

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

TVISION INSIGHTS, INC.,

Petitioner

v.

THE NIELSEN COMPANY (US), LLC,

Patent Owner

Inter Partes Review Case No. IPR2023-01014

U.S. Patent No. 11,470,243

DECLARATION OF DR. ELI SABER

TABLE OF CONTENTS

I. INTRODUCTION	1
A. BACKGROUND AND QUALIFICATIONS.....	2
B. MATERIALS REVIEWED	6
II. LEGAL FRAMEWORK.....	9
A. OBVIOUSNESS	10
III. LEVEL OF A PERSON HAVING ORDINARY SKILL IN THE ART..	11
IV. THE CLAIMED INVENTION OF THE '243 PATENT	12
V. OPINIONS ON CHALLENGED GROUND 1	16
A. OPINIONS REGARDING <i>TIAN</i> 'S RESOLUTION REDUCING FOR HEAD POSE ESTIMATION.....	16
B. OPINIONS REGARDING <i>LU-000</i> 'S IMAGE RECOGNITION SYSTEM	18
C. OPINIONS REGARDING <i>TIAN</i> 'S ALLEGED TEACHING OF "REDUCING A RESOLUTION"	22
1. <i>Tian Does Not Teach "Reducing a Resolution"</i>	23
D. OPINIONS REGARDING THE MOTIVATION TO COMBINE <i>TIAN</i> 'S HEAD POSE ESTIMATION WITH <i>LU-577</i>	26
1. <i>Petitioner's Proposed Modification</i>	27
2. <i>Modifying Lu-577 with Tian's Head Pose Estimation Does Not Achieve the Proposed Benefit of Improved Facial Recognition</i>	28
3. <i>Dr. Doermann's Cited Background References Do Not Support His Opinions</i>	44
4. <i>Opinions Regarding Claims 4, 11, and 18: Petitioner Provides No Credible Motivation for Analyzing Reduced-Resolution Images and Tian Teaches That Doing So is Undesirable</i>	58
5. <i>Opinions Regarding Petitioner's General Allegations of Obviousness</i> ..	73
VI. OPINIONS ON CHALLENGED GROUND 2	76
A. OVERVIEW OF THE CITED REFERENCES AND PETITIONER'S PROPOSED COMBINATION	76
1. <i>Nielsen-372's Local Metering System</i>	76
2. <i>Steinberg's Digital Image Processing System</i>	80
3. <i>Petitioner's Proposed Modification of Nielsen-372 With Steinberg</i>	81

B. OPINIONS REGARDING PETITIONER’S MODIFICATION OF <i>NIELSEN-372 WITH STEINBERG</i>	82
1. <i>Petitioner’s Modification Would Frustrate the Purpose of Nielsen-372’s Metering System</i>	83
2. <i>Petitioner’s Proposed Modification Would Change the Principle of Operations of Nielsen-372’s System</i>	86
3. <i>The Patent Office Found Similar Modifications To Be Non-obvious</i>	90
VII. OPINIONS ON CHALLENGED GROUND 3	92
A. OPINIONS REGARDING WHETHER THE COMBINATION OF <i>NIELSEN-372, STEINBERG, AND TIAN</i> RENDER OBVIOUS CLAIMS 4- 6, 11-14, AND 18-20	92
VIII. OPINIONS REGARDING WHETHER TIAN AND STEINBERG ARE IN THE SAME FIELD OF ENDEAVOR OR SOLVE THE SAME PROBLEM AS THE ’243 PATENT	92
IX. CONCLUSION	97

I, Eli Saber, declare as follows:

I. INTRODUCTION

1. My name is Dr. Eli Saber, and I am over 21 years of age and otherwise competent to make this Declaration. I make this Declaration based on facts and matters within my own knowledge and on information provided to me by others, and, if called as a witness, I could and would competently testify to matters set forth herein.

2. I have been retained as a technical expert witness in this matter by Counsel for the Patent Owner, the Nielsen Company (US), LLC, (“Patent Owner”) to provide my independent opinions on certain issues requested by Counsel for Patent Owner relating to the accompanying Patent Owner Preliminary Response to the Petition for *Inter Partes* Review of U.S. Patent No. 11,470,243. I am being compensated in this matter at my customary rate of \$500/hour. My compensation in this matter is not based on the substance of my opinions or on the outcome of this matter. I have been informed that TVision Insights, Inc., is the Petitioner in this matter. I note that I have no financial interest in TVision Insights, Inc., or in Patent Owner, and I have no other interest in the outcome of this matter.

3. I have summarized in this section my background and qualifications. I have also included a current version of my curriculum vitae as Exhibit 2008.

A. Background and Qualifications

4. I am a Professor in the Department of Electrical and Microelectronic Engineering (EME) and the Chester F. Carlson Center for Imaging Science at the Rochester Institute of Technology (RIT) in Rochester, New York. I also serve as the Director of the Image, Video and Computer Vision Laboratory. I joined RIT's full-time faculty in 2004. I have over 33 years of industry and academic experience, 28 of which are in the field of imaging.

5. I received a Ph.D. in Electrical Engineering from the University of Rochester in 1996. My concentration was in Signal/Image/Video Processing, Pattern Recognition, and Computer Vision. I received a Master of Science Degree in Electrical Engineering from the University of Rochester in 1992 in the same concentration as listed above, and I received a Bachelor of Science Degree in Electrical and Computer Engineering, Summa Cum Laude, from the State University of New York at Buffalo in 1988.

6. Before becoming a full-time professor, I worked for Xerox Corporation from 1988 to 2004. During my 16 years at Xerox, I was – generally speaking – responsible for: a) delivering color management solutions, image processing innovations, architectures, and algorithms for high-end color printing systems, b) xerographic sub-systems for a variety of color products and c) control systems for toner production facilities. Among other roles, as an Advanced Development

Scientist and Manager, I established the Advanced Design Laboratory — an Imaging/Xerographics lab — and provided technical and managerial leadership for the Electrical, Imaging and Xerographics Department. As a Product Development Scientist and Manager, I led the research and development of color management algorithms and image quality metrics for various product platforms. I also led the Image Science, Analysis, and Evaluation area, with twelve to fifteen direct reports and a budget of approximately \$2 million.

7. From 1997 until 2004, I was an adjunct faculty member at the Electrical Engineering Department of the Rochester Institute of Technology and at the Electrical & Computer Engineering Department of the University of Rochester. I was responsible for teaching undergraduate and graduate coursework in signal, image, video processing, pattern recognition, and communications and performing research in multimedia applications, pattern recognition, image understanding and color engineering.

8. Since joining RIT full time, I have been responsible for teaching several undergraduate and graduate courses in Digital Signal Processing, Digital Image Processing, Digital Video Processing, Engineering Analysis, Random Signal & Noise, Advanced Engineering Mathematics, Matrix Methods, Pattern Recognition, Communications, Modern Control Theory, and Linear Systems.

9. I proposed, developed and introduced the Digital Video Processing and Pattern Recognition courses to the Electrical Engineering Department at RIT. These graduate courses were instrumental in providing MS/PhD students with the proper foundation and knowledge in all aspects of the subject areas in order to perform in-depth research in the field and continue to advance the state of the art.

10. The Digital Video Processing course is designed to provide a thorough and comprehensive coverage of the field. Topics covered include the mathematical representation of digital video including motion and time-varying image formation models, spatio-temporal sampling/reconstruction, sampling structure conversions, motion estimation using optical flow, block-based, pel-recursive, Bayesian and point correspondence type methodologies, segmentation, motion tracking, filtering, restoration, super-resolution and compression.

11. The Pattern Recognition course, on the other hand, covers pertinent concepts in Bayes decision theory, linear and nonlinear classifiers, artificial neural networks, feature selection/generation/extraction and clustering.

12. I am currently working to develop a graduate course titled: "Advanced Neural Networks and Deep Learning: Theory, Mathematics and Algorithms". The course is intended to provide an in-depth comprehensive study of the theory, mathematics and algorithms for advanced neural networks and their usage in deep learning. It is primarily targeted for upper graduate students that are planning to

pursue advanced studies/research in the field. The course provides a thorough study of shallow and deep neural networks architectures and algorithms, activation functions, backpropagation, initialization, optimization strategies, etc. It then progresses to convolutional and recurrent style neural networks and, if time permits, will provide an overview of attention mechanisms/transformers and deep reinforcement learning. The course is set to debut in Spring 2024.

13. My current research focuses on image/video processing for multimedia, military & biomedical applications; computer vision and three-dimensional scene reconstruction; and machine learning for object tracking/remote sensing applications. As a principal investigator (PI) or co-principal investigator, I have acquired research funding in excess of \$5 million since joining RIT and have managed multiple government grants from the Department of Defense as well as several corporate grants from Hewlett-Packard, Lenel, and Data Physics. I am currently managing (as PI) a multiyear government grant from the Department of Defense title “Target Detection/Tracking and Activity Recognition from Multimodal Data.” In 2012, I was awarded the Prestigious Trustees Scholarship, the highest award at RIT with regard to research recognition.

14. I am the author or co-author of 38 peer-reviewed journal publications.

15. I have also authored or co-authored 102 conference and workshop publications and a book entitled Advanced Linear Algebra for Engineers with

MATLAB, published by CRC Press in February 2009, and am a named inventor on multiple U.S. and foreign patents.

16. I am a senior member of the Institute of Electrical and Electronic Engineers and a member of the IEEE Signal Processing Society, the Electrical Engineering Honor Society, and Eta Kappa Nu.

17. In my consulting career, I have provided both deposition testimony and trial testimony.

B. Materials Reviewed

18. In reaching my opinions in this matter, I relied on the following materials:

- Petition for *Inter Partes* Review of U.S. Patent No. 11,470,243 (“Petition”);
- U.S. Patent No. 11,470,243 (“the ’243 Patent”) (Ex. 1001);
- File History for U.S. Patent No. 11,470,243 (“’243 File History”) (Ex. 1002);
- Declaration of Dr. David Scott Doermann (Ex. 1003);
- U.S. Patent Application Publication No. 2002/0059577 (“Lu-577”) (Ex. 1004);

- *Evaluation of Face Resolution for Expression Analysis*, Ying-li Tian, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2004 (“Tian”) (Ex. 1005);
- U.S. Patent Application Publication No. 2010/0274372 (“Nielsen-372”) (Ex. 1006);
- U.S. Patent Application Publication No. 2010/0066822 (“Steinberg”) (Ex. 1007);
- U.S. Patent No. 4,858,000 (“Lu-000”) (Ex. 1008);
- U.S. Patent Application Publication No. 2011/0081052 (“Bigioi”) (Ex. 1010);
- U.S. Patent Application Publication No. 2011/0265110 (“Weinblatt”) (Ex. 1012);
- *Robust Face Detection at Video Frame Rate Based on Edge Orientation Features*, Bernhard Fröba and Christian Küblbeck, Proceedings of the 5th IEEE Conference on Automatic Face and Gesture Recognition (Ex. 1013);
- U.S. Patent No. 7,315,631 (“Corcoran”) (Ex. 1014);
- European Patent Application No. 0932114A2 (“Hong”) (Ex. 1015);

- *Watching Americans Watch TV*, Erik Larson, the Atlantic Monthly, March 1992 (Ex. 1016);
- *Head Pose Estimation in Computer Vision: A Survey*, Erik Murphy-Chutorian and Mohan Manubhai Trivedi, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, No. 4, April 2009 (Ex. 1020);
- *Face Recognition: A Literature Review*, Zhao et al., ACM Computing Surveys, Vol. 35, No. 4, December 2003 (Ex. 1021);
- *Evaluation Report*, Blackburn et al., DoD Counterdrug Technology Development Program Office, February 2001 (Ex. 1024);
- *Face Image Resolution Versus Face Recognition Performance Based on Two Global Methods*, Wang et al., 2004 (Ex. 1027);
- *Facial Expression Recognition: A Brief Tutorial Overview*, Claude Chibelushi and Fabrice Bourel, Staffordshire University, 2002 (Ex. 2009);
- *Facial Expression Recognition*, Maja Pantic, Springer-Verlag Berlin Heidelberg, 2009 (Ex. 2010);

- *Automatic Analysis of Facial Expressions: the State of the Art*, Maja Pantic and Leon Rothkrantz, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, No. 12, December 2000 (Ex. 2011);
- *Mechanisms of Face Perception*, Doris Y. Tsao and Margaret S. Livingstone, Annu Rev Neurosci. 2008, 31, 411-437 (Ex. 2012); and
- Any other materials I cite in support of this Declaration.

II. LEGAL FRAMEWORK

19. I am a technical expert and do not offer any legal opinions. However, counsel has informed me that in proceedings before the USPTO the claims of an unexpired patent are to be given their broadest reasonable interpretation in view of the specification from the perspective of one skilled in the art. The broadest reasonable interpretation does not mean the broadest possible interpretation. Rather, the meaning given to a claim term must be consistent with the ordinary and customary meaning of the term (unless the term has been given a special definition in the specification), and must be consistent with the use of the claim term in the specification and drawings. Further, the broadest reasonable interpretation of the claims must be consistent with the interpretation that those skilled in the art would reach.

A. Obviousness

20. I understand that, in this *inter partes* review, the Petitioner has argued that certain claims of the '243 Patent are obvious. Counsel explained that a patent claim is invalid as obvious where the differences between the subject matter claimed and what was known in the prior art would have been obvious to a person of ordinary skill in the relevant field at the time the invention was made.

21. Counsel explained that an analysis of obviousness requires consideration of four factual inquiries. Counsel explained that these inquiries are: (1) the scope and content of the prior art; (2) the differences between the prior art and the claims; (3) the level of ordinary skill in the relevant field; and (4) whether what counsel described as “secondary considerations” or “objective indicia” of non-obviousness are shown by the evidence.

22. Counsel explained that a person of ordinary skill in the relevant art is not a robot (or "automaton"), but instead a person with ordinary creativity and common sense who would apply that creativity and common sense, along with relevant background knowledge, to the problems at hand.

23. I understand that a patent may be invalid as obvious in view of a single prior art reference or multiple prior art references together, and that the background knowledge of, and common-sense inferences that would be made by, a person of ordinary skill in the art may be considered in determining whether a claim is obvious.

24. Counsel informed me that there need not be a specific teaching, motivation, or suggestion in the prior art to combine or modify prior art references in a particular way. For instance, the background knowledge or common sense of a person of skill could make a claim of an obvious variation on what was already known. However, there must be an articulable reason as to why a particular combination would be made or a specific reference modified.

25. I have also been informed that it is improper, in an obviousness analysis, to apply hindsight, and in particular to use the patent claims as a roadmap to assemble a collection of references or modifications to references. I understand that the proper question is what would have been obvious to a person of ordinary skill in the art without the benefit of hindsight and specifically without access to the patent at issue.

III. LEVEL OF A PERSON HAVING ORDINARY SKILL IN THE ART

26. Petitioner proposes that a person having ordinary skill in the art at the time of the '243 Patent would have been:

a person with a bachelor's degree in computer science, electrical engineering, or a similar field with at least two years of experience in media analysis or a person with a master's degree in computer science, electrical engineering, or a similar field with a specialization in media analysis.

Petition, 7.

27. I do not dispute Petitioner's proposed POSITA definition.

28. As of the earliest priority date of the '243 Patent, which counsel has informed me is December 15, 2011, I was, at the least, a POSITA applying Petitioner's definition. As set forth above in the discussion of my qualifications, I had a Ph.D., master's degree, and bachelor's degree in electrical engineering. My Ph.D. focused on Signal/Image/Video Processing, Pattern Recognition, and Computer Vision. Thus, I specialized in media analysis.

29. Additionally, I have over 25 years of work experience related to media analysis, including performing image and video processing research for multimedia applications and performing research in pattern recognition and machine learning. I have authored several journal publications related to implementing pattern recognition for detecting faces and/or relevant content in images prior to December 15, 2011. I also had at least two years of work experience in developing face detection and facial feature extraction approaches.

IV. THE CLAIMED INVENTION OF THE '243 PATENT

30. I have been asked to provide my opinions on the invention disclosed and claimed in the '243 Patent. The '243 Patent discloses an audience measurement system. *'243 Patent*, Ex. 1001, Abstract, 1:37-39, 2:26-29. The audience measurement system recognizes the identities of people watching TV by performing facial recognition on images of audience members. *'243 Patent*, 2:42-45, 5:14-19.

High-resolution images are captured in the TV environment by a high-resolution image sensor to perform facial recognition. *'243 Patent*, 2:50-56, 5:39-43, 8:55-60.

31. Cameras were known to implement an illumination source within audience measurement systems when capturing high-resolution images. *'243 Patent*, 2:57-61. Frequently activating the illumination source to capture high-resolution images caused several disadvantages (i.e., resource consumption, audience annoyance). *'243 Patent*, 3:4-30. To address these issues, the audience measurement system of the *'243 Patent* selectively activates the illumination source only for frames that “require high-resolution image data,” such as for facial recognition processes. *'243 Patent*, 3:31-37. Frames not requiring high resolution include frames used for only counting audience members. *'243 Patent*, 3:41-44.

32. I note that Dr. Doermann, Petitioner’s declarant, opines the “selective activation of illumination” is the “purported new approach” of the *'243 Patent*. *Declaration of Dr. Doermann*, Ex. 1003, ¶ 20. In my opinion, the claimed system in the *'243 Patent* improves audience measurement systems by providing a way to determine head orientation to designate captured images for facial recognition. For example, in claim 4, head orientation determination is performed based on the reduced resolution image, with the result being that if the head orientation is suitable for face recognition (e.g., because the person in the image is facing the camera), then activation of the illumination and capture of a high resolution image for facial

recognition can be performed. *'243 Patent*, 9:4-22, 9:26-31, 9:54-59, 10:29-38, 10:67-11:24, 11:34-36. Based on the results of the head orientation, the system determines if frames are to undergo facial recognition. If so, the illumination source is activated. *'243 Patent*, 10:1-7. If not, illumination is not activated to avoid unnecessary light that may “annoy[]” the audience. *'243 Patent*, 3:4-30, 9:40-47. The audience measurement system reduces the resolution of images captured without illumination activated so the head orientation within the image may be accurately determined and used to determine if succeeding frames are to undergo facial recognition. *'243 Patent*, 10:67-11:8, 11:19-23, 11:34-36;10:31-38.

33. The *'243 Patent* thus discloses and claims (in claims 4, 11, and 18) determining the orientation of the heads in the captured images based on a reduced-resolution image. The step of determining the head orientation based on the reduced-resolution image then allows for analyzing images captured in low light to determine when to activate the illumination source for images that will undergo facial recognition. All of the challenged dependent claims include a limitation reciting “determin[ing] the orientation of the head with respect to the camera.” *'243 Patent*,

claims 4, 11, 18¹. Contrary to Dr. Doermann's opinion, because the "new approach" (per Petitioner) for selectively activating illumination relies on determining the head orientations, the Challenged Claims do indeed recite a combination of features not shown in the prior art. Ex. 1003, ¶¶ 16-20. Namely, the determination of the head orientations provides an improvement to the audience measurement system, i.e., improving the analysis of images captured in low light by performing head orientation determination on reduced-resolution images and determining the head orientation in order to perform the selective illumination of images designated to undergo facial recognition.

34. For the reasons above and provided herein, it is my opinion that the Challenged Claims recite a combination of novel and non-obvious features. Ex.

¹I note the remaining challenged dependent claims incorporate the limitations from either claim 4 (claims 5-6), claim 11 (claims 12-14), or claim 18 (claims 19-20). '243 *Patent*, col. 15-col. 17. Thus, all dependent claims include head orientation determination therein. Furthermore, the dependent claims depend from independent claims that recite determining "an orientation of the head," and therefore incorporate head orientation determination via the independent claim limitation. '243 *Patent*, 15:36-37, 16:35-36, 17:30-31.

1003, ¶ 20. Namely, the Challenged Claims recite determining a head orientation and matching the head to a known person to “determine audience identification information,” with the head orientation determination being based on a reduced-resolution image (see, e.g., claim 4).

V. OPINIONS ON CHALLENGED GROUND 1

A. OPINIONS REGARDING *TIAN*’S RESOLUTION REDUCING FOR HEAD POSE ESTIMATION

35. Petitioner maps *Tian* as allegedly teaching the claimed “reduc[ing] a resolution of a first image of the one or more of the images of the media exposure environment to obtain a reduced-resolution image” and “determin[ing] the orientation of the head with respect to the camera based on the reduced-resolution image.” *Petition*, 26-27 (citing § VI.A.2[1e] at p. 22-23). I have been asked to provide my opinions on whether a POSITA would have been motivated and found it obvious to combine *Tian* with *Lu-577* (as modified to include *Lu-000*’s recognition system) for teaching the above limitations. Before providing such opinions, I initially provide a discussion of a POSITA’s understanding of *Tian*.

36. *Tian* is concerned with “automatic **facial expression analysis** (AFEAs) systems,” and, more specifically, the “effects” of different image resolutions on various steps of “**facial expression analysis**” performed by AFEAs. *Tian*, Ex. 1005, Abstract. *Tian* is interested in the effects of various image resolutions to determine

how AFEAs perform in “**real environments**” where “**the face image is often in lower resolution and with head motion.**” *Id.* To determine the effects of low resolution images from a “real environment” on facial expression analysis, *Tian* employs a “head pose estimation” algorithm as one technique for “face acquisition.” *Tian*, 000005. To perform the “head pose estimation” algorithm, *Tian* detects the head in order to handle the “full range of head motion.” *Tian*, 000002. Once the head is detected, the head pose estimation algorithm estimates the “head pose” by placing the estimated head poses within three categories: “1) frontal or near frontal view, 2) side view or profile, 3) others such as back of the head or occluded face.” *Tian*, 000002. Recognition of **facial expressions** is only performed on “frontal and near frontal view faces.” *Id.* Thus, *Tian* categorizes the head poses among the three categories to filter out images that are not suitable for facial expression analysis. *Tian* provides a Table to illustrate exemplary images and their respective categorization in the three classes via the head pose estimation algorithm:

Examples		 	 
	 		

Tian, Table 2 (left column depicting “frontal or near frontal” pose class, middle column depicting “side view or profile” pose class, and right column depicting “others” pose class).

B. OPINIONS REGARDING *LU-000*’S IMAGE RECOGNITION SYSTEM

37. As a preliminary matter, I note *Lu-000* is “incorporated by reference” into *Lu-577*. *Lu-577*, Ex. 1004, [0039] (describing *Lu-577*’s audience measurement system “may employ known head location and face recognition software (e.g., as taught by Lu in U.S. Pat. No. 4,858,000)), [0013] (disclosing the teaching of *Lu-000* being “incorporated by reference”). Petitioner notes this incorporation by reference. *Petition*, 4. I have been asked to provide my opinion to the combination of *Lu-577* and *Tian*. I have been instructed by counsel that Petitioner’s combination of *Lu-577* with *Tian* includes combining “Lu-000 and Tian.” *Petition*, 8; *see also*, ¶¶ 53-54. Thus, to provide my opinions on the combination, I analyzed *Lu-000*’s teachings, as discussed below.

38. The *Petition* maps *Lu-000* for several claim limitations, including “detect[ing] a head appearing” in one or more images and determining an audience member’s identity by matching the head to someone “associated with the media exposure environment.” *Petition*, 19-21, 23-25. Petitioner states *Lu-000* “captures and analyzes images to locate the heads” of audience members and “extracts and

compares pattern image signatures with stored signatures to identify the audience members.” *Petition*, 8; Ex. 1003, ¶ 30.

39. *Lu-000* discloses an image recognition system that comprises an audience scanner containing a video camera. *Lu-000*, Ex. 1008, Abstract, 1:14-19, 3:27-30, 3:61-63. The video camera captures video images of the audience within a “monitored area.” *Lu-000*, 3:36-41, 3:61-4:1. The audience scanner also comprises infrared sensors 30 that produce infrared signals within the monitored area. *Lu-000*, 3:63-64, 4:9-19. Digital representations are generated from the infrared signals “to identify directions within the monitored area corresponding to the possible locations of individual audience members.” *Lu-000*, 3:63-64, 4:64-5:2.

40. Initially, the recognition system operates in a “learning operational mode.” *Lu-000*, 6:27-29. In the learning mode, the digital representations are applied to a “face image location block,” normalized, and applied to a feature signature extraction block 98 that extracts “pattern image signature[s]” from the digital representations. *Lu-000*, 6:32-50. The extracted pattern image signatures are stored in an “individual face storage library (IFL).” *Lu-000*, 6:50-51. During the learning mode, “multiple face images are learned for each of the audience members” by processing a “series of video images,” amounting to multiple signatures being stored in the IFL for each audience member. *Lu-000*, 7:31-38.

41. *Lu-000* also discloses “An individual audience member face image may be learned multiple times,” generating various “different extracted signatures, **“depending on any changed facial expressions or various profiles of the audience member”².**” *Lu-000*, 8:36-45. In my opinion, a POSITA would have understood *Lu-000*’s disclosure of multiple signatures being extracted for a single audience member “depending” on various “profiles” indicates the multiple stored signatures are for multiple different head orientations. The term “profile”—when used in the context of facial recognition—refers to head orientations where the head is angled to the side. For example, profiles encompass various orientations of the head, including “quarter profiles,” “half profiles,” and “full profiles” (e.g., when head is angled 90 degrees to right or left). *Face Recognition: A Literature Review*, Ex. 1021, 000036 at col. 2, par. 2, 000041 at col. 1, par. 2, 000052 at col. 2, par. 1 (describing a “full profile” view of the head at “90°” and a “quasiprofile” view of the head being at “45°”). Thus, a POSITA would have understood the various profile views of the person’s head are stored as signatures, based on the orientation of the head within the image. Additionally, because more than one profile is stored for each person, a POSITA would have understood the stored signatures for each person at

² All emphases are added unless otherwise noted.

least includes multiple different profile views, such as various angled profile views and left/right profile views. Furthermore, the “series of video images” would contain various poses if the audience member was moving their head (e.g., to talk to an audience member sitting next to them). In other words, *Lu-000*’s recognition system accounts for various head orientations of the audience member captured in the video images by storing signatures for various head orientations (i.e., various different profiles) of the audience member’s head in the IFL.

42. The recognition system then switches to a “recognition mode” to recognize the audience members captured within the video images taken of the monitored area. *Lu-000*, 7:39-41. To recognize the audience members, a signature is extracted from the image and compared to “**each of the pattern image signatures stored in the individual face library**” in order to identify a match. *Lu-000*, 7:50-55. Because the extracted signature is compared to “each” stored signature, a POSITA would have understood *Lu-000*’s system to employ a brute-force method in identifying audience members. Namely, the extracted signature is compared to every signature in the IFL until a match is found or not found. *Lu-000*, 7:50-65 (describing comparing the extracted signature with each stored signature until a particular audience member is “identified” or left “unidentified”).

43. I note neither the learning mode nor the recognition mode includes a step of determining a head orientation within the images. Rather, the system just

extracts the signature from whatever head orientation is in the image and either (1) stores it in the IFL (in learning mode) or 2) compares it to stored signatures (in recognition mode). *See Lu-000 generally.*

C. OPINIONS REGARDING *TIAN*'S ALLEGED TEACHING OF "REDUCING A RESOLUTION"

44. Challenged Dependent Claims 4, 11, and 18 all recite the substantially similar step of "reduc[ing] a resolution of a first image of the one or more of the images of the media exposure environment to obtain a reduced-resolution image." '243 *Patent*, 15:56-58, 16:49-51, 17:41-43. Each of the remaining Challenged Dependent Claims either depends directly or indirectly from one of these dependent claims. Thus, all Challenged Dependent Claims include this step of "reduc[ing] a resolution."

45. Petitioner applies *Tian*'s performance of head pose estimation on a "down-sampl[ed]" image as teaching the reducing resolution limitation. *Petition*, 26-27 (citing "*Supra* §VI.A.2[1d-1e]," which applies *Tian*'s head pose estimation for teaching the claimed head orientation determination). Petitioner purports "a POSITA would have found it obvious to apply *Tian*'s teaching such that Lu-577's system reduces the resolution of a first image of the one or more images and performs face detection **or head pose estimation on the reduced-resolution image.**" *Petition*, 26; Ex. 1003, ¶ 120. I have been asked to provide my opinions as to whether *Tian* teaches

the “reduce a resolution” limitation. As discussed in detail below, it is my opinion *Tian* **does not teach** actively “reduc[ing] a resolution” of an image in a media exposure environment.

1. *Tian* Does Not Teach “Reducing a Resolution”

46. As discussed above, *Tian*’s purpose is to “explore” the effects of different image resolutions for each step of facial expression analysis. *Tian*, Abstract. To explore the effects of various resolutions, *Tian* analyzes “a total of five different resolutions” of images in a “widely used public database.” *Tian*, Abstract, 000002. *Tian* discloses that the “lower resolution images are down-sampled from the originals,” such that the lower resolution images resemble their higher resolution counterparts, except for the resolution difference. *Tian*, Abstract, 000002. In other words, various resolutions of a given image are analyzed for further action.

47. Petitioner relies on *Tian*’s analysis of down-sampled images to show *Tian* allegedly teaches the claimed step of reducing a resolution of a first image. *Petition*, 26. However, Petitioner states *Tian*’s passive statement (i.e., that images “are down-sampled”) supports Petitioner’s claim that *Tian* actively “**creates lower-resolution images** by down-sampling from original images.” *Id.* Based on the context of Petitioner’s selected quotes from *Tian* and *Tian*’s express goal of the exploration, it is my opinion that a POSITA would not conclude that *Tian* teaches

the facial expression analysis system actively performs the down-sampling of the images (i.e., reducing the images to create the reduced-resolution images).

48. As an example of *Tian's* context, Table 1 of *Tian* shows “An **example of five different resolutions from the Cohn-Kanade database**.” *Tian*, 000002. Petitioner cites a portion of this Table 1 caption and Petitioner’s Declarant, Dr. Doermann, cites the entire Table 1 for providing support of *Tian's* alleged creation of reduced-resolution images. *Petition*, 26; Ex. 1003, ¶ 119. Yet, neither Petitioner nor Dr. Doermann addresses *Tian's* express statement that the various resolutions (including lower resolutions) come “from” the database. A POSITA would not have understood the Table 1 caption to support that *Tian* teaches the expression system creating reduced-resolution images, as the statement indicates the database may store the necessary resolutions. Thus, *Tian* **does not teach** the facial expression system creating lower-resolution images or performing any reducing resolution step.

49. Additionally, when identifying the image resolutions that are explored, *Tian* states “five different resolutions of the head region are studied (288x384, 144x192, 72x96, 36x48, 18x24) **based on Cohn-Kanade AU-Coded Face Expression Image Database**.” *Tian*, 000002. The chosen resolutions being “based on” what the “Database” would not have indicated to a POSITA that the *Tian* system is performing the down-sampling. If *Tian's* system performed the down-sampling, *Tian* would not need to explain the resolutions being “based” from any database, as

Tian could just arbitrarily choose any “different resolutions” for exploring each step of facial expression analysis thereon. In other words, if *Tian* actually taught actively reducing the resolution of an image (as required in claims 4, 11, and 18), there would be no reason for *Tian* to describe using five different resolutions “based on” (i.e., coming from) the Cohn-Kanade Database. Because, if *Tian* were creating the reduced-resolution images, *Tian* could choose any random resolution to down-sample the images to. This creation of reduced-resolution images would not require basing the selected resolutions on any “database.”

50. Indeed, the purpose of *Tian*’s resolution effects exploration supports *Tian* using resolutions of readily available images. As discussed above, *Tian* explores the effects of different image resolutions for each step of facial expression analysis” for the purpose of studying the performance of facial expression analysis “in real environments” where the face image is “often in lower resolution.” See ¶ 36 (citing *Tian*, Abstract). Thus, *Tian* wants to explore how facial expression analysis performs on images taken “in real environments,” i.e., images that **start** with a low resolution. In fact, *Tian* has no interest in analyzing images that are not from a “real environment,” as “most” AFEA systems already analyze images from a “controlled laboratory situation.” *Tian*, Abstract, 000002. Based on *Tian*’s express purpose, a POSITA would understand *Tian* to teach analyzing facial expressions in images at whatever resolution they were captured with, e.g., whatever resolutions are available

from the database. Because *Tian* attempts to explore resolutions at which the images were captured with, a POSITA would understand *Tian* would be discouraged from actively **further** reducing the resolution of an image for the purpose of performing expression analysis.

51. For the above reasons, it is my opinion *Tian* fails to teach or suggest the claimed step of actively “reduc[ing] a resolution of a first image.”

**D. OPINIONS REGARDING THE MOTIVATION TO COMBINE
TIAN’S HEAD POSE ESTIMATION WITH LU-577**

52. I have been asked to provide my opinions as to whether a POSITA would have been motivated to modify *Lu-577*’s audience measurement system (as modified to include *Lu-000*’s recognition system) with *Tian*’s head pose estimation. Petitioner relies on the modification for dependent claims 4, 11, and 18. *Petition*, 27 (*citing* 22-23 for claim 4(b)), 37 (*citing* claim 4(b) mapping), 42 (*citing* mapping of 16(e) (i.e., “determining,... an orientation of the head with respect to the camera)). However, all of the motivations provided by Petitioner fail because either: (1) the motivations are based on irrelevant problems (e.g., a problem associated with face detection that is not commensurate with the claim scope of the ’243 Patent), or (2) the alleged improvements are flawed and would not, in fact, improve *Lu-577*’s facial recognition (as modified to include *Lu-000*’s comparison of extracted signatures with every stored signature). Thus, a POSITA would not have been motivated to

modify *Lu-577's* system (as modified with *Lu-000*) with *Tian's* head pose estimation/classification.

1. Petitioner's Proposed Modification

53. Petitioner asserts the combination of *Lu-577* and *Tian* for Ground 1. *Petition*, 3. As discussed above, *Lu-577* incorporates *Lu-000* by reference. See ¶ 37. Within the proposed combination of *Lu-577* and *Tian*, Petitioner cites to disclosures in *Lu-000* as teaching several limitations. *Petition*, 14-15, 18-21, 23-24. Petitioner also asserts the modification of *Lu-577* with *Tian* explicitly involves modifying *Lu-000's* image recognition system. *Petition*, 8 (stating “[a] POSITA would have found it obvious to **combine Lu-000 and Tian** because they are similarly directed to image processing and facial analysis” and stating *Tian's* techniques would “improve **Lu-000**”); see also, Ex. 1003, ¶¶ 57-58. Thus, based on Petitioner's express modification, it is my understanding that the disclosures relied upon from *Lu-000* for teaching various limitations within Ground 1 at least form the system being modified by *Tian*.

54. For example, Petitioner relies upon *Lu-000's* comparison of extracted pattern image signatures with stored pattern image signatures to determine matches and thereby identify audience members in mapping the claimed determining an identification “based on a match of the head to a known person.” *Petition*, 23-24 (citing *Lu-000*, Abstract, 2:20-38, 7:39-8:6). As discussed above, during recognition

mode, *Lu-000*'s system compares the extracted signature to **each stored signature** until a match is found (or no match is found). *See*, ¶ 42. I note both Petitioner and Dr. Doermann cite this disclosure of *Lu-000* comparing the extracted signature "with each of the stored pattern image signatures to identify a particular one of the predetermined audience members." *Petition*, 24 (citing Abstract, 2:20-38); *see also*, Ex. 1003, ¶ 108; *Lu-000*, 7:52-55. Thus, Petitioner's proposed modification of *Lu-577* at least includes modifying *Lu-000*'s recognition system with *Tian*'s head pose estimation/classification.

2. Modifying *Lu-577* with *Tian*'s Head Pose Estimation Does Not Achieve the Proposed Benefit of Improved Facial Recognition

55. Petitioner alleges modifying *Lu-577*'s facial recognition (as incorporated from *Lu-000*) with *Tian*'s head pose estimation used for facial expression analysis would "improve[] the performance of face recognition" by determining "a person's head pose to determine whether an image is suitable for face recognition." *Petition*, 10 (citing Ex. 1003, ¶¶ 65-66); *see also*, 9 (asserting a POSITA would have been motivated to modify *Lu-577* with *Tian* to "improve the confidence level of person identification and allow the selective performance of face recognition on suitable images (e.g., where a person opens her eyes and faces the camera."); Ex. 1003, ¶ 66 (Dr. Doermann opining the modification "improves the accuracy of identification of audience members"). However, contrary to Petitioner's

assertion, a POSITA would not have been motivated to incorporate *Tian*'s head pose estimation into modified *Lu-577* because the combination:

- would not improve the accuracy of *Lu-577*'s facial recognition (as modified to include *Lu-000*'s signature comparison);
- would add an unnecessary step to the recognition process without any improvement in accuracy, cost reduction, or other benefit; and
- would decrease the performance of *Lu-000*'s face recognition by determining certain images to be unsuitable for facial recognition that would've actually achieved successful recognition.

I. The Modification Does Not Improve Facial Recognition Accuracy

56. It is my opinion Petitioner's proposed modification of including *Tian*'s head pose estimation would not improve the facial recognition accuracy of *Lu-577*'s system (as modified to incorporate *Lu-000*'s image recognition process). As discussed previously, *Lu-000* implements a brute-force technique when matching the extracted image signature with every stored signature to find a match and thereby identify the audience member. *Lu-000*, 7:41-55; *see* ¶ 42. Additionally, the stored signatures include multiple signatures stored for each audience member to account for various angles of the head captured during recognition mode. *See* ¶¶ 40-41 (*citing Lu-000*, 7:31-38, 8:36-45). In other words, not only does *Lu-000*'s matching process

check every stored signature, but it also checks multiple head profile signatures for each audience member.

57. *Lu-000's* matching procedure accounts for various head profiles that may be captured for an audience member by comparing the extracted signature with all stored profile signatures. Therefore, modifying *Lu-000's* recognition system to include *Tian's* head pose estimation is **redundant**. For this reason, a POSITA would have understood modifying *Lu-000's* brute force method to determine a head orientation via *Tian's* head pose estimation would not improve the face recognition accuracy, as Petitioner and Dr. Doermann purport. Rather, at best, the modification **would result in the same level of facial recognition accuracy provided by *Lu-000's* un-modified system.**

58. Indeed, because the brute force method already accounts for various head profiles of the audience member, the need for any head pose determination is **eliminated**. When *Lu-000's* recognition process tries to check the extracted signature against **all** stored profile signatures to find a match, various head orientations (even for each audience member) are already accounted for. This makes any additional head pose determination, let alone *Tian's* head pose estimation, unnecessary within the facial recognition process of *Lu-000*.

59. *Lu-000's* system stores multiple profile signatures for a single person **“depending on any changed facial expression or various profiles.”** *Lu-000*, 8:36-

45; *see also* ¶ 41. This dependency on varying profiles within an image indicates *Lu-000*'s system **wants** to store multiple different head orientations for each person to be able to correctly match different profiles of each audience person that may be captured in the feature-extracted image. In this way, because multiple profiles are stored, the feature-extracted image will match with one of the stored signatures.

60. Thus, a POSITA would have understood the feature-extracted images from digital representations of captured video images in *Lu-000* already serve as “suitable” images for face recognition. This suitability means there is no need to employ *Tian*'s head pose estimation or any head orientation determination within *Lu-000*'s recognition process. Furthermore, because *Lu-000* already accounts for various head orientations in the feature-extracted image through matching the image to the multiple stored profile signatures, *Lu-000* already provides Petitioner's purported benefit of improved facial recognition accuracy. Thus, a POSITA would not have been motivated by improving facial recognition accuracy to modify *Lu-000* to determine a head orientation, let alone employ *Tian*'s head pose estimation to allegedly determine such head orientations.

II. Modifying Lu-000 as Proposed Would Add an Additional Step to Facial Recognition Without any Worthwhile Benefit

61. As outlined in the previous section, *Lu-000*'s recognition system already provides accurate face recognition, and this accuracy does not increase by

including *Tian's* head pose estimation. Thus, in the proposed modification, an additional step would be added to *Lu-000* for performing the facial recognition, without any benefit such as a lower cost or decreased processing time. In fact, neither Petitioner nor Dr. Doermann provide any other benefit (e.g., lower cost, decreased processing time) resulting from incorporating *Tian's* head pose estimation, because there would be no such benefit. *Petition*, 10; Ex. 1003, ¶ 65-66.

62. Rather, within the proposed modification, several steps would need to be added into the recognition process. For example, algorithms would have to be included into the system to determine orientations of the head. Per *Tian*, and to achieve Dr. Doermann's purported benefit of finding "suitable" images, the recognition mode would require modification to include algorithms for filtering out captured images that are not found suitable, based on the head pose estimation. Dr. Doermann expressly cites to *Tian's* **exclusive use of frontal/near frontal poses** as indicating these poses are "suitable" for facial **expression analysis**. Ex. 1003, ¶ 66. Under Dr. Doermann's interpretation, to achieve the proposed benefit, *Lu-000's* recognition system would require reconfiguration of, e.g., the head location block 104 to (1) determine the head orientation; and (2) remove any captured images that do not contain a frontal/near frontal pose. *Lu-000*, 7:39-45.

63. These steps would also be required within the learning mode of *Lu-000's* recognition system such that only "suitable" (e.g., frontal/near frontal) pattern

signatures are stored. For example, the face image location block 92 would require reconfiguration to determine the head pose (per *Tian*) within the image and only perform subsequent processing on frontal/near frontal head images for storage. After all, the proposed modified system would only be interested in attempting to match frontal/near frontal feature-extracted images. In short, several reconfigurations of multiple components within *Lu-000*'s recognition system would be required according to Petitioner's proposed modification. These reconfigurations implementing additional steps would not only increase the complexity of the system, but would do so for no apparent benefit that is not already provided by *Lu-000*'s non-modified facial recognition.

64. In fact, the additional steps required would increase the processing required to perform facial recognition. The increased processing would merely result in a waste of processing time and energy within the modified system. Thus, adding *Tian*'s head pose estimation would require several additional steps be included in *Lu-000*'s recognition system to determine the orientation of the head within (1) the images to be stored; and (2) the images to be matched to identify the "suitable" frontal/near frontal images.

65. Additionally, I note that Petitioner's combination **does not** substitute *Lu-000*'s brute-force method with *Tian*'s head pose estimation, as evidenced by Petitioner relying on steps within the brute-force matching process for teaching

certain limitations for Ground 1 (i.e., limitation reciting matching a head to a “known person” to identify them). Furthermore, even if *Lu-000*’s storage of multiple profile signatures was substituted for *Tian*’s head pose estimation or a method of head orientation determination, one would still end up with, at best, the same facial recognition accuracy as provided by *Lu-000*’s non-modified system. This is because the non-modified *Lu-000* already accounts for various head profiles when recognizing faces via brute-force comparison with various stored profiles.

***III. The Modification Would Decrease the Performance of
Lu-000’s Facial Recognition***

66. As discussed in the previous section, modifying *Lu-000*’s facial recognition to include *Tian*’s head pose estimation would add additional steps without providing any commensurate benefit. Additionally, the proposed modification would **hinder** the performance of *Lu-000*’s recognition system. Namely, using *Tian*’s head pose estimation for finding “suitable” images (i.e., frontal view images) would decrease the amount of successful facial recognitions produced relative to *Lu-000*’s unmodified process, due to the modified system only performing facial recognition on frontal/near frontal images.

67. Petitioner asserts a POSITA would have found it obvious to determine a head pose “to determine whether an image is suitable for face recognition.” *Petition*, 10. Dr. Doermann supports this statement by opining, in the modified

system, a “person’s head pose is first determined in an image and a decision is made as to whether the image will be used for face recognition.” Ex. 1003, ¶ 66. Dr. Doermann cites *Tian*’s teaching of performing **facial expression analysis “only for frontal or near frontal view faces.”** Ex. 1003, ¶ 66 (*citing Tian*, 000002). Specifically, *Tian* teaches filtering the images prior to performing facial expression analysis based on head pose classification. *See* ¶ 36. Thus, the proposed modification includes applying *Tian*’s head pose estimation and classification in determining only “suitable” **frontal and near frontal images to apply facial recognition thereto.**

68. However, applying this pre-filtering in the modified system would negatively affect *Lu-000*’s facial recognition abilities. Namely, the overall amount of successful facial recognitions would decrease. *Lu-000*’s brute-force matching comparison allows various head profiles to create a successful face recognition. Contrastingly, in the proposed modified system, only frontal and near frontal images would be chosen for performing facial recognition. A POSITA would have understood this change to decrease the amount of facial recognition processes performed. The video camera of *Lu-000*’s facial recognition system would have to capture images with the head in a frontal or near frontal profile for the system to even attempt to identify an audience member.

69. Furthermore, because *Lu-000*’s non-modified system successfully matches various head orientations (including non-frontal) to multiple stored

orientation signatures, the frequency of successful facial recognition operations would decrease in the modified system. Understandably, matching less extracted signatures to *Lu-000's* individual face libraries would create less successful facial recognitions when compared to the non-modified system. Additionally, as discussed above, the modified system would store less signatures, as only frontal or near frontal images would be required for matching to identify faces. *See* ¶¶ 62-63. Thus, the decreased number of candidate images **and** the decreased number of stored signatures would decrease not only the number of facial recognition operations, but the number of successful recognitions. A POSITA would have understood this modification would thereby recognize fewer faces overall compared to the non-modified recognition system.

70. Dr. Doermann opines determining “suitable” images for facial recognition “would have improved the accuracy of the audience identification.” Ex. 1003, ¶ 61. Dr. Doermann supports his opinion regarding **facial recognition accuracy** with a reference describing challenges occurring in **face detection** for the purpose of **face tracking**. Ex. 1003, ¶ 61 (*quoting Bigioi*, Ex. 1010, [0003]); *Bigioi*, [0003] (describing “**typical face detector[s]** can only **accurately detect faces** in a semi-frontal pose.”). Dr. Doermann also attempts to conflate **expression analysis** with face recognition, citing *Tian's* performance of facial expression analysis on only “frontal and near frontal view faces” as supporting improved accuracy for facial

recognition. Ex. 1003, ¶ 66. However, these three processes (face detection, facial expression analysis, facial recognition) are distinct from one another and have distinct problems and constraints associated therewith, as I discuss below.

71. Facial expression analysis is a distinct process from facial recognition and has different constraints and feasibilities associated therewith. *Facial Expression Recognition: A Brief Tutorial Overview*, Ex. 2009, 1 at col. 2, par. 2 (describing facial expression recognition contrasting from facial recognition because “personal identity information conveyed by a face **is an unwanted source of variability in expression recognition**” and “[c]onversely, **variability arising from facial expression is unwanted in face recognition**, where the uniqueness of the face is the central recognition criteria”). For example, to recognize facial expressions, you may look at the movement of relevant features of the face (e.g., eyebrows, mouth, and eyes), as various movements of features are associated with making different facial expressions. *Facial Expression Recognition*, Ex. 2010, 1 (describing the process of facial expression recognition comprising “[a]nalyzing the motion of facial features and/or the changes in the appearance of facial features and classifying this information into some facial-expression-interpretative categories such as facial muscle activations like smile or frown, emotion (affect) categories like happiness or anger...”); *Automatic Analysis of Facial Expressions: the State of the Art*, Ex. 2011, 1425 at col. 2, par. 3 (describing a problem of facial expression

analysis is producing a visual representation of the “shown facial expression,” and that the visual information provided by a face “could be made clear by describing the movements of points belonging to the facial features (eyebrows, eyes, and mouth) and then by analyzing the relationships between those movements”). Contrastingly, recognizing a face may require fewer facial features or different facial features than expression analysis, in order to match the image with a stored representation of a person. Ex. 1021, 000008 at col. 1, par. 1 (describing facial features often used for facial recognition are the “eyes, nose, mouth”), 000006 (describing studies that show the nose may play an “insignificant role” in face recognition, and that the eyes and mouth “have been determined to be important for perceiving and remembering faces”). Consequently, it may be feasible to perform facial recognition on images that are not feasible for successful facial expression analysis.

72. Additionally, unlike expression analysis, facial recognition does not include analyzing the movement of facial features, but rather may use distances between facial features in an image in order to identify the face. *Mechanisms of Face Perception*, Ex. 2012, 3-4 (generally describing how facial feature positions are “identified and used to compute various geometric ratios” when using “feature-based methods” for identifying a face). Thus, the position of the facial features is important to facial recognition, not the movement of those facial features. This is because facial

recognition does not care about what expression the face is making, but cares about who the face belongs to. Based on the above distinctions (among others), the conditions required, and problems faced for facial expression analysis **do not support any motivation related to facial recognition accuracy**.

73. The problems and constraints affecting face detection are also distinct from those associated with facial recognition and those associated with facial expression analysis. Firstly, face detection only requires detecting a face structure, while facial recognition requires analyzing the face within the image and comparing it to another image to determine a match. *Mechanisms of Face Perception*, Ex. 2012, 2 (describing face detection as “simply detecting the presence of a face,” while identification (i.e., recognition) compares measured features “to stored templates”), *see also*, 2 (describing that face detection and “identification” (i.e., recognition) have “**opposing demands**: The identification of individuals requires a fine-grained analysis to extract the ways in which each face differs from the others despite the fact that all faces share the same basic T-shaped configuration, whereas detection requires extracting what is common to all faces”). Furthermore, face detection is known to have other constraints, such as, for example, checking for movement of the face to ensure the captured image is of a person and not of, e.g., a face in a painting. Ex. 1016, 000009 (describing the metering system scanning the room for “movement, the hallmark of humanness” when detecting a face in a room). This

constraint is not necessary to facial recognition, as the image (regardless of where it came from) either matches or does not match a person's stored face image.

74. Additionally, in face detection, having more of the face visible may provide more confidence that the structure in the image is, in fact, a face. For example, Ex. 1010 states a “typical face detector” used for face detection can “only accurately detect faces in a semi-frontal pose,” and cannot detect a face when the face is in a “non-frontal or profile pose, facing instead either too much up, down, left or right.” Ex. 1010, *Bigioi*, [0003]. Thus, having more facial features visible (such as in a semi-frontal pose) improves the accuracy of face detection when compared to a “non-frontal pose” where not as many facial features are visible. Contrastingly, in recognizing a face, less facial features may still provide an accurate result because the system merely compares the facial features visible with a stored image containing those facial features to determine a match/no match to a person. Knowing these distinctions, a POSITA would also not have found problems in face detection to relate to facial recognition, let alone support a motivation to modify a facial recognition system. Dr. Doermann himself does not provide any explanation as to why a POSITA would have looked at these distinct problems in facial expression analysis and face detection when improving facial recognition. Ex. 1003, ¶¶ 65-66. Thus, neither *Bigioi* nor *Tian* support Dr. Doermann's opinion regarding the accuracy of **facial recognition**.

75. Rather, contrary to Dr. Doermann's opinion, a POSITA would have understood head poses other than "frontal" or "semi-frontal" would provide successful facial recognition. In fact, one of Dr. Doermann's cited references, Ex. 1021, states as much. Ex. 1003, ¶ 66 (*citing* Ex. 1021). Ex. 1021 is a rather extensive publication providing a "Literature Survey" on "Face Recognition." *Face Recognition: A Literature Survey*, Zhao et al., Ex. 1021, 000001. The publication discloses that identifying people via "personal identification system[s] based on **analysis of frontal or profile images of the face is often effective**." Ex. 1021, 000002. Thus, performing facial recognition on profile images (i.e., side view images, side-angled images, etc.) was known to be successful, in addition to performing on frontal images. Indeed, for certain applications, recognizing profile images is "important," such as in mugshot matching. Ex. 1021, 000006 (describing "using profiles" in "face recognition" may be "important in mugshot matching applications, where profiles can be extracted from side views"). Indeed, certain methods for feature extraction used in facial recognition systems were known to recognize profile views by applying "bilateral symmetry of the face or using learned information." Ex. 1021, 000010; *see also*, 000021, col. 1, par. 3 (describing applying techniques using eigenfaces methods for performing facial recognition on "side view" images and images from "multiple views"), 000034, col. 2, par. 2 (describing estimating a "3D shape vector of a face" from "a set of 2D face images **in different**

views”). In conclusion, Dr. Doermann’s own cited reference provides evidence that head poses other than frontal and semi-frontal were known to result in successful facial recognition. Thus, finding “suitable” images of only frontal or “semi-frontal” head poses would not have improved the accuracy of face recognition in *Lu-000*, as Dr. Doermann opines, because other head poses were known to provide successful recognition results. Ex. 1003, ¶ 66.

76. *Steinberg* (cited by Petitioner) also provides support that images capturing head poses other than frontal provide successful facial recognition. Petitioner cites *Steinberg* as allegedly teaching the claimed determining “an orientation of the head” and the claimed matching the head “to a known person” to identify the person. *Petition*, 56-61. Namely, Petitioner purports *Steinberg* determines an orientation of the head “as it reflects different rotation angles of the head in an image.” *Petition*, 57 (citing *Steinberg*, Ex. 1007, [0167], FIGs. 6(a)-6(e)). To generate a “face print” to match with stored “known faceprints,” *Steinberg* **normalizes head poses other than frontal captured in images**, in addition to utilizing frontal images. *Steinberg*, [0169-0171], [0178]. In fact, *Steinberg* discloses a “second implementation” of the “normalization” where the system does no re-alignment of the head pose within the image and, rather, compares that head pose to “a plurality of face recognition datasets.” *Steinberg*, [0173]. Each dataset “is associated with a particular discrete range of pose angles,” including left/right

rotations from 0 degrees up to 90 degrees. *Id.* Thus, Petitioner's own cited art discloses using various **non-frontal head poses** (e.g., angled left/right by up to 90 degrees) captured in images to **successfully perform facial recognition**. Accordingly, a POSITA would not have understood frontal or semi-frontal images to be the only "suitable" images for providing accurate facial recognition, and thus would not have been motivated to "improve" *Lu-000's* facial recognition accuracy by only performing the recognition process on frontal or semi-frontal images. Ex. 1003, ¶ 66.

77. I also note *Lu-000's* non-modified system indicates successful facial recognition on "profile" images, as signatures are stored for "various profiles" which are then compared to extracted signatures to identify an audience member. *Lu-000*, 7:50-55, 8:36-45. Based upon these references, a POSITA would have understood face recognition would successfully be performed on head orientations other than frontal. Thus, limiting a system (as Petitioner proposes) to only perform recognition on frontal/near frontal would limit the successful identification of a person. Additionally, this limiting would provide no accuracy improvement to *Lu-000's* modified system.

78. For the reasons discussed above, a POSITA would have found no reason to modify *Lu-000's* recognition system to include *Tian's* head pose

estimation to determine “suitable” images for facial recognition. *Petition*, 10; Ex. 1003, ¶ 66.

3. Dr. Doermann’s Cited Background References Do Not Support His Opinions

79. Dr. Doermann applies several background references for supporting his opinions regarding the motivation to combine *Tian*’s head pose estimation into *Lu-577*’s facial recognition (as modified by *Lu-000*’s recognition system). Ex. 1003, ¶¶ 66-67. However, as discussed in detail below, upon closer analysis, **none** of the background references support Dr. Doermann’s opinion regarding **improving facial recognition**.

I. Ex. 1021

80. Dr. Doermann relies on Ex. 1021 to support his obviousness opinions. Namely, Dr. Doermann seemingly relies on Ex. 1021 to support his opinion that modifying *Lu-577* (as modified by *Lu-000*) to include *Tian*’s head pose estimation “addresses the problem that pose variations adversely affect recognition performance.” Ex. 1003, ¶ 66 (*citing* Ex. 1021, 000042). However, Ex. 1021 **does not** support Dr. Doermann’s opinion.

81. Dr. Doermann seems to be relying on the Section 6 Introduction in Ex. 1021 that states “pose change can cause serious performance degradation.” Ex. 1021,

000042. Placing these statements in the intended context, the “performance degradation” expressly occurs in “uncontrolled environment[s]”:

The best face recognition techniques reviewed in Section 3 were successful in terms of their recognition performance on large databases in well-controlled environments. However, face recognition in an uncontrolled environment is still very challenging. [...] Though many existing systems build in some sort of performance invariance by applying preprocessing methods such as histogram equalization or pose learning, significant illumination or pose change can cause serious performance degradation. [...]. These problems are unavoidable when face images are acquired in an uncontrolled, uncooperative environment, as in surveillance video clips.

Ex. 1021, 000042.

82. Dr. Doermann ignores this context. When read with this understanding, the “advers[e] affect[s]” Dr. Doermann relies on for motivating the combination of *Tian*’s head pose estimation/classification into *Lu-000*’s recognition system are irrelevant. The “pose change” problem in Ex. 1021 correlates to face recognition using images captured in “an uncontrolled, uncooperative environment,” such as, e.g., surveillance images. Ex. 1021, 000042. Contrastingly, *Lu-000* captures video images of audience members’ faces in a “monitored area.” *Lu-000*, 3:27-30, 3:36-41, 3:65-4:3. Such a “monitored area” would include, for example, a room where

people are watching TV, such as a living room. *Lu-000*, 1:13-18 (describing the invention “field” as image recognition systems “for use with **television audience measurement**” in identifying individual members “of a **viewing audience**”). A POSITA would have understood in such an environment, the audience members viewing the TV are facing straight forward (i.e., toward the TV) and are cooperative.

83. A POSITA would have understood *Lu-000*’s “monitored area” to be a controlled environment, rather than one “uncontrolled” (e.g., taking images with surveillance camera). For example, audience members would be sitting fairly still, within a relative range of distances from the camera, and facing in generally one direction in order to be engaged with and view the TV program. Additionally, because *Lu-000*’s system employs “covert” illumination in the monitored area while people are watching TV, a POSITA would have understood that the areas maintain relatively constant lighting. All of these factors indicate *Lu-000*’s monitored area is well-controlled. Thus, based on Ex. 1021, *Lu-000*’s environment is precisely the environment that provides “successful” recognition performance using existing facial recognition “techniques.” Ex. 1021, 000042. For this reason, the cited problem in Ex. 1021 that occurs in “uncontrolled” environments would not motivate a modification of *Lu-000*.

84. I note Dr. Doermann provides no detail for his interpretation or understanding of “pose variation” or “pose change” as it is used in his quote from

Ex. 1021. In collectively reviewing Ex. 1021's teachings, "pose variation" and "pose change" refers to a captured image having a different pose from a pose in a stored image to match the captured image therewith for recognition. Ex. 1021, 000036 (the testing of a facial recognition system's ability to identify faces including a third test that "measured the effect of pose changes" on the performance of matching an image to a stored image), 000038 (testing facial recognition systems for recognizing people "under pose variations"), 000053 (describing results of "tracking" a person's "face across large pose variations"). Indeed, the general "problem" addressed in the publication is that of identifying "one or more persons" from "still or video images" using a "stored database." Ex. 1021, 000002. Thus, because an image of a person is being compared to a "database of faces" to identify them, the issue of "pose changing" would happen when the pose of the captured image is different from the pose of the image stored in the database. *Id.* In fact, Figure 17 provides, as an example, the "FERET" database that comprises images of various poses of a single individual:



Fig. 17. Images from the FERET dataset; these images are of size 384×256 .

Ex. 1021, 000037. The tests performed on the FERET database (including images of a same person with various poses) revealed “major problem[s]” within facial recognition, including recognizing people “under pose variations.” Ex. 1021, 000038. Because the database contained images of the same person in different poses, these tests identifying the pose variation problem involved attempting to match one pose image with an image of a different pose of the same person.

85. For additional “pose change” context, Ex. 1021 discusses results from a “Face Recognition Vendor Test,” including results related to “the effect of pose angle” on facial recognition performance. Ex. 1021, 000038. The test found a difference of 25 degrees (“ $\pm 25^\circ$ ”) “does not significantly affect performance.” *Id.* Essentially, when a captured pose is within 25 degrees of the pose angle captured in the stored reference image to match with, there was no significant effect on the system’s ability to successfully perform recognition. A 40 degree pose angle difference between the captured head pose and stored image head pose, however,

was found to “significantly” affect recognition performance. Ex. 1021, 000038. Thus, “pose variation” means the variation between a pose of a person in a captured image and a pose of that same person in a stored image to be matched with.

86. Indeed, in the context of viewing images captured in “uncontrolled” environments, it would be important to explore how a change in pose of an image taken of an unknown person would affect matching that image to a known image in a database to identify them. For example, a POSITA would appreciate that, in facial recognition, a surveillance video that may have captured a burglar running away from the camera (i.e., capturing a side view or semi-obstructed view of the face) could be matched against mugshot images (e.g., frontal pose images) to identify the criminal. Ex. 1021, Table 1 (providing an exemplary application of facial recognition as “shoplifting, suspect tracking and investigation”); 000042, col 1, para 4 (describing the problem of “pose change” occurring when analyzing “surveillance video clips” for facial recognition); 000021, col 1, para. 3 (describing applications of facial recognition that utilize mugshots). Thus, in these applications of face recognition (i.e., where images are captured in an “uncontrolled” environment), a POSITA would have understood the problem of “pose change” to relate to attempting to match an uncontrolled head pose (e.g., head posed in any random direction) captured in an image of an unknown person to a reference image (containing a different head pose) stored in a candidate library.

87. Because the “pose variation” problem in Ex. 1021 refers to pose differences between the captured image and the stored reference image to match therewith and relates to “uncontrolled” imaging environments, this problem does not motivate the modification of *Lu-000*’s recognition system, as Dr. Doermann opines. Ex. 1003, ¶ 66. *Lu-000*’s method is not applicable to “uncontrolled” settings (e.g., surveillance video applications) of image capture where pose variation occurs. For example, *Lu-000* implements the brute force facial recognition method for “predetermined” individuals in a television audience within a “monitored area.” *Lu-000*, 2:20-24, 3:33-36 (generally describing the image recognition system transferring stored data to a “television audience data” collection system”). In such a “monitored area,” a POSITA would have understood the TV is located a certain distance from the audience members and placed at a particular location within the monitored area. Thus, the audience members being recognized would face a known direction (i.e., the TV) and be within a range of distance from the TV (e.g., based on how far back the seats are from the TV). This relatively controlled setting is quite different from an uncontrolled setting where, for example, a stranger walks in front of a surveillance camera at any distance from the camera and could be looking in any direction.

88. Additionally, **there is no pose variation problem associated with *Lu-000***, because *Lu-000* compares an extracted feature signature to several stored

signatures of various poses for each audience member. As discussed above, *Lu-000* stores multiple pattern signatures of various head poses for each person. See ¶¶ 40-41. *Lu-000* then attempts to match an extracted signature with each pose signature in the library to perform successful facial recognition. Thus, *Lu-000* matches the extracted signature with a **same pose in one of the stored signatures**, rather than attempting (and potentially failing) to match an extracted signature with a different pose in a stored image (i.e., the crux of the pose variation problem). Indeed, the insignificant difference between pose angle changes of $\pm 25^\circ$ further indicates *Lu-000* will find a match with the various pose signatures stored in the library. *Lu-000*'s recognition system can store a relatively small library of various poses (e.g., "various profile[]" head poses), and still successfully identify an audience member. *Lu-000*, 7:25-38, 8:36-45. Namely, the extracted signature could differ from the stored pose signatures by up to 25 degrees and still result in a match. Thus, these comparable recognition results provide evidence that the "pose variation" problem Dr. Doermann alleges would motivate the combination has no effect on *Lu-000*'s recognition system. Ex. 1003, ¶ 66. And, accordingly, a POSITA would not have been motivated by this problem to modify *Lu-000*.

II. Ex. 1012

89. Dr. Doermann also cites to *Weinblatt* (Ex. 1012) for supporting Petitioner's proposed motivation for modifying *Lu-577*'s system (employing *Lu-*

000's face recognition system). Ex. 1003, ¶ 67 (*citing* Ex. 1012 (“Weinblatt”)). As a second motivation, Dr. Doermann opines a “POSITA would have been motivated to apply Tian’s teachings of head pose estimation to that of Lu-577 **“such that the television audience measurement system can determine whether a member of the audience is watching television.”** *Id.* Dr. Doermann supports his opinion by citing *Weinblatt* to provide an exemplary system that allegedly used “head pose to determine individual’s watching activity.” Ex. 1003, ¶ 67. Specifically, Dr. Doermann cites a quote from *Weinblatt* regarding using gaze tracking (i.e., what the subject is ‘looking at’) to determine if someone is watching TV:

For example, the facial recognition unit 11 **may use the measurements of the angle of a face, if the eyes of the face are open or closed, and/or if the face of the person is ‘looking at’ the camera** or at the broadcast program to determine if the person is watching the broadcast program.

Weinblatt, [0039].

90. Though *Weinblatt* may teach using gaze tracking techniques, such as measuring face angles, eye positioning, and detecting eyes open/closed to determine a person’s “watching activity,” these measurements **are not part of Tian’s head pose estimation**. In fact, *Tian* does not suggest performing any of these measurements. Thus, when viewing *Weinblatt*, a POSITA would not have been

motivated to modify *Lu-577's* audience measurement system (employing *Lu-000's* facial recognition) with *Tian's* head pose estimation to determine watching activity.

91. Rather, Petitioner's proposed modification would not result in a system that can determine whether an audience member is watching TV. *Tian* estimates head poses by classifying each image into 3 fairly broad categories: frontal/near frontal, side view, and others. See ¶ 36. *Tian* classifies these poses for the sole purpose of determining which images are suitable for facial expression analysis. *Id.* Contrastingly, in determining whether someone is watching TV, *Weinblatt* employs three different (and comparatively, sophisticated) methods of gaze tracking: (1) measuring the angle of the face; (2) determining if the eyes are open or closed; and (3) determining whether the face is "looking at" the TV. Thus, a POSITA would have understood *Tian's* head pose estimation and classification system does not determine the information measured in *Weinblatt*.

92. Namely, when viewing *Tian's* exemplary images placed within the three pose categories, this inability becomes quite clear. For example, when viewing the "frontal" category (left column), three exemplary images are provided:

Poses	Frontal or near frontal	Side view or profile	Others
Examples			

Tian, Table 2 (annotated). Though these three images are placed in the **same pose category**, one image is looking directly toward the camera (blue) and the other images are looking to the left of the camera (red). This distinction indicates *Tian* cannot differentiate between different angles of the face, as the head in the blue image is angled ~ 0 degrees from the camera, the left red image is angled a few degrees (e.g., ~ 5 degrees) to the left of the camera, and the right red image is angled even more toward the left (e.g., ~ 15 or more degrees). Thus, a POSITA would have understood *Tian* does not measure “the angle of the face” when determining head poses. Furthermore, under similar reasoning, a POSITA would understand *Tian* also does not determine if the face is “looking at” the camera, because two images with the face looking **to the left of the camera** are similarly categorized with an image having **the face looking directly toward the camera**.

93. Similar angle differentiations occur between the images placed in the same “side view” category and “others” category. Lastly, when viewing *Tian*’s pose

categories, the eyes being closed or opened is not a factor affecting the estimation of the pose. Namely, none of the categories are defined as having “eyes closed” or “eyes open.” *Tian*, 000003. Additionally, a POSITA would have understood **both an image with closed eyes OR open eyes could equally be classified in any of the three pose categories**. Namely, a face with closed eyes could be in a “frontal” pose, a “side view,” or some “other” pose. Thus, the head pose classification of *Tian* does not determine whether eyes are open or closed and distinguish between the two alternatives.

94. Thus, though Dr. Doermann relies on *Weinblatt's* description of determining when someone is watching TV to support the motivation, this description only includes measurements that are **not** performed in *Tian*. This reference would, in fact, not motivate a POSITA to modify *Lu-577's* system (as modified to include *Lu-000's* recognition system) with *Tian's* head pose estimation for the purpose of determining “whether a member of the audience is watching TV,” at least because the modification would not actually achieve the purported benefit. Ex. 1003, ¶ 67. Furthermore, a POSITA would not have been motivated to implement *Tian's* head pose classification for being able to determine when someone is watching TV, because *Tian's* head pose classification cannot provide such a determination.

III. Ex. 1010

95. The last reference Dr. Doermann cites as allegedly supporting the motivation “to determine whether an image is suitable for face recognition” is *Bigioi*. *Petition*, 10; Ex. 1003, ¶ 66 (*citing* Ex. 1010, *Bigioi*). Specifically, Dr. Doermann cites *Bigioi* as supporting his opinion that “a POSITA would have understood that certain poses are not suitable for **face recognition**.” *Id.* Dr. Doermann goes on to opine “in a side view, one or more facial features of a person may not be visible, which may make face recognition with a high accuracy unfeasible,” citing *Bigioi*, [0003] that discloses “a typical face detector can only accurately detect faces in a semi-frontal pose.” Ex. 1003, ¶ 66 (*citing Bigioi*, [0003]).

96. From the quotation alone, one can see *Bigioi*’s description of accuracy **is not related to facial recognition accuracy, but rather is discussed in the context of face detection**. *Bigioi*, [0003]. As discussed previously in ¶¶ 71-75, problems and constraints associated with face detection may not be relevant to the process of facial recognition. *See* ¶¶ 73-74. Therefore, the problem that **face detection** is limited to semi-frontal poses addresses a different process than **facial recognition** (as required by the claims and as urged by Dr. Doermann); namely, that of face detection. *Bigioi*, [0001-0002]. Dr. Doermann’s citation to a reference describing face detection for supporting a statement regarding facial recognition further shows his conflation of the disparate fields.

97. Petitioner's own cited art (*Steinberg*) contradicts Dr. Doermann's alleged motivation of finding "suitable" images, as discussed earlier. *See* ¶ 76. Additionally, Ex. 1021 also discusses various head poses, other than frontal and "semi-frontal," being suitable for performing accurate facial recognition thereon. *See* ¶ 75. For example, Ex. 1021 explicitly describes performing facial recognition "using profiles (which **may be important in mugshot matching applications, where profiles can be extracted from side views**)." Ex. 1021, 000006. Thus, Dr. Doermann's own background reference contradicts his opinion that "side view[s]" of faces "make face recognition with a high accuracy unfeasible." Ex. 1003, ¶ 66.

98. Ex. 1021 even provides several techniques that may be used to employ profile head poses (i.e., poses where part of the face is "invisible in an image") to produce accurate facial recognition results. Ex. 1021, 000034 (describing "[w]hen part of a face is invisible in an image due to rotation," the invisible part of the face may be "recovered" using "bilateral symmetry of faces."). Further discussions of profile images (e.g., side view images) being "effective" for performing facial recognition were previously discussed. *See* ¶¶ 75-77. Thus, the factual evidence contradicting Dr. Doermann's opinion indicates that, on the contrary, poses other than "semi-frontal" would be "suitable" for face recognition.

4. Opinions Regarding Claims 4, 11, and 18: Petitioner Provides No Credible Motivation for Analyzing Reduced-Resolution Images and Tian Teaches That Doing So is Undesirable

99. I have been asked to provide my opinions regarding Petitioner's proposed motivation for combining *Tian*'s alleged teaching of reducing a resolution into *Lu-577*'s system (modified to employ *Lu-000*'s recognition system), as mapped for dependent claims 4, 11, and 18. *Petition*, 26-27. Petitioner applies *Tian*'s head pose estimation for allegedly teaching determining a head orientation based on the "reduced-resolution image" (also recited in claims 4, 11, and 18). *Petition*, 27. As discussed above, Petitioner asserts a POSITA would have found it obvious to apply *Tian*'s teaching to *Lu-577*'s system to "reduce[] the resolution of a first image" and perform "head pose estimation on the reduced-resolution image." See ¶ 45. Petitioner relies on *Tian* for allegedly teaching the reducing resolution step, and asserts that *Lu-577* in combination with *Tian* teaches dependent claims 4, 11, and 18. *Petition*, 26-27, 36, 42; Ex. 1003, ¶¶ 117-120.

100. In my opinion, *Tian* does not teach a step to "reduce the resolution of a first image." See ¶¶ 46-51. Separately, a POSITA would **not have been motivated by Petitioner's proposed motivation** to modify *Lu-577*'s system (as otherwise modified to implement *Lu-000*'s recognition system) to perform *Tian*'s head pose estimation on a reduced-resolution image. Namely, Petitioner's arguments are

directed to a motivation to perform **face detection** on a reduced-resolution image, rather than providing a motivation to **perform the claimed head orientation determination on a reduced-resolution image**. Thus, Petitioner's motivation does not actually address the claim language that Petitioner is mapping the combination to; namely, that a POSITA would have been motivated to modify *Lu-577's* system (as otherwise modified to implement *Lu-000's* recognition system) to reduce a resolution of a first image and determine a head orientation based on the reduced-resolution image, as allegedly taught by *Tian*.

101. Additionally, a POSITA would not have been motivated to make the proposed modification, because a POSITA reviewing *Tian* would understand reduced-resolution images are **less desirable** than higher resolution images when performing expression analysis, and thus would be discouraged from the path of **intentionally** reducing an image resolution to perform head pose estimation on the reduced-resolution image.

102. Thus, Petitioner's motivation to combine *Tian's* alleged teaching of determining a head orientation (i.e., *Tian's* mapped head pose estimation) based on a reduced-resolution image into *Lu-000's* recognition system fails to motivate a POSITA to make the proposed modification, and, in fact, a POSITA would not be motivated to make the proposed modification.

I. Tian's Mapped Teachings

103. As discussed above, *Tian* explores the effects of various image resolutions on facial expression analysis. See ¶ 36. *Tian* accomplishes this by attempting to perform facial expression analysis on pre-processed face images that are “down-sampled” from original images to generate “lower resolution images.” *Tian*, Abstract, 000002. *Tian* provides examples of the “five different resolutions from Cohn-Kanade database,” shown in Table 1:

Table 1: An example in five different resolutions from Cohn-Kanade database. The head region is cropped for display purpose only. The lower resolution images are down-sampled from the originals.



Tian, Table 1.

104. Petitioner maps *Tian*'s “down-sampling” to create “lower resolution images” for teaching the claimed reducing “the resolution of a first image.” *Petition*, 26-27. As discussed above, Petitioner maps *Tian*'s head pose estimation for teaching determining the head orientation. *Petition*, 27; see Section V.D.1. *Tian* discloses performing head pose estimation on all tested resolutions and places the findings in Table 3. *Tian*, 000002, 000005, Table 3 (providing a “summary” of results on the

effects of “different resolutions” on, among other processes, “head pose estimation”). Dr. Doermann cites this Table 3 and the summarized results for supporting the mapping of *Tian* for determining a head orientation “based on the reduced-resolution image,” as claimed. Ex. 1003, ¶¶ 121-122.

105. Petitioner states a POSITA would have been motivated to modify *Lu-577* (employing *Lu-000*’s recognition system) with *Tian*’s reduced-resolution head pose estimation because reducing the resolution “can reduce the time and resource consumption of tasks such as **face detection and head pose estimation.**” *Petition*, 10; 27 (asserting the modification “would have saved processing resources”); Ex. 1003, ¶ 122 (opining “performing the head pose estimation on the reduced-resolution image would have saved processing resources and time.”). Petitioner further states that because the “**face detection and head pose estimation**” accuracy can be maintained for “reduced-resolution images,” reducing the resolution would not “substantially reduc[e] the accuracy of these tasks.” *Petition*, 10; Ex. 1003, ¶ 69 (Dr. Doermann opining “performing face detection **or** head pose estimation on a reduced-resolution image provides savings in processing resources or time, but does not substantially reduce the accuracy of these analysis tasks”). As discussed further below, the background references applied for supporting the motivation of performing head pose estimation on reduced-resolution images to reduce resource consumption **do not** provide support this motivation. Additionally, a POSITA

reviewing *Tian* would have been discouraged from purposely reducing the resolution of a captured image.

II. References Cited by Dr. Doermann Do Not Provide Support for Reduced Resource Consumption in Head Pose Estimation

106. Dr. Doermann cites three background references (Ex. 1013-1015) to support his motivation of reducing “time and resource consumption” by performing head orientation determination on reduced-resolution images, as allegedly taught by *Tian*. Ex. 1003, ¶ 69. In my opinion, these references **do not provide support for the opinion that head pose estimation** is resource intensive and should be performed on reduced-resolution images.

107. Dr. Doermann alludes to this discrepancy himself, as he opines “tasks such as **face detection**” (not head pose estimation) “**can** be resource-consuming.” Ex. 1003, ¶ 69. However, face detection being resource-consuming **does not address whether head pose estimation is resource-intensive**. Indeed, *Tian* recognizes these two processes are distinct from one another, as *Tian* separately evaluates “face detection” from “head pose estimation.” *Tian*, 000002, 000005 (stating head pose estimation resulted in different accuracy than face detection—**“compared to face detection, head pose estimation can detect 98% faces...”**). In fact, both processes result in different facial expression accuracies. *Id.* Thus, per *Tian*, face detection is a distinct and separate step from head pose estimation.

108. *Tian* even distinguishes head detection (i.e., “detect[ing] the head”) from head pose estimation. *Tian*, 000002 (describing “[a]fter” the head is detected, the image is converted and then a neural network is “employed to estimate the head pose”). Specifically, *Tian* discloses the head is first detected, and then the separate step of head pose estimation is performed on the image. *Id.* Thus, *Tian*’s head pose estimation is **distinct from face detection** and head detection.

109. The first cited reference (Ex. 1013) states “view-based [face detection] methods...are reported to have a good detection performance but the drawback to be computationally rather expensive.” Ex. 1013, 000001; Ex. 1003, ¶ 69 (quoting Ex. 1013). Dr. Doermann relies on this quote for supporting his opinion that “**face detection** can be resource-consuming.” Ex. 1003, ¶ 69. However, Ex. 1013 does not disclose **all** face detection techniques are computationally expensive. Rather, only certain approaches (i.e., view-based methods) are computationally “expensive.” Ex. 1013, 000001. In viewing Ex. 1013, a POSITA would not have understood the claimed “determine the orientation of the head” to be “resource-consuming,” as Ex. 1013 is discussing the distinct process of **face detection**. Thus, this problem of resource consumption **would not have motivated** a POSITA to make Petitioner’s proposed modification, given the knowledge of Ex. 1013. In fact, even Dr. Doermann admits the “resource-consuming” operation is **face detection, not head pose estimation**. Ex. 1003, ¶ 69.

110. Turning to Ex. 1014, Dr. Doermann further opines “[s]ome approaches of **face detection** may require analyzing all pixels of an image,” and that “processing resources and time can be reduced” by reducing the resolution of the image. Ex. 1003, ¶ 69 (*citing* Ex. 1014, *Corcoran*, 2:43-53). Once again, Ex. 1014 may support there being a problem of resource consumption in **face detection** (*Corcoran*, 2:43-53), but does not provide support for head orientation determination having any problem of resource consumption. Instead, *Corcoran* discloses avoiding certain calculations of a “highest resolution integral image” to thereby reduce, in some instances, “**processing overhead for face detection and tracking.**” *Corcoran*, 2:43-53. Thus, *Corcoran*, at best, supports that reducing an image resolution may reduce “processing overhead” for **face detection**. *Corcoran* has no mention of determining a head orientation or estimating a head pose. *See generally*, Ex. 1014. Thus, Ex. 1014 does not remedy the deficiency of Dr. Doermann’s cited reference, Ex. 1013, in supporting his opinion that **head pose estimation** (mapped for teaching the head orientation determination) is resource intensive.

111. The last background reference Dr. Doermann cites for allegedly supporting the proposed motivation is Ex. 1015 (“*Hong*”). Ex. 1003, ¶ 69 (*citing* Ex. 1015, *Hong*). Dr. Doermann does not provide any explanation for his cite, but merely quotes *Hong*’s disclosure that “by reducing the **resolution of the saturation of the image**, computational requirements are substantially reduced, and a relatively

simple method may be used.” Ex. 1003, ¶ 69 (*citing* Ex. 1015, *Hong*, [0022]). Dr. Doermann cites this quote to allegedly support his opinion that “processing resources and time can be reduced if the resolution of the image is reduced.” *Id.* *Hong* discloses “a method for **detecting face-like regions in a colour image.**” *Hong*, Abstract. *Hong* implements this method by performing **face detection**. *Hong*, [0001] (describing the “invention” relates to “detecting a face-like region,” and that the method be used with “other methods for detecting a face in an image”). For the same reasons discussed above, a POSITA would not have been motivated by a benefit associated with **face detection** to modify *Lu-577* (as modified by *Lu-000*) to determine a head orientation based on a reduced-resolution image. *See* ¶¶ 107-110.

112. The method includes “reducing the resolution of the colour image by **averaging the saturation to form a reduced resolution image.**” *Hong*, [0009], [0011], [0020]. The saturation averaging does not reduce the computational requirements, but rather allows more efficient methods of **face detection** to be implemented. *Hong*, [0020]. The efficient face detection methods provided by implementing the saturation averaging reduces “computational” requirements. *Hong*, [0022]. Additionally, a POSITA would have understood the averaging operation to add computational processing to the face detection procedure. Namely, the steps involved in the averaging process are extra steps implemented in order to perform the more efficient face detection methods of *Hong*. Additionally, the

averaging operation would increase the blurring of the image, as taking an average of the saturation would reduce the sharpness and detail in the image. *Hong*, [0031-0032] (describing after the step of “averaging” the saturation, the reduced-resolution image comprises regions or “blobs of uniform saturation”).

113. Furthermore, specific equations and algorithms are used to implement the saturation averaging and face detection performed on the reduced-saturation images. *Hong*, [0010-0011], [0018], [0020]. Thus, if Ex. 1015 motivates any modification relating to reducing a resolution to reduce “computational requirements,” the modification would include switching to a completely different algorithm for performing face detection that works well for reduced-saturation images. However, neither Petitioner nor Dr. Doermann asserts any process in *Lu-000* being switched with a face detection method catered to reduced-saturation images when modifying to include *Tian*’s head pose estimation. Thus, for this reason alone, Ex. 1015 does not support any motivation for modifying *Lu-000* to include *Tian*’s reduced-resolution head pose estimation, such as the motivation of decreasing processing requirements and time.

114. Additionally, employing the reduced “computational requirements” motivation from *Hong* would actually **increase the amount of modification required** beyond that proposed by Petitioner and Dr. Doermann. Namely, to reap the benefit of reduced “computational requirements,” a POSITA would have

understood reconfiguration of *Tian*'s head pose estimation is required to perform the face detection method on saturation-averaged images, rather than detecting a head in an image. *Tian*, 000002. Furthermore, additional steps would be required for averaging the saturation of *Lu-000*'s images. These increased modifications may actually provide the opposite effect of increasing the processing required, rather than reducing computations.

115. Even if Dr. Doermann envisioned this extensive modification, the saturation-averaging modification does not make sense when modifying *Lu-000*'s recognition system. *Lu-000*'s extracted signatures that are compared to the various stored pose signatures are extracted from normalized and **binarized** face images. *Lu-000*, 6:40-51 (disclosing an "extracted pattern image signature" is extracted from a "binary" B-sub "face image"), 6:57-62 (describing the extraction of a signature from a binarized B-sub image is utilized "in the recognition operational mode"), 7:47-55 (disclosing extracting the signature from the binary level B-sub image and comparing it to the stored signatures to identify a match). A POSITA would have understood because the images are binarized, the signatures are **extracted from two-tone images in black and white**. Thus, a POSITA would not have looked at *Hong*'s reduction in the saturation of a **colored image** as motivating a modification to *Lu-000*'s recognition system that employs black and white images. Accordingly, Ex.

1015's reduced saturation would not have motivated a POSITA to modify *Lu-577* (as modified by *Lu-000*).

116. To summarize, none of the background references support Dr. Doermann's conclusory opinion that performing "head pose estimation on a reduced-resolution image provides savings in processing resources or time." Ex. 1003, ¶ 69. All of the references (Ex. 1013-1015) cited by Dr. Doermann are directed to **face detection**, which is not relevant to Petitioner's proposed modification of performing **head orientation determination** on a reduced-resolution image (as allegedly taught by *Tian's* head pose estimation). Thus, a POSITA would not have understood *Tian's* head pose estimation is resource intensive. Accordingly, a POSITA would not have been motivated by the supposed benefit of reduced computational requirements to modify *Lu-577's* system (employing *Lu-000's* recognition system) with *Tian's* teaching of reducing an image resolution and performing head pose estimation on the image (as Petitioner maps for teaching the claimed "determine the orientation of the head...based on the reduced-resolution image.").

III. Tian Teaches Performing Facial Analysis on Lower Resolution Images Is Undesirable

117. As discussed previously, *Tian* desires to explore the effects of various image resolutions on expression analysis. See ¶ 36. However, *Tian's* exploration is

taken with the understanding it was well-known performing facial image analysis on low resolution images poses additional challenges. Indeed, an image with low resolution does not show as many facial details as an image of higher resolution, and this loss of detail affects the ability of systems to perform successful facial image analysis. For example, Ex. 1027 is a paper that explored how image resolution affects facial image analysis, such as face recognition performance. *Face Image Resolution Versus Face Recognition Performance Based on Two Global Methods*, Ex. 1027, Abstract (describing the results “depict[ing] how face recognition rate changes with face image resolution”). Ex. 1027 found a lower recognition rate associated with lower resolution images. Ex. 1027, col. 1, par. 1 (stating “before 64 x 48 the recognition rate increases roughly with the increasing resolution”), Fig. 7 (illustrating low recognition rates for resolutions of 16x12, 32x24, and 48x36, to name a few). Similarly, the ’243 Patent itself discloses the “accuracy of the identification” of audience members via facial recognition “increase with an increase in resolution of the image data,” meaning low resolution images create low facial recognition accuracy. ’243 Patent, 2:50-56. This is why (as discussed above) the audience measurement system of the ’243 Patent designates high-resolution images for facial recognition, not low resolution images. *See*, ¶ 31. Additionally, Ex. 1024 states “there is always some low resolution at which the face image will be of

sufficiently small size that the face is unrecognizable.” *Facial Recognition Vendor Test 2000*, Ex. 1024, 000046.

118. It was well-known that analyzing low resolution images also negatively affects geometric methods of determining head pose from facial features in an image (another exemplary facial image analysis process). *Head Pose Estimation in Computer Vision: A Survey*, Ex. 1020, col. 1, par. 2 (“Far-field imagery,” in the context of handling outlying or missing feature detections, “is **problematic since the resolution can make it difficult or impossible to precisely determine the feature locations.**”). Thus, it was well-established that low resolution images affect the performance of various facial image analysis processes in negative ways, and thus low resolution images are undesirable for use in facial image analysis.

119. *Tian*, understanding that low resolution images are less desirable images for facial image analysis than higher resolution images, explores the effects of various resolutions (including low resolutions) on facial expression analysis because it is recognized that, in some real-world “environments, the face image is often in lower resolution.” *Tian*, Abstract. In other words, *Tian* is exploring the facial expression analysis performance when “in real applications, the face region is often in lower resolution.” *Tian*, 000002. When viewing *Tian*, a POSITA would be dissuaded from unnecessarily increasing the difficulty (and potentially decreasing

the success) of facial expression analysis by purposely reducing the resolution of an image prior to performing the expression analysis steps.

120. Petitioner ignores the above issues with applying facial image analysis on low resolution images. Rather, Petitioner and Dr. Doermann point to results from *Tian* that show, at certain resolutions, performing head pose estimation “**can** detect 98% faces,” and from these results, asserts determining a head orientation based on a reduced-resolution image can be performed “without substantially reducing the accuracy of the determination.” *Petition*, 27; Ex. 1003, ¶¶ 68-69. Worse still, neither Petitioner nor Dr. Doermann addresses the aspects of *Tian* that would actually **discourage** a POSITA from purposely and unnecessarily reducing the resolution of an image.

121. *Tian* discloses “**low resolution input images**” is one factor “**that complicate[s] facial expression analysis.**” *Tian*, 000002, 000006. Additionally, “real environments” in which “lower resolution” images are captured make real-life facial expression analysis “**more difficult.**” *Tian*, Abstract (stating “in real environments, the face image is often in lower resolution”), 000002 (stating “Real-life facial expression analysis,” such as analysis that would be performed on images taken in real environments, “is **much more difficult than the posed actions studied predominantly to date.**”), *see also*, 000006. Thus, *Tian* repeatedly states low resolution images make facial expression analysis more difficult. To combat this

difficulty, *Tian* even suggests techniques that can be used to increase image resolution to improve expression analysis accuracy. *Tian*, 000006 (stating “multiple cameras” can be “expected to provide improved accuracy;” “[a]ctive cameras can be used to **acquire high resolution face images;**” and “techniques of **super-resolution can be used to get high resolution images** from multiple low resolution images”). Thus, though *Tian* does find certain steps of facial expression analysis to be able to be performed on low resolution images (000005), *Tian* **still** recognizes the **problems caused by low resolution images** and even discloses methods to address these problems that increase the image resolution.

122. *Tian*’s repeated discussion of low resolution images making facial expression analysis more difficult and his inclusion of various techniques to improve accuracy by increasing resolution would have dissuaded a POSITA from intentionally lowering the resolution of an image to be analyzed and, accordingly, increasing the difficulty of facial expression analysis. In other words, though *Tian* explores methods for mitigating the difficulties associated with analyzing low resolution images when **only low resolution images are available**, a POSITA would be discouraged from taking a high resolution image and reducing the resolution of the image. Furthermore, because performing various facial image analysis processes on low resolution images was known to negatively affect

performance and accuracy, a POSITA would have been dissuaded from purposely reducing a resolution of an image for facial analysis processes.

123. Because *Tian* discusses low resolution images making facial analysis processes more difficult, a POSITA would be discouraged from performing a step that reduces a resolution of the image prior to performing facial analysis processes. Thus, when viewing *Tian*, a POSITA would have been discouraged from modifying *Lu-577*'s system (otherwise modified by *Lu-000*'s teachings) to reduce the resolution of the media exposure environment image prior to determining head orientation. Accordingly, a POSITA would not have been motivated to make such a modification, when viewing *Tian* and understanding low resolution images negatively affect facial analysis processes.

5. Opinions Regarding Petitioner's General Allegations of Obviousness

124. As part of the obviousness argument for modifying *Lu-577* (otherwise employing *Lu-000*'s recognition system) with *Tian*'s "techniques," Petitioner provides general assertions. For example, because *Lu-577* and *Tian* are allegedly "similar," Petitioner states a POSITA would have found it obvious to combine them. *Petition*, 8 (asserting *Tian* "similarly teaches techniques for image processing and facial analysis"), ("A POSITA would have found it obvious to combine *Lu-000* and *Tian* because **they are similarly directed to image processing and facial**

analysis.”). As another example, Petitioner states “[a] POSITA would have had a reasonable expectation of success” when combining *Tian*’s teachings (e.g., head pose estimation/classification) into *Lu-577* “given the **overlap between Lu-000 and Tian.**” *Petition*, 11. Dr. Doermann provides the same general reasoning, opining *Lu-000* and *Tian* are “similar” and directed to “similar functionalities.” Ex. 1003, ¶¶ 57-59, 71-72.

125. In my opinion, *Tian* is **not** “similar” to *Lu-000* or *Lu-577*. As discussed above, *Tian* explores the effects of various image resolutions on the **facial expression analysis process**. See ¶ 36. Contrastingly, *Lu-577*, as modified by *Lu-000*, “employ[s] known head location and **face recognition software**” to identify “viewing persons” in the audience. *Lu-577*, [0039]. The head location and face recognition software of *Lu-000* is implemented within an image **recognition** system that identifies members of a “viewing audience” by recognizing the face of each audience member. *Lu-000*, Abstract, 7:25-55 (generally describing comparing an image feature signature with each stored face signature in a “face library” to identify the audience member).

126. *Tian* is **not** performing any **facial recognition** to determine the identity of a person, but rather is recognizing facial expressions. *Tian*, Abstract, 000002. Indeed, recognizing a person’s identity is quite irrelevant to *Tian*’s purpose of recognizing expressions. You do not need to know an audience member is Suzy if

you merely want to determine if the person is happy or sad. Thus, *Tian* is not “similar” to *Lu-000*’s recognition system, *Lu-577*’s audience measurement system, or the combined system.

127. I note Dr. Doermann seems to conflate the “similarity” of *Lu-000* and *Tian* by opining the references teach techniques for “image processing and facial analysis.” Ex. 1003, ¶¶ 57-58. However, this rough clumping of the references ignores the clear disparities between the techniques themselves and what the techniques are used for. Namely, *Tian*’s image processing techniques are implemented for **expression analysis**, while *Lu-000*’s image processing is employed for **facial recognition**. A POSITA would have understood the problems associated with these processes are distinct and have distinct constraints, and thus would not have found *Tian*’s facial expression analysis techniques applicable to *Lu-000*’s face recognition system. See ¶¶ 71-72. Indeed, as discussed above, *Tian*’s expression analysis does not identify the person making the recognized expression. See ¶ 126. Compressing the references into “facial analysis” or “image processing” makes them seem “similar.” But, on closer inspection, *Tian* is not “similar” to *Lu-000* or *Lu-577*.

128. Thus, it is my opinion a POSITA would not have found *Tian* to be similar to *Lu-577* or *Lu-000*. Namely, because *Tian* **recognizes expressions without identifying people** making the expressions, while *Lu-577* (as modified by *Lu-000*) **recognizes faces to identify people**. For these reasons, Petitioner’s and Dr.

Doermann’s general allegations that the references are “similar” would not have made the combination obvious to a POSITA.

129. To summarize, a POSITA would not have been motivated or found it obvious to modify *Lu-577* (employing *Lu-000*’s recognition system) to perform head orientation determination, let alone performing *Tian*’s head pose classification on reduced resolution images. This is because *Lu-000* already recognizes various orientations of the head to successfully identify people, thereby already using “suitable” images for facial recognition. See ¶¶ 56-60. Any modification to *Lu-000* to determine head orientation would thus be superfluous, provide no added benefit, and add meaningless complexity to the recognition system. See ¶¶ 61-65. Modifying *Lu-000* with *Tian*’s head pose classification would actually decrease the success of *Lu-000*’s facial recognition system. See ¶¶ 66-77. Additionally, *Tian* is not even similar to *Lu-000*, because *Tian* does not recognize faces, but recognizes expressions on faces without identifying the face. See ¶¶ 124-128.

VI. OPINIONS ON CHALLENGED GROUND 2

A. OVERVIEW OF THE CITED REFERENCES AND PETITIONER’S PROPOSED COMBINATION

1. *Nielsen-372*’s Local Metering System

130. *Nielsen-372* discloses “a metering system” that provides “viewing and metering information for program content” presented, e.g., by a TV. *Nielsen-372*, EX. 1006, [0036]. The metering system includes a “metering unit 124” and a

“display device ON/OFF detector 128.” *Id.* The display device 120 the “ON/OFF detector” may be detecting is, for example, a television. *Nielsen-372*, [0040]. The metering unit 124 “generate[s] viewing records and other viewing information” by logging and processing data. *Nielsen-372*, [0040]. The display device ON/OFF detector 128 determines “whether the display device 120” is “operating in an ON (active state) or an OFF (inactive) state.” *Nielsen-372*, [0043]. Thus, the ON/OFF device of the metering system functions to determine whether the TV (exemplary display device 120) is ON or OFF.

131. *Nielsen-372* discloses the ON/OFF device 128 is coupled to the metering unit 124 and **“may be used to more accurately process the viewing information and viewing records determined by the metering unit 124.”** *Nielsen-372*, [0043]. *Nielsen-372* explains the accuracy of processing increases because the ON/OFF device may detect when the TV is turned off and thus no one is watching TV, even when the set top box is “inadvertently” left ON. *Id.* In this scenario, the set top box may still output program content, even though no one is watching the TV. *Id.* The ON/OFF device provides information so the “central facility” receiving the output program content can correctly identify the content as not actually being watched (i.e., because the display device is turned off). *Id.* Thus, the ON/OFF device augments the viewing records to more accurately determine

whether content output by the set top box is “actually presented by the display device 120.” *Id.*

132. The metering system also includes a separate “people meter 162” that captures “information about the audience.” *Nielsen-372*, [0048]. Namely, the people meter 162 receives information from a control device comprising input keys. *Id.* The people meter prompts audience members to indicate they are present by pressing the appropriate input key. *Id.*

133. The ON/OFF device 128 comprises several sensors 144. *Nielsen-372*, [0050]. One such sensor is a video sensor 208, such as, e.g., a “camera.” *Id.* The video sensor 208 is “positioned to detect the display area of the display device 120.” *Id.* In other words, the video sensor 208 is positioned to face the TV screen. That way, the ON/OFF device 128 may process the video signals from the video sensor 208 to determine if the display is on or off, among processing other sensor signals. *Nielsen-372*, [0052]. Specifically, the ON/OFF device 128 processes the output video signals to “determine characteristics of the input video signal 234” that “may be used to ascertain whether the information presenting device monitored by the display device ON/OFF detector” is “turned ON and operating in an active state.” *Id.* These characteristics are determined by detecting that extracted regions of video frames of the display device “differ” from successive video frames, and accordingly detecting the TV is ON and active. *Nielsen-372*, [0116]. The differences indicate a

changing video presentation on the TV screen, and “thus, that the monitored information presenting device is ON.” *Id.* If no changes are detected between the extracted regions and successive video frames, the ON/OFF device determines the display device 120 to be OFF. *Id.*

134. In Petitioner’s modification of *Nielsen-372*, Petitioner proposes remodeling the video sensor 208 of the ON/OFF device 128 for performing audience identification. *Petition*, 44 (asserting a POSITA would have found it obvious to modify *Nielsen-372*, such that “audience members are identified by face detection and recognition” by implementing *Nielsen-372*’s “needed hardware”), Ex. 1003, ¶ 200 (Dr. Doermann identifying *Nielsen-372*’s video sensor 208 as the needed hardware for identifying audience members). However, this modification would render the ON/OFF device completely inoperable for its intended purpose of detecting when the **display** is turned ON or OFF. *Nielsen-372*’s ON/OFF device is an important feature of the metering system, as the ON/OFF device may be used in and benefit several “applications,” including parenting tools, media monitoring systems, building monitoring systems, etc. *Nielsen-372*, [0002-0006]. As an example, building owners would benefit from building monitoring systems including the disclosed ON/OFF device such that the state of various electronic appliances in the building can be detected, even when the building owner is remote from the building. *Nielsen-372*, [0004]. Thus, understanding the importance of the

ON/OFF device for its intended function of detecting when a display screen is turned ON/OFF, a POSITA would have understood the ON/OFF device to be an integral component of *Nielsen-372*'s metering system.

135. However, for Petitioner's proposed modification of the ON/OFF device's video sensor (e.g., camera), the video sensor is oriented towards the audience rather than "positioned" toward the display device (as taught in *Nielsen-372*). *Nielsen-372*, [0050]. With this modification, the video sensor 208 would not be operable to detect when the display device is ON or OFF. This would render the ON/OFF device and, in turn, the metering system of *Nielsen-372* useless for its stated purpose of detecting the ON/OFF state of the display device. *Nielsen-372*, [0002].

2. *Steinberg*'s Digital Image Processing System

136. *Steinberg* discloses a system applying "digital image processing" using facial recognition techniques to identify people within images that "may be part of a personal or other image collection." *Steinberg*, [0090], [0092-0094]. The system uses a face detection module to determine if a face is in an image. *Steinberg*, [0118]. Once a face is detected, the image is passed to a face normalization module that analyzes, among other things, the pose of the faces in the image. *Steinberg*, [0120], [0176]. The system then compares the image with various stored images within face classes to identify the face and place the image in the correct class for storage.

Steinberg, [0156-0157], [0159] (describing comparing the image with “recorded” known faceprints stored in various “Face class[es]” and, once matched, storing the image with a link to a “database profile” for the matched person). Thus, *Steinberg* discloses an image processing system that performs facial recognition on faces within captured images to accurately store the image in the correct face class within an image collection. *Steinberg*, [0090], [0003] (field of invention being related to “grouping and classification of images in a database or image collection” based on the identification of faces in the images).

137. *Steinberg* expressly notes the primary “context” of the image processing system to be for image “collections of consumer digital images which would be generated by a typical user of a digital camera.” *Steinberg*, [0010]. This “context” in which the image processing system is applied is not pertinent to audience measurement systems, and in fact has no link to measuring audience data. *See* ¶¶ 159-160.

3. Petitioner’s Proposed Modification of *Nielsen-372* With *Steinberg*

138. Petitioner’s modification incorporates *Steinberg*’s facial recognition into *Nielsen-372*’s metering system. *Petition*, 44. Specifically, as discussed above, Petitioner proposes utilizing *Nielsen-372*’s “**needed hardware** (e.g., a computer and a camera for implementing *Steinberg*’s approach.” *Petition*, 44; *see also*, ¶¶

134-135. Dr. Doermann also specifically applies *Steinberg's* approach on *Nielsen-372's* own hardware, such as the video sensor 208 of the ON/OFF device. Ex. 1003, ¶¶ 199-200. Thus, Petitioner's combination explicitly modifies *Nielsen-372's* ON/OFF device to implement *Steinberg's* facial recognition. I have been asked by counsel to provide my opinions on the modification of *Nielsen-372's* own hardware (e.g., video sensor 208) for implementing *Steinberg's* facial recognition.

B. OPINIONS REGARDING PETITIONER'S MODIFICATION OF *NIELSEN-372* WITH *STEINBERG*

139. I have been asked to provide my opinions on whether a POSITA would have found it obvious and been motivated to perform Petitioner's proposed modification. Petitioner's Ground 2 requires the combination of *Nielsen-372* with *Steinberg*. *Petition*, 45-49, 53-70. Additionally, Petitioner's Challenged Dependent Claims of Ground 3 require the combination of *Nielsen-372* with *Steinberg* (otherwise modified by *Tian*). *Petition*, 71-88. As discussed further below, because Petitioner relies on *Nielsen-372's* camera in the ON/OFF device for capturing images and identifying audience members in those images, the camera must be oriented toward the audience members. Yet, *Nielsen-372's* camera is oriented toward the display device that the audience is watching, so the metering system can accurately determine media content information. Thus, Petitioner's proposed modification would frustrate the purpose of the metering system of *Nielsen-372* by

rendering it unable to determine such media content information and would change the principle of operation of the metering system.

1. Petitioner's Modification Would Frustrate the Purpose of *Nielsen-372's* Metering System

140. As discussed above, the camera (i.e., video sensor 208) of *Nielsen-372* is positioned toward the display device so the ON/OFF device camera can detect the display area. *See* ¶ 133. It is my opinion that a camera explicitly positioned to detect the display **is not positioned to detect the audience viewing such display**. Indeed, for people to form an “audience” watching a TV, they themselves must be facing the TV screen. Thus, a camera positioned to face the TV screen would not **also somehow** be positioned to face the audience. As *Nielsen-372's* camera is positioned to face the TV screen, at best, the camera would be positioned to detect actor's faces on the TV program to identify the actors. However, these faces would not belong to “audience members.” *Petition*, 44 (stating the combination would identify “audience members”).

141. The stated function and purpose of *Nielsen's* video sensor 208 (the “camera” in the proposed combination) and the ON/OFF device 128, as a whole, is to determine whether the display showing the TV program is powered on or turned off. *See* ¶¶ 130-131, 133. This is why it is important that the camera face the display, **not the audience watching the display**. Thus, a POSITA would understand if the

camera were positioned to take images of audience members, the camera would be useless for its purpose of detecting whether the display is ON or OFF.

142. Additionally, when positioned toward the display, a POSITA would have understood the video sensor 208 of the ON/OFF device 128 would not take images of the audience members, for the reasons discussed above. *See* ¶ 140. *Nielsen-372*'s metering system already includes a separate device (i.e., people meter) for performing the function of identifying audience members. *See* ¶ 132. Thus, the camera of *Nielsen-372* does not need to take images of the audience for identifying the audience members, but **does need to face the display screen to perform the intended function of the ON/OFF device.**

143. Yet, this is exactly how the proposed modification requires *Nielsen-372*'s video sensor 208 to be positioned: **facing the audience**. The modification requires this video sensor 208 orientation because the video sensor is being used for performing *Steinberg*'s facial recognition to identify audience members. *Petition*, 44 (the combination using *Nielsen-372*'s camera "such that audience members are identified by face detection and recognition."). Due to the placement of audience members in relation to the display screen (i.e., TV screen), in this modification, the video sensor 208 would no longer be positioned "to detect" the ON/OFF state of the display device. *See* ¶ 140; *Nielsen-372*, [0050].

144. Thus, in the proposed modification, the video sensor 208 **would not be able to function for its purpose within the ON/OFF device 128 of detecting when the display screen is turned ON or shut OFF.** *Nielsen-372*, [0116] (describing the video frames being analyzed and compared to determine whether the presenting device is ON or OFF). As discussed above, this purpose of the video sensor 208 (i.e., detecting when the display is ON or OFF) is important to the metering system because it provides the benefit of increasing the accuracy of recorded viewing records. *See* ¶ 131; *Nielsen-372*, [0043]. A POSITA would understand this importance, and thus would be dissuaded from modifying *Nielsen-372*'s video sensor 208 in a way that would detrimentally affect the functioning of the ON/OFF device for providing such a benefit.

145. The proposed modification to *Nielsen-372* requiring the camera face the audience instead of the display would thereby significantly and fundamentally change how *Nielsen-372*'s metering system operates. Namely, making the proposed modification would fundamentally affect the ability of the metering system to detect the ON/OFF state of the display screen. This would be a significant change, as the accuracy of the viewing records generated by the metering system depends on detecting whether the display is ON or OFF. Indeed, performing the proposed modification would render the ON/OFF device unable to detect when the display screen was ON or OFF. Because this detection increases the accuracy of the viewing

records, performing the modification would also render the entire metering system unsatisfactory for the intended purpose of providing accurate viewing information.

2. Petitioner's Proposed Modification Would Change the Principle of Operations of *Nielsen-372*'s System

146. Modifying *Nielsen-372*'s video sensor 208 (i.e., camera) in the ON/OFF device to face the audience and capture images to automatically identify audience members via *Steinberg*'s facial recognition techniques would change the principle of operations of the metering system in two significant ways. Firstly, as discussed above, the camera would no longer take video images of the display screen. *See* ¶ 140, 142-143. Thus, the metering system would no longer be able to rely on video frames from the video sensor 208 to determine the ON/OFF state of the display. Rather, the metering system would have to rely on other techniques for determining whether the display was ON or OFF.

147. Though *Nielsen-372* discloses other sensors 144 within the ON/OFF device 128 for detecting whether the display device is turned ON or OFF, these other sensors may not be as accurate as the video sensor 208. *Nielsen-372*, [0050]. For example, the audio sensor 204 of the ON/OFF device 128 would generate false positives by picking up various sounds within the room, including, e.g., audience members talking. *Id.* (describing the audio sensor 204 as “one or more microphones”). Thus, on its own, the audio sensor 204 would not provide accurate

information as to whether the sound came from an ON display screen. Additionally, the emission sensor 212 could pick up current, temperature, or EM remnants from the display that linger even after the TV has been turned off. *Nielsen-372*, [0050]. In other words, the other sensors, both together and alone, cannot accurately detect **when a picture is presented on a TV screen and moving (to indicate watching a show) to accurately detect when the display screen is ON and actively presenting program content.** And, accordingly, these sensors are less accurate in detecting whether the display device is ON and presenting content or turned OFF.

148. *Nielsen-372*'s ON/OFF device includes all of these various sensors to provide redundancies and increased accuracy in detecting the state of the display device. Namely, because each sensor has its own error associated therewith (examples provided above), including multiple sensors increases the probability that the detection outcome is accurate. *Nielsen-372* describes combining the outputs from all of the sensors when reaching an ON/OFF detection decision. *Nielsen-372*, [0133]. Also, the addition of multiple sensor types allows a user to curate the ON/OFF device to their specific needs, by weighting some sensor outputs higher than others in making the final ON/OFF determination. *Nielsen-372*, [0063], [0067], [0074], [0133], Fig. 2.

149. Petitioner's proposed modification to the video sensor 208 of the ON/OFF device would stop the video sensor 208 from detecting the state of the

display screen, and thus entirely remove a sensor class from *Nielsen-372's* combination of sensors. As discussed above, the video sensor 208 is the only sensor capable of detecting the picture on the TV screen and whether the picture is moving. *See* ¶ 147. Namely, the video sensor is the only visual-based sensor. This makes the video sensor 208 an important class of sensor utilized, as it allows the ON/OFF device to determine whether the display device is presenting changing content (e.g., TV show) to the audience. As discussed above, the other sensors may still read signals indicating the TV is on when the screen has actually been turned OFF (e.g., audio sensor picks up people talking, emission sensor picks up remnant currents, EM, or temperature after the TV is turned OFF). *See* ¶ 147. Thus, if you take away the video sensor 208, the metering system would change from one that can detect when a TV is presenting content to an audience (i.e., active and ON) to one that cannot tell the difference between a TV that is blank (i.e., OFF or inactive) and when a screen is actually presenting a TV program.

150. Regarding the second significant change to the metering system, the proposed combination would change the system from one that is **voluntary** into one that is **automatic** and tracks audience members without their input (and potentially without their knowledge or consent). Petitioner's proposed combination is motivated by substituting *Nielsen-372's* prompting of the audience members to identify themselves (via the input keys of the people meter) with an automatic identification

using *Steinberg's* facial recognition. *Petition*, 44; Ex. 1003, ¶ 201. Dr. Doermann himself acknowledges situations in which an audience member “refuses to provide the input showing his or her presence.” Ex. 1003, ¶ 201. This audience member may not want to be tracked, and thus, appreciates the ability to opt out of being identified. When you take away the voluntary nature of *Nielsen-372*, you fundamentally change the way in which the system operates, and audience members may be dissuaded from having such an automatic TV tracking system in their living rooms.

151. Additionally, *Nielsen-372's* non-modified metering system provides users with the ability to choose *when* they identify themselves. For example, audience members may not want to be identified when viewing certain guilty pleasure shows or when viewing TV content while they are supposed to be doing something else (e.g., a household chore). For these reasons, audience members would appreciate the ability to choose when to provide viewing data and when not to.

152. Furthermore, some viewers may be uncomfortable with automatic image-taking as a method of audience identification. Viewers appreciate the comfort provided of watching TV in their own home, under their own conditions. Automatically capturing images may be a dissuading factor to viewers that otherwise would implement *Nielsen-372's* system, as viewers may feel the need to look their best for the mere task of watching TV on their couch. Additionally, audience

members may not want to provide an image of their face to a household statistical system, as they may be uncomfortable with the fact that their face images are stored and associated with their names. Audience members would appreciate that this automatic face recognition system could be hacked, and then their identification and face image could potentially be stolen.

153. For these reasons, a POSITA would have understood changing *Nielsen-372's* metering system from **voluntary** identification to **mandatory** identification would fundamentally change the principle of how the metering system operates. Even further, this change would make the system less desirable for audience members that appreciate the privacy and control afforded through voluntary identification.

3. The Patent Office Found Similar Modifications To Be Non-obvious

154. I have been asked to provide my opinion on the prior art cited in the File History of the '243 Patent. In the Notice of Allowance for the '243 Patent, the Examiner cited U.S. Patent Application Publication No. 2011/0164188, "D1," as the closest prior art. *File History of the '243 Patent*, Ex. 1002, 293. As the Examiner notes, D1 discloses a camera "configured to **capture images of the viewing environment**" in order to estimate a "head orientation of viewer." Ex. 1002, 293-294. The images of the "viewing environment" include images of the "**screen**

assembly or other objects or points of interest normally viewable by a viewer.”

Ex. 1002, 294. Thus, the reference cited by Examiner is similar to *Nielsen-372's* camera, as both cameras image the “screen” being viewed by the viewer.

155. The Examiner goes on to state the “prior art of record” (including D1), alone, “fails to teach or suggest the determination of audience identification information based on a match of a head to a known person.” Ex. 1002, 294. The Examiner further indicates even if a reference was found that actually takes images of the audience themselves (i.e., of the “media exposure environment”), a POSITA would **not have found it obvious to substitute D1's objects of interest (i.e., objects displayed on the screen) with the audience in the media exposure environment.** Ex. 1002, 294.

156. Petitioner's proposed combination is attempting to make this exact modification; namely, modifying the ON/OFF camera that takes video frames of the **screen** to instead take images of the **audience members**. Thus, for similar reasons articulated by the Examiner, a POSITA would not have found it obvious to modify *Nielsen-372's* camera from imaging the “screen” viewable by the viewers to instead image the audience members themselves.

VII. OPINIONS ON CHALLENGED GROUND 3

A. OPINIONS REGARDING WHETHER THE COMBINATION OF *NIELSEN-372*, *STEINBERG*, AND *TIAN* RENDER OBVIOUS CLAIMS 4-6, 11-14, AND 18-20

157. I note Petitioner essentially puts forth the same obviousness arguments used within Grounds 1 and 2 for Ground 3 combining *Nielsen-372*, *Steinberg*, and *Tian*. Compare *Petition*, 8-11 with 71-73 (providing the same obviousness arguments for *Tian*); see also *Petition*, 71 (using the same arguments from Ground 2 (“§§ VI.B.1-2”) for showing obviousness of “*Nielsen-372* and *Steinberg*”). Thus, it is my opinion Ground 3 fails to render obvious claims 4-6, 11-14, and 18-20 for the same reasons noted above for Grounds 1 and 2. See *Section V-VI*.

VIII. OPINIONS REGARDING WHETHER *TIAN* AND *STEINBERG* ARE IN THE SAME FIELD OF ENDEAVOR OR SOLVE THE SAME PROBLEM AS THE ’243 PATENT

158. I have been asked to provide my opinions on whether *Tian* and *Steinberg* are in the same field of endeavor or solve the same problem as that disclosed in the ’243 Patent. I note Petitioner applies either *Tian* or *Steinberg* as secondary references in all Challenged Grounds. *Petition*, 3. For the reasons discussed below, it is my opinion neither *Tian* nor *Steinberg* are: (1) in the same field of endeavor or (2) solve the same problem as the ’243 Patent. The ’243 Patent is solving problems in the field of “audience measurement,” and neither *Tian* nor *Steinberg* have anything to do with audience measurement.

159. The '243 Patent discloses the “field” of the invention to “relate[] generally to audience measurement and, more particularly, to methods and apparatus to capture images.” *'243 Patent*, 1:36-39. The description of the invention further indicates the field is “audience measurement,” disclosing embodiments of an “audience measurement device.” *'243 Patent*, 4:52-56; *see also, generally, the '243 Patent*. Furthermore, the claims themselves recite an “audience measurement system” (claim 1 and 9) and a “method for obtaining exposure data for a media exposure environment” (claim 16). *'243 Patent*, claims 1, 9, 16.

160. Contrastingly, *Steinberg's* “Field of the Invention” is “digital image processing, **particularly to the field of automatic or semiautomatic grouping and classification of images in a database or image collection based on the occurrence of faces in the images and the identification and classification of such faces.**” *Steinberg*, [0002-0003]. *Steinberg's* primary context of applying the image processing system is for classifying images within stored image collections, with these images being ones taken by a user of a digital camera. *See* ¶ 137. Nowhere does *Steinberg* mention applying the image processing system for audience measurement. *See, generally, Steinberg*. Thus, *Steinberg* is not within the same field of endeavor as the '243 Patent. Accordingly, a POSITA would not have looked to *Steinberg's* image processing system that classifies stored image collections for combining with the primary reference *Nielsen-372*.

161. *Tian* is clear in its field of endeavor and its problem to solve, disclosing “[t]his paper explores **the effects of different image resolutions for each step of facial expression analysis**” and exploring such effects on “automatic facial expression analysis (AFEA) systems.” *Tian*, Abstract. Thus, as discussed previously, *Tian* explores the effects of different image resolutions on the performance of AFEA systems to determine how resolutions affect facial expression analysis. See ¶ 36. *Tian* has no mention of an audience or performing audience measurement. See, generally, *Tian*. Thus, *Tian* is not within the same field of endeavor as the ’243 Patent. Accordingly, a POSITA would not have looked to *Tian*’s facial expression analysis techniques and combined them with the primary reference *Lu-577*.

162. Petitioner claims both *Tian* and *Steinberg* share the same field of endeavor of the ’243 Patent, asserting the field of endeavor is “media analysis.” *Petition*, 3, 5-6; Ex. 1003, ¶¶ 33-36, 42-45. However, this interpretation blends two completely distinct fields; that of (1) media **exposure** analysis; and (2) image processing. The media being “analyzed” in “media exposure analysis” is explicitly being exposed to an audience, thus specifically analyzing media an audience is watching. ’243 Patent, 1:43-64 (generally describing “audience measurement” as implementing analysis of “media content” being exposed to an audience). This media exposure analysis is implemented for “audience measurement.” *Id.*

Contrastingly, image processing generally relates to processing any images, and **does not require the processing to be performed on media being watched by an audience.** Notably, *Tian's* field of facial expression analysis and *Steinberg's* field of classifying collected images via facial recognition fall under the distinct field of image processing. Thus, because image processing is a field distinct from audience measurement, neither of the secondary references relied upon by Petitioner are directed to the field of endeavor of the '243 Patent.

163. Additionally, it is my opinion that neither *Tian* nor *Steinberg* are reasonably pertinent to the problem being solved in the '243 Patent. This is because the problem being solved in the '243 Patent is one found in audience measurement. Namely, the '243 Patent is solving a problem caused by capturing images of audience members in an **“audience measurement system,”** where the dark environment makes frequent illumination “annoy[ing]” to audience members, among causing other disadvantages. '243 Patent, 2:57-3:35. For example, the '243 Patent discloses frequent illumination annoys “members of a panel that are exposed to the illumination source **while in the presence of the audience measurement system.**” '243 Patent, 3:12-16. Furthermore, red glow emitted from an IR LED illumination source in an audience measurement system may annoy audience members because the red glow “faces the panel members **while the panel members are, for example, watching a television.**” '243 Patent, 3:16-24. This annoyance

may prove “detrimental to an ability of the **audience measurement system** to maintain persons in the panel **and/or to collect as much data as possible.**” ’243 Patent, 3:21-24.

164. Contrastingly, *Tian* is solving a problem known to occur in **facial expression analysis**. Namely, *Tian* investigates how different resolutions affect facial expression analysis to determine the lowest resolution where you can still accurately recognize an expression. *Tian*, Abstract, 000005-6. An inventor solving the ’243 Patent’s problems of audience measurement would not have logically looked to a reference investigating the minimum resolution applicable for successful facial expression analysis. In other words, *Tian* would not have logically come to the attention of an inventor concerned with problems in audience measurement.

165. Similarly, *Steinberg* would not have logically come to the attention of an inventor concerned with audience measurement problems. Namely, because *Steinberg* is solving problems related to classifying images within an image database. *Steinberg*, [0006] (describing utilizing “face recognition” for “grouping, classification, management, presentation and access to collections of digital images in databases” would be “desirable”), [0010]. Thus, a reference, such as *Steinberg*, that is broadly related to classifying collections of images from consumer cameras would also not have logically come to the attention of an inventor solving audience measurement problems.

IX. CONCLUSION

166. I declare that all statements made herein of my 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 so made are punishable by fine or imprisonment, or both, under Section 1001 and Title 18 of the United States Code.

Dated: October 10, 2023

By: 
Eli Saber