained in the database.”	52
IX. THE TEACHINGS OF STUBLER IN COMBINATION WITH WANG DO NOT SATISFY ALL ELEMENTS OF ANY OF CLAIMS 2-5 OR 9-12 OF THE ’807 PATENT.	53
X. ADDITIONAL REMARKS	53

I, Nikolaos Papanikolopoulos, of Minneapolis, Minnesota, declare that:

I. ASSIGNMENT

1. I have been retained by Robins Kaplan LLP on behalf of 138 East LCD Advancements Ltd. (“138 East” or “Patent Owner”) as an expert with regard to this matter. I understand that Petitioner Google LLC (“Google” or “Petitioner”) is requesting that the Patent Trial and Appeal Board (“Board”) institute *inter partes* review (IPR) of U.S. Pat. No. 7,486,807 (“the ’807 patent”) (Ex. 1001).

2. In my capacity as an expert, I have been asked to provide my independent analysis and opinions regarding the ’807 patent, the prosecution history of the ’807 patent, the alleged prior art relied upon by Petitioner, the meaning of various claim terms from the ’807 patent from the perspective of a person of ordinary skill in the art (“POSITA”), and the opinions offered by Petitioner’s expert, Dr. Robert Stevenson, in his declaration (“Stevenson Declaration,” Ex. 1003). My opinions and conclusions are fully discussed in later sections of this declaration.

3. I am an independent consultant, and I am not, and have never been, an employee of 138 East or Google. I received no compensation for this declaration beyond my normal hourly compensation (\$450) based on my time spent analyzing the ’807 patent, the prosecution history of the ’807 patent, the prior art publications relied upon by Petitioner, the Stevenson Declaration, and other related materials. I

will not receive any added compensation based on the outcome of any IPR or other proceedings involving the '807 patent.

4. This declaration and my opinions are in support of Patent Owner's Response, and I understand that the content of this declaration may be expressed and reproduced in the POR.

II. BACKGROUND AND QUALIFICATIONS

5. I am the McKnight Presidential Endowed Professor and Distinguished McKnight University Professor in the Department of Computer Science and Engineering of the University of Minnesota in Minneapolis, Minnesota. I am also the Director for the Minnesota Robotics Institute.

6. I have been retained as an expert in the fields of image processing/computer vision, sensing, and related technologies. I am being compensated at my normal consulting rate of \$450 per hour. My compensation is not dependent on and in no way affects the substance of my statements in this declaration.

7. I have been a professor at the University of Minnesota (originally as an assistant professor, and then as an associate professor) since the Fall 1992. Between Fall 2001 and Spring 2004, and between Fall 2010 and Spring 2013, I was the Director of Undergraduate Studies of the Department of Computer Science and Engineering.

8. In 1992, I received my Ph.D. in Electrical and Computer Engineering from Carnegie Mellon University. My thesis was entitled “Controlled Active Vision” and focused on using computer vision in a controlled fashion to detect, track, and manipulate objects in the environment. In 1988, I also received my M.S. in Electrical and Computer Engineering from Carnegie Mellon University. My B.S. in Electrical Engineering was received in 1987 from the National Technical University in Athens, Greece.

9. For over thirty years, my research and teaching work has focused on image processing, computer vision, intelligent transportation systems, and robotics. This research has included detection and tracking of humans and objects such as faces using information from stationary or moving cameras. It has also included main object detection and recognition including work with artificial intelligence and pattern recognition systems for surveillance applications.

10. I currently teach or have been teaching three courses related to image processing and computer vision: (i) CSci 5561 Computer Vision, (ii) CSci 5511 Artificial Intelligence, and (iii) CSci 5551 Introduction to Intelligent Robotic Systems. In particular, I have been teaching the class in Computer Vision since 1992.

11. I am currently serving as a Senior Editor in the IEEE Transactions on Intelligent Transportation Systems with responsibility in the area of imaging. I was Associate Editor in the same journal for more than twenty years.

12. I have received research grants for various projects involving computer vision and image processing as applied to medicine and surveillance, including face detection and tracking of humans. My industrial experience includes founding a robotics company in 2005 named ReconRobotics Inc. that develops different robotic platforms. I have also been a consultant since 1997 for numerous companies and provided technical expertise in developing computer vision algorithms for human tracking, object recognition, fingerprint recognition, vehicle tracking, camera calibration, inspection, and precision agriculture. In one of my consulting assignments, I had to develop computer vision algorithms for inspection of industrial parts by using different illumination sources (Banner Engineering). I have licensed my algorithms to several industries with the most recent one being Sentera Inc., which is using image processing and image segmentation for precision agriculture.

13. My research has produced more than 380 journal and conference publications. More than 90 publications are in refereed journals including top journals in imaging such as IEEE Trans. on Pattern Analysis and Machine Intelligence, IEEE Trans. on Image Processing, and International Journal of

Computer Vision. Many of my publications and research grants relate to human detection and tracking systems. Some examples include:

Walczak, N., Fasching, J., Cullen, K., Morellas, V., and Papanikolopoulos, N.P., "Toward Identifying Behavioral Risk Markers for Mental Health Disorders: An Assistive System for Monitoring Children's Movements in a Preschool Classroom," *Machine Vision and Applications*, Volume 29, No. 4, May 2018, PP. 703-717.

Cook, D., Morris, T., Morellas, V., and Papanikolopoulos, N., "An Automated System for Persistent Real-Time Truck Parking Detection and Information Dissemination," *Proceedings of the 2014 IEEE Int. Conference on Robotics and Automation*, Hong Kong, China, May 31 – June 7, 2014, PP. 3989-3994.

Somasundaram, G., Sivalingham, R., Morellas, V., and Papanikolopoulos, N.P., "Classification and Counting of Composite Objects in Traffic Scenes Using Global and Local Image Analysis", *IEEE Trans. on Intelligent Transportation Systems*, Volume 14, No. 1, March 2013, PP. 69-81.

Veeraraghavan, H., Atev, S., Bird, N., Schrater, P., and Papanikolopoulos, N., "Driver Activity Monitoring through Supervised and Unsupervised Learning," *Proceedings of the IEEE 8th Int. Conf. on Intelligent Transportation Systems (ITSC 2005)*, PP. 895-900, Vienna, Austria, September, 2005.

Wahlstrom, E., Masoud, O., and Papanikolopoulos, N., "Vision-based Methods for Driver Monitoring," *Proceedings of the IEEE 6th Int. Conf. on Intelligent Transportation Systems (ITSC 2003)*, PP. 903-908, Shanghai, China, October 12-15, 2003.

Masoud, O., and Papanikolopoulos, N.P., “A Novel Method for Tracking and Counting Pedestrians in Real-time Using a Single Camera,” *IEEE Trans. on Vehicular Technology*, Volume 50, No. 5, September 2001, PP. 1267-1278.

Singh, S., and Papanikolopoulos, N. P., “Monitoring Driver Fatigue Using Facial Analysis Techniques,” Proceedings of the 1999 IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems, PP. 314-318, Tokyo, Japan, October 5-8, 1999.

14. I have worked extensively in the use of sensors to monitor driver faces and fatigue. The work extended to monitor driver fitness to drive. We have developed a computer vision-based system that detects faces and analyzes facial features and eye movements to detect signs of drowsiness and driver fatigue. By monitoring factors such as eye closure, head pose, and blink patterns, the system can alert drivers when they show signs of fatigue and suggest appropriate actions (this research started in the late 90s and evolved through the next decade). It should be noted that the proposed algorithms and systems considered the continuously evolving light conditions that vehicular systems operate within.

15. It should be noted that our group has made significant contributions to the field of face detection and tracking (the face is the main object segmented), particularly in the context of security and transportation applications. Our research includes the development of robust face detection algorithms capable of operating in complex, real-world environments with varying lighting conditions and

orientations. My team has focused on creating real-time tracking systems that can maintain performance even with partial occlusions or rapid movements.

16. Our research has explored multi-modal approaches, combining visual and audio cues for more accurate tracking, and has addressed privacy concerns by developing techniques that can detect and track faces without necessarily identifying individuals. In the realm of security and surveillance, we have applied our face detection and tracking techniques to crowd monitoring and anomaly detection. Moreover, we have implemented adaptive learning techniques to improve face detection over time, extended this work to 3D face modeling, and explored emotion and expression recognition. Throughout my career, I have consistently emphasized the importance of scalability and efficiency, working to optimize algorithms for use in resource-constrained environments. Some results from the aforementioned efforts are shown in the figures below.

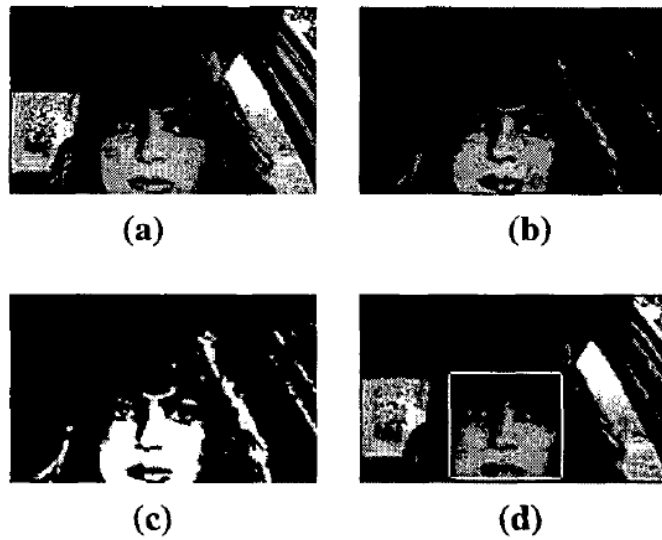


Figure 1: It is taken from the paper “Monitoring Driver Fatigue Using Facial Analysis Techniques” and shows detection of face using skin color information.



Figure 2: It is taken from the paper “Vision-based Methods for Driver Monitoring and shows driver features being tracked.



Figure 3: It is taken from the paper “Driver Activity Monitoring through Supervised and Unsupervised Learning” and shows skin color detection (bottom row) on various images (top). Skin color is indicated in black. The results are post-processed by a sequence of morphological erosions and dilations.

17. My work in vehicle detection and tracking has focused on using image and video processing techniques, including object detection, tracking, and recognition, to accurately track vehicles in real-time. These algorithms have been applied to different scenarios, such as urban traffic monitoring and surveillance. Moreover, machine learning techniques have helped in the improvement of these methodologies.

18. My work has been extended to tracking autonomous vehicles, where the challenges are different from traditional vehicle tracking that uses information from static sensors (e.g., cameras). My interests include developing tracking algorithms and systems that can handle the complexities of autonomous vehicles,

including predicting their future movements and understanding their intentions.

The work has been extended to monitor the intentions of pedestrians (where pedestrians are segmented as the main objects).

19. These vehicle tracking systems enable continuous monitoring and tracking of vehicles, providing valuable information for traffic management, law enforcement, and security purposes.

20. My paper "Detection and Classification of Vehicles," published in the IEEE Transactions on Intelligent Transportation Systems in March 2002 is one of the most cited papers in the history of this journal (more than a thousand citations). The paper among other contributions shows how to segment the main object from videos and images in low-light conditions. I also have nine patents in pertinent areas. One of them is the "Crowd Counting and Monitoring" with P. Kilambi and O. Masoud. It discloses a system and method for accurately estimating the number of people in a crowded area and monitoring their behavior. The invention utilizes computer vision techniques and machine learning algorithms to analyze video feeds from cameras overlooking the target area. It employs techniques such as background subtraction, main object detection, and tracking to identify individual people within the crowd.

21. The system uses density estimation methods to handle areas where individuals are difficult to distinguish. Advanced features might include the ability

to detect and track movement patterns, identify unusual behaviors or potential security threats, and predict crowd flow.

22. Our research work has been funded by several government funding agencies. For example, we have received funding from the Department of Homeland Security to monitor threats at Mass Transit Sites like the 30th Street Station in Philadelphia, PA, and the Light Rail Station in the Minneapolis-Saint Paul International Airport. Tracking humans was an essential task for which we developed a variety of image processing algorithms and tools.

23. For more than a decade, we have been developing computer vision-based systems to assess truck parking availability at truck rest areas. The funding came from FHWA, MnDOT, WiscDOT, and KansasDOT. The approach uses multiple cameras to detect trucks and other vehicles and determine empty spaces. The system operates in all weather conditions around the clock.

24. I have received numerous honors and awards for my research and contributions. I have been a Distinguished McKnight University Professor at the University of Minnesota since 2007 and have been a McKnight Presidential Endowed Professor in Computer Science since 2016. Also in 2007, I was nominated and became an IEEE Fellow. In 2016, I received the IEEE RAS George Saridis Leadership Award in Robotics and Automation as well as the Center for Transportation Studies Research Partnership Award. I have also received the IEEE

VTs 2001 Best Land Transportation Paper Award for the paper "A Novel Method for Tracking and Counting Pedestrians in Real-Time Using a Single Camera" (with Osama Masoud).

25. As a result of my work and research, I am familiar with the design, operation, and functionality of systems such as those described in the '807 patent.

26. Further details regarding my employment and academic history are included in my curriculum vitae, attached as Exhibit A.

III. LEVEL OF SKILL IN THE ART AND BASIS FOR OPINIONS

27. Mr. Stevenson contends that a POSITA "would have had at least a bachelor's degree in electrical engineering, computer science, or related discipline, and two years of experience in digital image analysis and processing." Ex. 1003 ¶ 49. I apply Dr. Stevenson's definition of a POSITA for purposes of this declaration, which does not affect my opinion that the Petitioner fails to show that the challenged claims are unpatentable in view of the asserted prior art references.

28. Based on my experience and record of accomplishments, I am qualified to render opinions regarding the technology claimed and described in the '807 patent. Based on my expertise and qualifications, I am qualified to provide an opinion as to what a POSITA would have understood, known, or concluded as of approximately 2003.

29. In reaching my opinions and conclusions, I have relied upon my education, my work experience in the relevant field, and my training, and considered and relied upon my review and analysis of the Petition and all Exhibits.

IV. RELEVANT LEGAL PRINCIPLES

30. For the purposes of this declaration, I have been informed about certain aspects of the law that are relevant to my analysis and opinions. I have applied these legal principles in rendering my opinions below.

A. Claim Construction

31. I understand that the ordinary and customary meaning of a claim term is the meaning that the term would have to a POSITA at the time of the invention when read in light of the specification and the prosecution history of the patent.

32. In the absence of an express intent on the part of the inventor to give a special meaning to the claim terms, the words are presumed to take on the ordinary and customary meanings attributed to them by a POSITA when read in light of the specification and prosecution history.

33. I understand that it is the use of the words in the context of the written description, and as customarily used by those skilled in the relevant art, that accurately reflects both the ordinary and the customary meaning of the terms in the claims.

34. I understand that the basis for a term's ordinary and customary meaning may be derived from a variety of sources, including the words of the claims themselves, the remainder of the specification, the prosecution history (collectively referred to as "intrinsic evidence"), and "extrinsic evidence" concerning relevant scientific principles, the meaning of technical terms, and the state of the art at the time of the invention.

35. I have been instructed that dictionary definitions or definitions from technical references can be used as extrinsic evidence to inform or confirm the ordinary and customary meaning of words found in a claim, but that in construing claim terms, the general meanings gleaned from reference sources, such as dictionaries, must always be compared against the use of the terms in the context of the claim itself and the intrinsic evidence.

36. I understand that a patent applicant is entitled to be his or her own lexicographer (in other words, provide his or her own meaning to a word or phrase) and may rebut the presumption that claim terms are to be given their plain and ordinary meaning. To do so, the applicant must clearly set forth a definition of the term that is different from its ordinary and customary meaning. Where the applicant provides an explicit definition for a term, that definition will control interpretation of the term as it is used in the claim in which it appears. I understand that the specification can also be relied on for more than just explicit lexicography

to determine the meaning of a claim term. For example, I understand that the meaning of a particular claim term may also be determined by implication, that is, according to the usage of the term in the context of the specification.

B. Anticipation

37. I understand that under U.S. patent law, 35 U.S.C. § 102, a claim is invalid as anticipated if the claimed invention was disclosed in a single prior art reference before the invention thereof by the patentee.

38. I am informed that anticipation requires that a single reference describe the claimed invention with sufficient precision and detail to establish that the subject matter existed in the prior art.

39. I understand that ambiguous references do not, as a matter of law, anticipate a claim.

C. Obviousness

40. I understand that under U.S. patent law, 35 U.S.C. § 103, a claim is invalid as obvious if the differences sought to be patented 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 said subject matter pertains.

41. I am informed that an obviousness analysis requires an assessment of the scope and content of the prior art, the differences between the art and the

claims at issue, and the level of ordinary skill in the art. I am told that it is against this backdrop that obviousness is assessed.

42. I am informed that a POSITA is a hypothetical person who is presumed to be aware of all the pertinent prior art. I am also informed that an obviousness analysis may take account of the inferences and creative steps that a POSITA would employ.

V. THE '807 PATENT

A. Summary of the '807 Patent Inventions

43. The '807 patent claims improvements in the computing field of classifying and retrieving digital image files to facilitate adding and retrieving images using keywords. In particular, the inventions claimed in the '807 patent improve the way that image retrieving devices associate and store manipulable keyword data characterizing digital image files. The inventions claimed in the '807 patent improve the way that prior image retrieving devices operate and handle keywords. Specifically, the inventive image retrieving device stores objects or data representing objects in digital images with associated keywords in a database, detects data representing objects in new, incoming image data, proposes keywords for the new, incoming image data to the user based on the detected data representing objects in the incoming image data and the objects stored in the database, and updates the keyword metadata for stored images based on the user's

response. The '807 patent's implementation of a computing system with an improved data structure—particularly the way in which the inventions detect objects in new, incoming image data and propose associated keywords based on the detection and the user update operations—improve computing systems compared to other approaches. The '807 patent's claimed inventions improve upon, for example, prior computing systems that required users to provide their own keywords to associate with, or describe, digital image data, which was laborious and could result in inconsistent keywords being associated with images containing the same objects.

B. Technical Problems Addressed by the '807 Patent Inventions

44. The '807 patent addresses several technical problems known in the field of classifying and retrieving digital image files with characterizing data. The '807 patent describes five known devices and techniques for manipulating characterization-based metadata in a database for storing digital images, but notes the limitations of each of these prior art systems. EX1001 at 3:1-48. The '807 patent identifies a need for new, improved devices and techniques for computing systems that associate and store manipulable keyword metadata with digital image files.

45. As one example, finding a class of objects is different from finding a particular instance of this class of objects. Finding a class of objects, for instance,

involves training computer vision systems to recognize general categories like “table,” “chair,” or “car” based on shared visual characteristics across many examples. These object detection systems learn to identify common features that defined the class regardless of variations in size, color, or style. This approach may rely on large datasets containing thousands of labeled examples to create models that could generalize well to new instances of the class. In such class identification systems, the system did not need to know anything about specific individual objects - it simply needed to recognize the visual patterns that make, for example, a table distinguishable from other furniture. Another example is the detection of objects like vehicles in traffic images. Yet another example would be detection of a human being in an image. Classes of objects are nearly infinite.

46. A face is an example of an object of interest in the '807 patent. Finding a specific face like “Jiro’s face” in the '807 patent represents a fundamentally different challenge than identifying a general object in the “face” class. Finding “Jiro’s face” requires instance recognition (far more complex than class recognition). This task demanded that the system identify one particular face that was seen before and was part of the database with unique characteristics that distinguished it from all other faces. The system must learn the distinctive visual features of “Jiro’s” specific face - perhaps its unique texture, structure, skin color, or identifying marks - rather than the general “face.”

C. Technical Solutions and Benefits Provided by the '807 Patent Inventions

47. The '807 patent claims a technical solution to these problems: a new, improved computing system including a database component and related techniques for associating and storing manipulable keyword metadata with digital image files and objects appearing in the digital image files. The improved device stores, in a database, digital images and associated keywords describing objects in the images, detects objects (such as people) in the image when the image is classified, and efficiently proposes keywords for a newly inputted image by detecting and matching an object in the newly inputted image with an object stored in the database. *Id.* at 3:51-4:4.

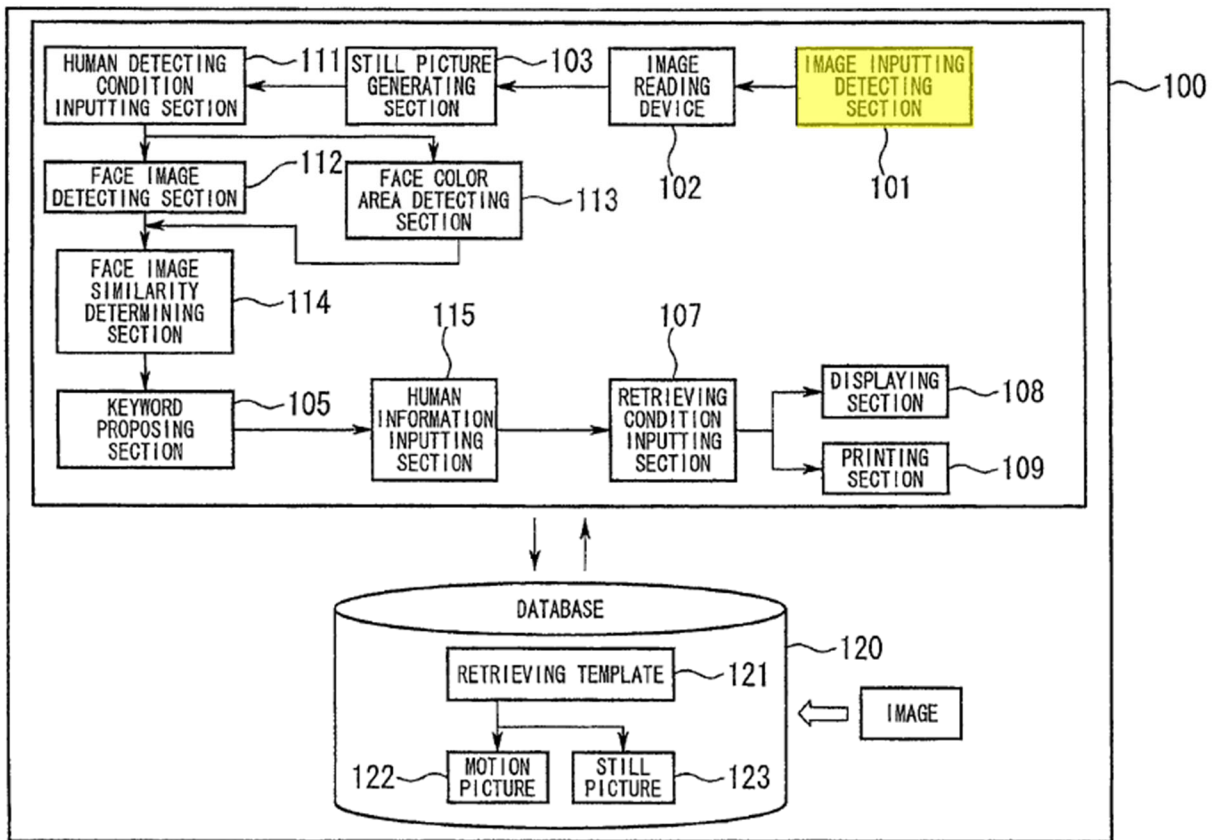
48. According to the '807 patent, “images which are supposed to be classified and detected are saved in [a] database together with the[ir] keywords by the image storing section.” *Id.* at 4:5-7. An exemplary database entry is shown in Fig. 6 of the '807 patent:

FIG.6

SUPERIOR KEYWORD 1	SUBORDINATE KEYWORD 2	NAME	BIRTHDAY	IMAGE
FAMILY	SECND ELDEST BROTHER	JIRO	JANUARY 1, 2003	IMAGE 1:TAKEN ON JANUARY 1, 2003 IMAGE 2:TAKEN ON FEBRUARY 1, 2003 IMAGE 3:TAKEN ON MARCH 1, 2003

49. In general, the system can acknowledge “a new predetermined object which is inputted” as part of a new image, for example, a human or a car. *Id.* at 4:7-10. Image detection may be performed by an image inputting detection section (annotated in yellow, below), as shown in Fig. 2.

FIG. 2

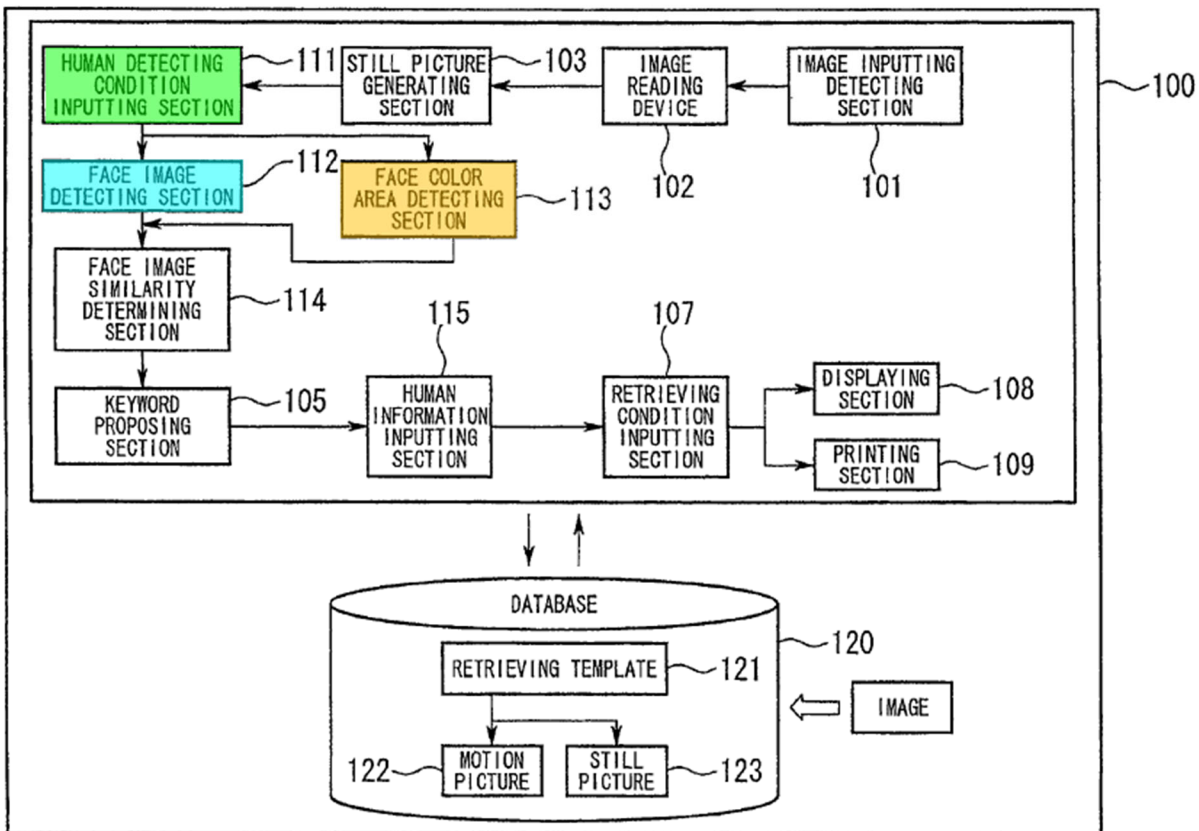


50. After detecting an image, the system's human detecting condition inputting section "sets up a condition for determining whether or not a human is contained in an image" (annotated in green below) *Id.* at 8:15-17. For example, the system detects the presence of a human face either by detecting a face image (annotated in blue below) or face color (annotated in orange below). *Id.* at 8:17-20. The face image detecting section "detects . . . whether or not a face image is contained in the image by analyzing the image." *Id.* at 8:23-25. The '807 patent describes particular rules for carrying out the face detection, including through

frequency analysis or matching with pre-stored patterns. *Id.* at 8:25-31. As a specific example, the '807 patent discloses that detecting the presence of a face may be done by using skin color detection as disclosed in Japanese Unexamined Patent Application, First Publication No. Hei 8-221547 (“Wakayama”).¹ *Id.* at 11:22-30.

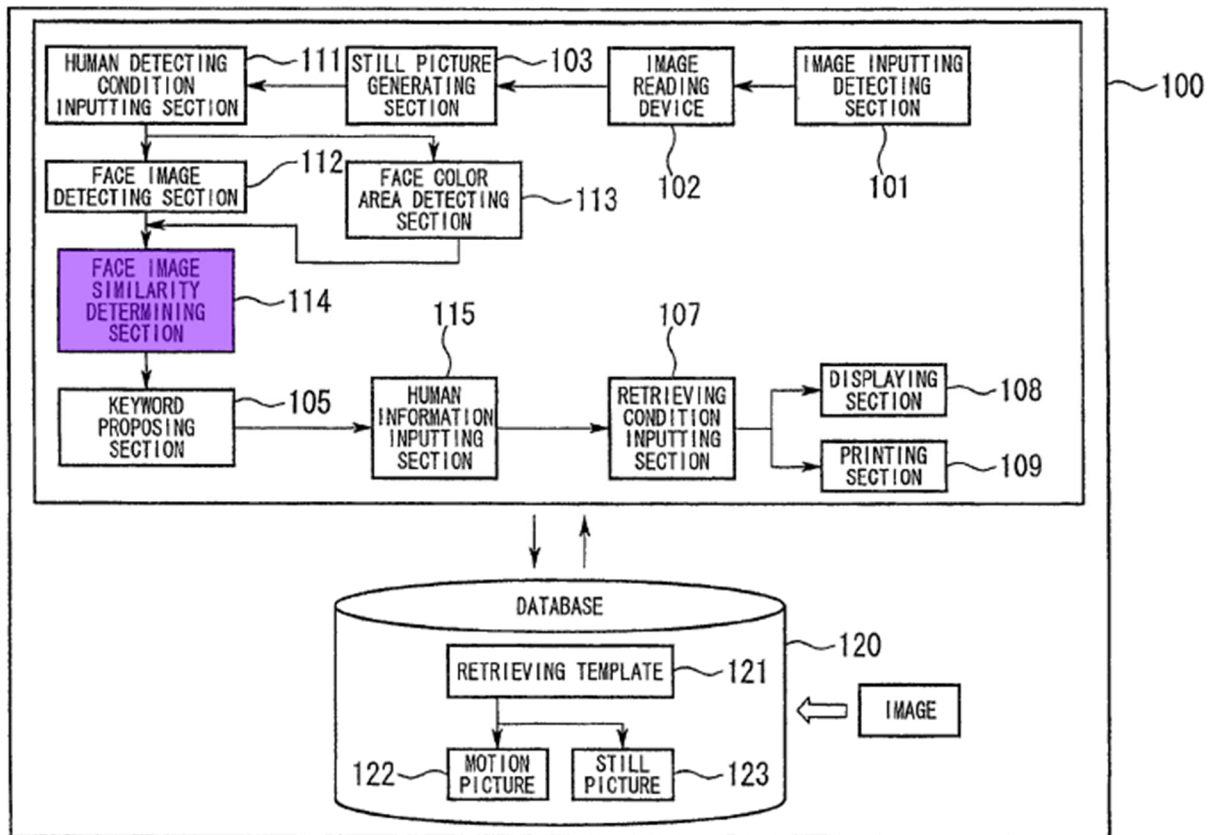
¹ I understand that a copy of Japanese Unexamined Patent Application, First Publication No. Hei 8-221547 is being submitted as EX2003. I also understand that a certified translation of Japanese Unexamined Patent Application, First Publication No. Hei 8-221547 is being submitted as EX2004.

FIG. 2



51. If the system detects a face, it applies further rules to “check whether or not a similar face image to the detected face image is contained” in the database; for instance, the system may match images using image density information. *Id.* at 8:60-67. This operation is performed, in the example of Figure 2, by the face image similarity determining section, highlighted in purple below.

FIG. 2



52. According to the '807 patent, the face image similarity determining section may generate a mosaic pattern of the face image (*e.g.*, a representation of an object) which can then be compared to other mosaic patterns (*e.g.*, representations of an object) stored in the database to identify specifically similar faces. *Id.* at 11:58-67. The '807 patent states "it is possible to employ a method which is disclosed in Japanese Unexamined Patent Application, First Publication No. Hei. 8-221547 for determining such similarity by forming a density image of a face so as to classify the category." *Id.* at 8:63-67; *See* EX2003, EX2004.

53. Thus, the invention of the '807 patent engages in a two-step process of first acknowledging or detecting the presence of an object in an input image, and then comparing the acknowledged object to other objects stored in the database to identify similar objects. According to the '807 patent, by this process “it is possible to retrieve the image which contains a *specific person* accurately and efficiently among numerous images.” EX1001 at 4:51-54 (emphasis added); *see also, id.* at 4:22-24 (“thus, it is possible to retrieve the image which contains a *specific object* among numerous images accurately and efficiently.” (emphasis added)). In other words, the comparison in the '807 patent is a separate step (*e.g.*, step 114 in Fig. 2) and goes beyond the generic similarity (*e.g.*, broad categorical similarities such as “person” or “car”) between objects determined in the first step (*e.g.*, steps 112 and 113 in Fig. 2). The image similarity determining step uses object similarity determination to expand the generic characterization to identify *specific* objects across images. In the parlance of the '807 patent, a “similar” object is a specific object such as a specific person or a specific car (*etc.*) and include those objects within images previously contained in the database that “coincide” with or are “very similar” to the object identified in the detected image. *Id.* at 9:18-22.

54. If the system acknowledges a similar, specific object (*e.g.*, a face of a specific person) in an input image, the system uses a priority-based ruleset to retrieve keyword metadata associated with that specific object from the database to

present to the user for confirmation or modification. *Id.* at 9:2-6. When prompted with specific proposed keywords retrieved from the database, the user can select or correct an existing keyword or add a new one associated with the input image. *Id.* at 12:5-10.

55. The '807 patent therefore recites a new, improved computing system and related techniques to associate and store manipulable keyword metadata with digital image files. In particular, the novel structure of the image input and retrieval system provides a specific, technical solution to recognized shortcomings in prior computing systems that required users to provide their own keywords to associate with digital image data. The system improves on these prior computing systems through a particular arrangement of components that (1) detect specific objects in images when they are first input to the system, (2) retrieve and propose associated keywords from a database based on similarities between the acknowledged objects and stored objects, and (3) propose, update, and/or correct the keywords according to user input as more images containing the same objects) are input to the system.

VI. CLAIM CONSTRUCTION

- A. **“Similar to the object of the image previously contained in the database” means “sufficiently similar to identify the specific object of the image previously contained in the database.”**

56. Claim 1 of the '807 patent recites “an object information inputting section for confirming by the user, adding, and correcting the keyword which is

proposed by the keyword proposing section when the inputted object acknowledged by the object acknowledging section is *similar to the object of the image previously contained in the database.*” EX1001 at cl. 1 (emphasis added). Independent claims 7 and 8 contain similar limitations. *See id.* at Cls. 7, 8. It is my opinion that a POSITA would have understood the phrase “similar to the object of the image previously contained in the database,” when read in light of the claims, specification, and prosecution history, to mean “sufficiently similar to identify the specific object of the image previously contained in the database.”

1. The Specification of the '807 Patent Supports the Construction.

57. It is my opinion that the specification of the '807 patent supports the construction of “similar to the object of the image previously contained in the database” to mean “sufficiently similar to identify the specific object of the image previously contained in the database.” In particular, it is my opinion that the '807 describes the invention in such a way that makes clear that the invention requires more than generic similarity between objects.

58. The specification describes “similarity” using language that makes clear to a POSITA that the level of “similarity” is higher than generic similarity; it is similarity to the level of a specific object. The '807 patent explains that according to the invention “it is possible to retrieve the image which contains a

specific object among numerous images accurately and efficiently.” EX1001 at 4:22-24. The exemplary embodiment related to detecting/classifying pictures of humans further supports the understanding that the invention relates to determining similarity of specific persons. The ’807 patent explains that, according to the invention, “it is possible to retrieve the image which contains a specific person accurately and efficiently among numerous images.” *Id.* at 4:51-53. Thus, it is my opinion that a POSITA would readily understand that references to a “specific object” and a “specific person” means more than high-level, categorical similarity.

59. It is my opinion that the specification of the ’807 patent reinforces this construction by stating that the image retrieving device determines whether “the face image coincides to the face image which is contained previously” or whether “the face image is very similar to the face image which is contained previously.” *Id.* at 9:18-20. By using words like “coincides” and “very similar,” it is my opinion that the specification makes clear to a POSITA that the ’807 patent uses the term “similar” to mean more than just generic, categorical similarity.

60. Consistent with the above, it is my opinion that the specification of the ’807 patent further supports the construction by consistently describing a two-step process by which incoming images are analyzed to identify similar (*i.e.*, specific) objects to those previously stored in the database. In the first described step, the image retrieving device detects a general type of object, such as a human

face.² EX1001 at 8:14-21; *see also id.* at Fig. 2, reference numerals 111, 112, and 113. As explained with reference to Fig. 3, “when a new image is contained in the image retrieving device 100, the image retrieving device 100 tries to detect a human in the contained image. If a human is detected, the image retrieving device 100 expedites a user to input the human information by a display form.” *Id.* at 9:8-13. Similarly, as explained with respect to Fig. 1, “[t]he object acknowledging section 104 acknowledges whether or not a specific object is contained in an image by analyzing the image.” *Id.* at 7:29-31.

61. Once a human is detected, the system proceeds to the second step where the detected human face is compared to determine whether a similar (i.e., specific or “very similar”) human face has previously been stored in the database. *Id.* at 8:60-67; *see also id.* at Fig. 2, reference numeral 114. Specifically, the ’807 patent teaches “[i]f it is determined that the face image coincides to the face image which is contained previously or that the face image is very similar to the face image which is contained previously, the information for such a human is proposed

² The specific exemplary embodiments of the ’807 patent are presented in the context of human detection. Accordingly, this Patent Owner Response also discusses the invention in that context. The ’807 patent makes clear, however, that the invention is not limited to the human object embodiment. EX1001 at 13:13-20.

so as to expedite the user to confirm.” *Id.* at 9:18-22. A specific method for determining similarity is Wakayama, cited in the ’807 patent. EX1001 at 8:63-67.

62. It is my opinion that the specification’s description of this two-step process further supports the construction because the second step would be unnecessary if “generic” similarity was all that was required. If mere “generic” similarity was all that “similar” means, then one could stop after detecting a human in the first step and simply present the information for all pictures in the database that also contain a human, regardless of whether or not those pictures contain the same or a very similar human. A user would then need to manually add and/or sort through keyword information for potentially hundreds or thousands of persons to determine which specific keywords should be added to the newly input picture. Such a process would be highly inefficient in contravention of the ’807 patent’s stated goal of “add[ing] keywords efficiently by detecting the object (such as a human) in the image.” *Id.* at 3:32-34. Thus, it is my opinion that a POSITA would understand based on the two-step process of identifying an object generally and then comparing it to previously stored objects to determine whether it is similar to any of the previously stored objects to support the construction of “similar to the object of the image previously contained in the database” to mean “sufficiently similar to identify the specific object of the image previously contained in the database.”

63. It is my opinion that the specification further supports the construction of “similar to the object of the image previously contained in the database” to mean “sufficiently similar to identify the specific object of the image previously contained in the database” through the exemplary database entry described with respect to Figure 6. As shown below, in Figure 6, database records are detailed to the level of identifying specific people. In Figure 6, the record is for “Jiro,” the second eldest brother:

FIG. 6

SUPERIOR KEYWORD 1	SUBORDINATE KEYWORD 2	NAME	BIRTHDAY	IMAGE
FAMILY	SECND ELDEST BROTHER	JIRO	JANUARY 1, 2003	IMAGE 1:TAKEN ON JANUARY 1, 2003 IMAGE 2:TAKEN ON FEBRUARY 1, 2003 IMAGE 3:TAKEN ON MARCH 1, 2003

64. According to the two-step process discussed above, when a new image of “Jiro” is detected and ready to be classified, the image retrieving device must first conclude that the image (1) contains an image of a human and (2) contains the specific person, “Jiro,” before recommending that that the new image be categorized as containing “Jiro.” As discussed above, the ’807 patent presents the information for “Jiro,” only “if it is determined . . . that the face image is very

similar to the face image which is contained previously.” *Id.* at 9:18-22 (emphasis added).

65. It is therefore my opinion that the specification of the ’807 patent consistently uses words and examples that convey to a POSITA that “similar to the object of the image previously contained in the database” means “sufficiently similar to identify the specific object of the image previously contained in the database.”

2. The Claims of the ’807 Patent Support the Construction.

66. It is my opinion that the claims of the ’807 patent further support the construction of “similar to the object of the image previously contained in the database” to mean “sufficiently similar to identify the specific object of the image previously contained in the database.”

67. For example, dependent claim 2 recites that the object acknowledging section includes “a face image similarity determining section” that “detect[s] a similar face image stored in the database.” A POSITA would understand that this recitation must go beyond mere generic similarity because substantially all faces share generically similar features (*e.g.*, two eyes, a nose, a mouth, all features placed in roughly the same location, within a certain range of color tones, etc.). Thus, if mere generic similarity were all that the claim required, every single face in the database would be determined to be “similar” to every other face.

68. It is my opinion that the claims make clear, however, that the claimed device must be able to tell that some faces are not similar to others because dependent claim 5, for example, provides additional operations for when just such a determination is made. Specifically, claim 5 recites that “in a case in which the similar face image is not detected by the face image similarity determining section” then the device proposes “the keyword to the image which was inputted last.” EX1001 at cl. 5. In other words, if the device does not determine that the newly detected image contains a specific face that has previously been stored in the database, then the device will simply propose the same keywords as were entered for the most recently categorized image. *Id.*

69. Because all faces are generically similar, and the claims explicitly encompass embodiments in which the device determines that two faces are not similar, then the claim term “similar to the object of the image previously contained in the database” must mean more than generic similarity. Thus, it is my opinion that the claims further support the conclusion that a POSITA would understand “similar to the object of the image previously contained in the database” to mean “sufficiently similar to identify the specific object of the image previously contained in the database.”

VII. ANALYSIS OF THE ASSERTED REFERENCES

A. Overview of Stubler

70. Stubler describes a method for generating “semantic labels” for an image based on similarities between metadata of the image and the metadata of previously stored images. EX1005 at ¶[0009]. Specifically, Stubler recites “a) acquiring an image . . . ; b) automatically extracting metadata from the acquired image . . . ; c) automatically selecting one or more stored images having metadata similar to the extracted metadata . . . ; and d) generating one or more captions or labels for the acquired image from one or more preexisting captions or labels associate with the selected stored images. *Id.*

71. Stubler identifies three different types of metadata that can be extracted. “Capture metadata” is information captured at the time the image captured, such as time/date information, GPS data, etc. *Id.* at ¶[0029]. “Global metadata” includes “whole image measurements based on color histograms [or] texture features.” *Id.* “Region metadata” includes “detected regional features, such as face, skin, sky, grass, open space [or] background vs. foreground.” *Id.* at ¶[0030]. Although Stubler does not itself set forth any detailed algorithm for detecting region metadata, it does incorporate by reference certain other references that purportedly disclose methods of detecting region metadata. *See id.*

72. According to Stubler, “acquired images are clustered or grouped by metadata similarity (120) based upon one or more of the foregoing objectively measurable or determinable features.” *Id.* at ¶[0031]. Notably, this clustering is done with respect to the “acquired images” before any comparison between the acquired images and stored images. *Id.* (“acquired images are clustered Once the data is clustered or grouped, the presence of any similarities are evaluated (130) between the clusters and groups and the images within the database.”). According to Stubler, the clustering (not the similarity determinations) may be based on unprocessed metadata or on processed metadata. *Id.* at ¶[0034] (“For other [sic] types of metadata, the data may already be represented in a form suitable for clustering In other cases, for example that of region metadata, the data may need to be processed for similarity by known algorithms. As an example, determination of facial Similarity is described in M. Turk and A. Pentland, “Eigenfaces for Recognition”, J. Cognitive Neurosci. 3(1) 1991 [(“Turk”)].”).³

³ I have been informed by counsel that Turk is not incorporated by reference in Stubler because Subler does not affirmatively incorporate any information from Turk, nor does it identify with detailed particularity what specific material it incorporates and clearly indicate where that material is found in the various documents. Therefore, I understand that Turk is not considered part of Stubler for

73. After the acquired images are clustered, the acquired images are compared with stored images in the database to determine whether there are any similarities between the clustered images and the images within the database. *Id.* at ¶[0031]. Although Stubler states that the database is searched based on “perceptually significant features,” Stubler does not identify any particular method for comparing the acquired images with the stored images in the database to determine whether they are similar or not. Stubler states, in relevant part:

If captioning or semantic labeling of the acquired images based on region metadata is desired (256), the database is searched (258) for any images with region metadata (e.g., face, sky, grass, skin, open space, etc.) similar to that of the acquired images, and repeated as needed (260) for similar regions until all images are analyzed.

Id. at ¶[0040]. While this passage states that the database is searched for images with “similar” region metadata, it does not explain what “similar” means, what level of similarity is required, or the particular methods of determining similarity used.

determining anticipation. Further, I have been informed that Petitioner did not include Turk in the specific grounds of unpatentability, and therefore may not rely on Turk to show that a claim feature was known.

B. Overview of Wang

74. Wang describes a “person identifier and management system.”

EX1006 at Title. The person identifier and management system includes an input/output system, an analysis system, and a profile database. *Id.* at 4:11-13. The input/output system includes a variety of input/output devices, including, for example, a display, a camera, a scanner, and a text input device. *Id.* at 5:13-26. The analysis system includes a face detection and feature extraction system to detect whether an image contains a face and to detect the eye locations of the face to compute face feature data. *Id.* at 6:35-42. The profile database stores a number of face images and associated face-identifying or person-identifying data. *Id.* at 3:37-39. When the analysis system detects a face and extracts face feature data from an image, it then uses the face feature data to search the profile database for matching face feature data. *Id.* at 5:53-60. If a matching face is found, then the system retrieves the person-identifying information data associated with the matching face feature data. *Id.* at 5:60-64. If no matching face feature data is found, the system presents the user with the option to input the new face in the profile database along with person-identifying information. *Id.* at 6:1-7.

VIII. THE TEACHINGS OF STUBLER DO NOT SATISFY ALL ELEMENTS OF ANY OF CLAIMS 1-14 OF THE '807 PATENT.

75. In my opinion, Stubler does not anticipate or render obvious all limitations of any challenged claim.

A. Stubler Fails to Disclose “an image storing section for storing the image which is supposed to be classified and retrieved together with a keyword in a database and an object of the image being previously contained in the database.”

76. I disagree with Dr. Stevenson’s opinion that Stubler discloses or renders obvious “an image storing section for storing the image which is supposed to be classified and retrieved together with a keyword in a database and an object of the image being previously contained in the database.” Ex. 1003 ¶¶ 76-82.

77. As explained in detail below, in my opinion, the claimed “object previously contained in the database” must be usable to determine whether the image to be classified includes an object that is “similar to the object of the image previously contained in the database,” as a POSITA would understand that phrase. This is true because in the later claim element, the “object information inputting section,” the claim states that certain operations are performed “when the inputted object acknowledged by the object acknowledging section is similar to the object of the image previously contained in the database.” As explained above with respect to claim construction, it is my opinion that a POSITA would understand “similar to the object of the image previously contained in the database” to mean

“sufficiently similar to identify the specific object of the image previously contained in the database.” Thus, the claimed “object previously stored in the database” must be usable to determine whether the image to be classified includes an object that is “similar to the object of the image previously contained in the database,” as a POSITA would understand that phrase.

78. It is my opinion that Stubler does not disclose such an object. Stubler discloses “region metadata” which merely consists of high-level textual descriptions of particular regions without semantic connections. These high-level textual descriptions lack sufficient information to determine whether a region in one image is sufficiently similar to region in another image to identify the *specific* objects in the region. In other words, while Stubler may be able to identify high-level, generic similarity (such as determining that two regions both generally contain images of human beings), Stubler’s region metadata lacks the necessary information to identify specific objects across images (such as determining that two images both contain the photographer’s eldest brother, “Jiro”). Because Stubler fails to disclose “an object of the image being previously contained in the database” as a POSITA would understand that term, it is my opinion that Stubler fails to anticipate or render obvious this element.

1. “An Object of the Image” Must be Usable to Determine Whether the Image to be Classified Includes an Object that is “Similar to the Object of the Image Previously Contained in the Database,” as a POSITA would Understand that Phrase.

79. Claim 1 of the '807 patent requires “an image storing section” that stores three pieces of information. EX1001 at cl. 1. First, the image storage section stores “the image which is supposed to be classified and retrieved.” *Id.* Second, the image storing section stores “a keyword in a database.” *Id.* Third, the image storing section stores “an object of the image being previously contained in the database.” *Id.*

80. According claim 1, the “object of the image being previously contained in the database” is compared to an object detected in an acquired image to determine whether the acquired image contains an object that “is similar to the object of the image previously contained in the database.” *Id.* As discussed above with respect to claim construction, it is my opinion that a POSITA would understand the phrase “similar to the object of the image previously contained in the database” to mean “sufficiently similar to identify the specific object of the image previously contained in the database.” My opinion is based on the fact that this phrase must be understood in light of the specification, which repeatedly and consistently states that the invention identifies “specific objects” or “specific persons.” EX1001 at 4:22-24, 4:51-53. My opinion is further based on the

specifications statements that the invention of the '807 patent determines whether faces “coincides with” or is “very similar” to an object (e.g., a face image) that was previously stored in the database. *Id.* at 9:18-20.

81. The object stored in the database may be an image of the object itself, or, alternatively, a representation of the object that includes sufficient information about the object to determine whether it is “sufficiently similar to identify the specific object of the image previously contained in the database.” Wakayama, cited by the '807 patent, provides an example of a representation of an object. Specifically, Wakayama teaches extracting “facial region mosaic feature data.” EX2004 at Abstract. Wakayama discloses extracting such mosaic data at a variety of granularities. For example, Wakayama teaches:

a head portion rough search step of searching for the general position of the head portion in roughly mosaicked human facial image data using a head portion rough search mosaic dictionary to obtain position information near the center of the head portion; a specific portion detailed search step of finely mosaicking a specific portion of the facial image data of a person based on the position information near the center of the head portion . . . and a facial region mosaicking step of accurately determining the position of the facial region on the facial image data based on the face-centered detailed position information, mosaicking the facial region of the facial image data using the obtained facial region position information.

EX2004 at ¶[0008]. By extracting increasingly detailed and finely-tuned mosaic data, the method of Wakayama builds a representation of a face object.

82. An example is instructive. Fig. 6 of the '807 patent (shown below) shows an exemplary database entry for a specific person: "Jiro," the second eldest brother. EX1001 at Fig. 6, 9:47-54.

FIG.6

SUPERIOR KEYWORD 1	SUBORDINATE KEYWORD 2	NAME	BIRTHDAY	IMAGE
FAMILY	SECND ELDEST BROTHER	JIRO	JANUARY 1, 2003	IMAGE 1:TAKEN ON JANUARY 1, 2003 IMAGE 2:TAKEN ON FEBRUARY 1, 2003 IMAGE 3:TAKEN ON MARCH 1, 2003

The '807 patent teaches "if a record which is shown in FIG. 6 is added such that images which relate to the 'second eldest brother' are contained, it is necessary to produce a keyword for the 'second eldest brother' as a subordinate keyword under the superior keyword 'family'." *Id.* at 9:58-62. In order to determine that an acquired image contains, specifically, the second eldest brother, the image retrieving device must be able to identify, with specificity, the presence of that specific person.

83. On the other hand, if the objects in the database are merely high level, generic descriptions of certain objects in an image, the image retrieving device would be unable to make such a specific determination. Stated differently,

information that simply describes objects at a high level or generically are not “objects” within the meaning of the claim. This is because high-level or generic information cannot be used to determine whether an object in an acquired image is “sufficiently similar to identify the specific object of the image previously contained in the database.” High-level, generic descriptions of objects merely indicate a general category of an object, such as “a person,” “a car,” “grass,” or “sky.” While these high-level descriptions may be useful for *broadly* categorizing images, they would not be themselves sufficient to identify *specific* objects (in other words, far more information and analysis are needed to identify a specific object). This is because there is substantial variation of objects within these broad categories. For example, there are billions of unique people, thousands of unique car models, thousands of types of grass, and a variety of atmospheric phenomena that can all be broadly categorized as “sky.”

84. Given the clear and consistent description of objects in the specification of the '807 patent, “an object of the image” must be usable to determine whether an object in an acquired image is “similar to the object of the image previously contained in the database,” as a POSITA would understand that phrase, namely, “sufficiently similar to identify the specific object of the image previously contained in the database.”

2. Stubler Discloses “Region Metadata,” which Merely Consists of High-Level Descriptions of Particular Regions.

85. Petitioner and Dr. Stevenson rely exclusively on Stubler’s disclosure of “region metadata”⁴ as disclosing “an object of the image being previously contained in the database.” Pet. at 20; EX1003 at ¶81.

86. I disagree with Petitioner and Dr. Stevenson. According to Stubler, region metadata includes regional features such as face, skin, sky, grass, open space, and background vs. foreground in a particular image region. EX1005 at ¶[0030]. Stubler, however, makes clear it is using the term “image region” specifically to refer to “significant characteristic low-level feature[s], such as ‘sky’ or ‘grass’.” *Id.* at ¶[0022]. Moreover, Stubler explains that regional processing

⁴ Stubler generally discusses two other types of metadata—capture metadata and global metadata. Neither the Petition nor Dr. Stevenson point to either of these types of metadata as allegedly disclosing the claimed “object previously contained in the database” that is stored along with the image to be classified and the keyword. To the extent Petitioner or Dr. Stevenson contend that capture metadata or global metadata disclose the claimed “object previously contained in the database,” I reserve the right to supplement my opinions to address such contentions.

yields “high-level feature description” and not the high-level features themselves (or a representation thereof). *Id.* In other words, Stubler’s region metadata is simply a textual description of a feature in a particular area of the acquired image. The region metadata is not the object of the image itself or a representation thereof.

87. This understanding (that region metadata consists merely of high-level textual descriptions) is consistent with the examples that Stubler provides for determining region metadata detection. Stubler states that its high-level region metadata is “derived from lower level data (e.g., flesh color, sky color and distribution, open space).” EX1005 at ¶[0028]. Stubler provides several examples for determining this region metadata, which are discussed below.

88. First, Stubler identifies Sung 1998 (EX1007) as providing an example of “face detection.” Sung explains that “face *detection* determines the location and size of each human face (if any) in an input image.” EX1007 at 1 (emphasis in original). Thus, face detection as described in Sung simply generates the location and size of an object (in this case, a face), but it is not the object itself. Sung contrasts face “detection” with a different, albeit “closely related” problem of facial recognition. *Id.* Sung is not concerned with facial recognition; it deals exclusively with facial detection. *Id.* (“We present an example-based learning approach for locating vertical frontal views of human faces in complex scenes.”). In practical terms, if the method of Sung is employed to detect the presence of a

face, the output of that determination is simply a binary “yes” or “no” as to whether a face is present at a particular location in an image. *Id.* at 2 (“the most critical part of our system is the algorithm for classifying window patters as ‘faces’ or ‘nonfaces.’”). At most, Sung is similar to sections 112/113 of Fig. 2 in the ’807 patent, in which the ’807 patent discusses detecting a face image or a face color area. Sung does not disclose an output of a specific face or representation of a face that could then be used to determine similarity with a specific face.

89. Second, Stubler points to Yang 1996 (EX1009) as an example of representing skin color in a chromatic space. Again, the resulting data generated in Yang is not the object itself, but a “chromatic color space” comprising normalized color information (*e.g.*, traditional R, G, and B values) in the absence of brightness data. EX1009 at 2. This data can then be used to detect or locate, generally, the presence of a face in the original image by comparing the chromatic color space information for each pixel to a skin color model. *Id.* at 4. Further, Yang discusses a transformation model over multiple images in a video to track the motion of a face. *Id.* at 5. At no point is the detected face or a representation thereof in Yang ever stored as its own object. Instead, the metadata is the *location* of the face in a series of images. *Id.* To the extent Yang discusses identifying the “same face” across images, it does so using “a motion model [that] is used to estimate image motion and to predict search window.” *Id.* at 1. Yang does not compare objects in multiple

images to determine whether they are similar. Yang, like Sung is more similar to sections 112/113 of Fig. 2 in the '807 patent, in which the '807 patent discusses detecting a face image or a face color area. Accordingly, Yang simply generates color information that can be used to detect the presence or location of a face and track motion of that face across images, but Yang does not disclose saving the specific object (*e.g.*, the face or a representation thereof) itself.

90. Third, Stubler points to U.S. Pat. App. No. 09/450,190 (“Luo”) (EX1008) as an example of detecting “semantic regions” such as the sky. Luo, like Sung and Yang, is concerned only with determining the location of objects. EX1008 at 3:58-60 (“It is an object of this invention to provide a method for detecting the location of main subjects within a digitally captured image . . .”). Luo is silent as to saving the object (or “main subject” in the terminology of Luo) as recited in claim 1 of the '807 patent. Thus, Luo also only discloses determining the location of a subject, and not saving the subject itself.

91. Therefore, it is my opinion that Stubler’s “region metadata” is merely high-level, textual descriptions of generic features within a region of an image. Such high-level textual descriptions are based on certain low-level features that are concerned with determining the location of certain objects.

3. Stubler’s “Region metadata” Is not Usable to Determine Whether the Image to be Classified Includes an Object that is “Similar to the Object of the Image Previously Contained in the Database,” as a POSITA would Understand that Phrase.

92. It is my opinion that Stubler’s region metadata is not an “object of the image” because region metadata is merely a high-level, textual description of the basic contents of a region of an image and is not usable to determine similarity between an image to be classified and an object previously contained in the database. Specifically, Stubler’s region metadata lacks sufficient information to determine whether a newly acquired image is “sufficiently similar to identify the specific object of the image previously contained in the database.” Each of the categories of “region metadata” discussed in Stubler and cited by Petitioner lacks sufficient information to identify a *specific* object in a newly acquired image.

93. As discussed above, Stubler identifies region metadata as containing “high-level feature description (clear sky, cloudy sky, mixed sky) and low-level features (color representation, area coverage, location).” EX1005 at ¶[0011]. Any “[s]imilarities between such regions would be calculated on the basis of a combination of high-level and low-level features.” *Id.* Neither high-level textual descriptions nor low-level features are sufficient to identify a specific object across images, as explained below.

94. First, high-level textual descriptions cannot be used to identify a specific object in an acquired image. As the examples in Stubler make clear, high level semantic descriptions are merely textual descriptions of generic categories of objections. Stubler at ¶[0028]. For example, Stubler identifies “face” as one possible high-level semantic description of a region. In other words, the region metadata indicates that an object in the region is *a* face. But Stubler is completely silent as to how a semantic indication that a given region includes *a* face would include sufficient information to compare that semantic description to an acquired image to determine whether the acquired image contains the *specific* face. Nor could it. At best, the only similarity that the high-level semantic description “face” could identify is that both the image with the region metadata “face” and an acquired image contain *a* face. This is insufficient because, as discussed above, the claimed “object” must include information to determine whether the image to be classified is “sufficiently similar to identify the specific object of the image previously contained in the database.”

95. The same is true of the other high-level semantic descriptions identified in Stubler. Specifically, general textual descriptions that state that a given region contains a “sky” or “grass” lack sufficient information to determine whether a new image to be classified contains a *specific* sky or *specific* grass.

96. Second, Stubler fails to explain how “low-level” features, such as those discussed above with respect to Sung, Yang, and Luo, can be used to determine whether a newly acquired image is “sufficiently similar to identify the specific object of the image previously contained in the database.” As discussed above, similar to the first step of the two-step process discussed in the ‘807 patent each of these low-level features analyzes a region of an image to make a binary determination of whether the region contains a particular high-level feature. While these references take into account information, such as skin color, which may be implicated in a process of identifying a specific face, neither Stubler nor the incorporated references disclose doing so to identify a particular instance of an object in a class. While Stubler includes a generic statement that calculation of similarities is “based on” (EX1005 at ¶[0028]) a combination of high-level and low-level features, Stubler provides no additional description explaining that these similarities are used to identify a *specific* object. Nor could it, the low and high-level features are determined in such vague detail that similarity could only be determined in broad categories and not to the granularity of specific instances of an object class.

97. Stubler’s disclosure with respect to Figures 3 and 5 demonstrate the shortcomings of Stubler. With respect to Figure 3, step 204, Stubler states “After the image representation is built, a search of the database is performed, and similar

images are retrieved (204) from the database, and a set of caption or semantic label options are prepared. The retrieval and preparation stage 204 will be further explained in relation to FIG. 5.” EX1005 at ¶[0037]. With respect to Figure 5, Stubler discloses:

If captioning or semantic labeling of the acquired images based on region metadata is desired (256), the database is searched (258) for any images with region metadata (e.g., face, sky, grass, skin, open space, etc.) similar to that of the acquired images, and repeated as needed (260) for similar regions until all images are analyzed.

Id. at ¶[0040]. That is the extent of Stubler’s explanation of how similar images are determined based on region metadata. There is no explanation as to how similar images are located, how similarity is calculated or determined, or what even constitutes “similarity.”

98. Therefore, it is my opinion that Stubler fails to disclose that “region metadata” is usable to determine whether the image to be classified includes an object that is “similar to the object of the image previously contained in the database,” as a POSITA would understand that phrase. Namely, Stubler fails to disclose that “region metadata” can be used to determine whether an acquired image is “sufficiently similar to identify the specific object of the image previously contained in the database.”

99. Claims 1, 7, and 8 each recite this limitation, and claims 2-6 depend from claim 1. Claims 9-14 depend from claim 8. Thus, my analysis applies to all of claims 1-14.

B. Stubler fails to disclose “an object information inputting section for confirming by the user, adding, and correcting the keyword which is proposed by the keyword proposing section when the inputted object acknowledged by the object acknowledging section is similar to the object of the image previously contained in the database.”

100. It is further my opinion that Stubler does not anticipate or render obvious the claimed “object information inputting section” because Stubler does not disclose a way to determine that “the inputted object acknowledged by the object acknowledging section is similar to the object of the image previously contained in the database,” as a POSITA would understand that phrase. Specifically, Stubler describes only storing region metadata, which is a high-level description or location of an object, but not the object itself or a representation thereof. Claim 1 recites that the object inputting section is for confirming, adding, or correcting a keyword “when the inputted object acknowledged by the object acknowledging section is similar to the object of the image previously contained in the database.” Thus, claim 1 determines whether the inputted object and the previously stored object are the same by comparing them. Under Stubler’s “region metadata,” this is not possible because there is no previously stored object with

which to compare an inputted object. Accordingly, Petitioner fails to demonstrate that Stubler anticipates or renders obvious the claimed “object information inputting section for confirming by the user, adding, and correcting the keyword which is proposed by the keyword proposing section when the inputted object acknowledged by the object acknowledging section is similar to the object of the image previously contained in the database.”

IX. THE TEACHINGS OF STUBLER IN COMBINATION WITH WANG DO NOT SATISFY ALL ELEMENTS OF ANY OF CLAIMS 2-5 OR 9-12 OF THE '807 PATENT.

101. In Ground 2, Petitioner and Dr. Stevenson assert that Stubler in view of Wang renders dependent claims 2-6 and 9-14 obvious. As discussed above, dependent claims 2-6 each depend from claim 1 and dependent claims 9-14 each depend from claim 8. Therefore, I understand that claims 2-6 include all of the limitations of claim 1 and claims 9-14 include all of the limitations of claim 8. Because Petitioner does not challenge independent claims 1 and 8 under Ground 2, Petitioner fails to demonstrate that claims 2-6 and 9-14 are unpatentable for the same reasons as claims 1 and 8.

X. ADDITIONAL REMARKS

102. I reserve any right that I may have to supplement this declaration if further information becomes available or if I am asked to consider additional information or to submit an additional declaration. Furthermore, I reserve any right

that I may have to consider and comment on any additional expert statements and testimony of Petitioner's expert in this matter.

103. I hereby 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. I further declare 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 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patents issued thereon.

A handwritten signature in black ink, appearing to be 'N. Papanikolopoulos', written over a light gray rectangular background.

Dated: April 21, 2025

Nikolaos Papanikolopoulos, Ph.D.