

IN THE 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.

Case No. IPR2023-01014

U.S. Patent 11,470,243

Declaration of Dr. David Scott Doermann

TABLE OF CONTENTS

	<u>Page</u>
I. OVERVIEW OF THE TECHNOLOGY.....	6
A. Television Audience Measurement.....	6
B. Audio Content Recognition.....	8
C. Image Processing and Facial Analysis.....	9
II. THE '243 PATENT.....	12
A. Overview	12
B. Prosecution History	15
C. Priority Date	16
D. Challenged Claims	16
III. SUMMARY OF THE PRIOR ART.....	17
A. Invalidity Grounds.....	17
B. U.S. Patent Application Publication No. 2002/0059577 (“Lu-577”) (Exhibit 1004)	18
C. <i>Evaluation of Face Resolution for Expression Analysis</i> (“Tian”) (Exhibit 1005).....	21
D. U.S. Patent Application Publication No. 2010/0274372 (“Nielsen-372”) (Exhibit 1006).....	22
E. U.S. Patent Application Publication No. 2010/0066822 (“Steinberg”) (Exhibit 1007).....	24
IV. LEVEL OF ORDINARY SKILL IN THE ART.....	25
V. CLAIM CONSTRUCTION	27
A. Media Exposure Environment.....	27
VI. INVALIDITY	29
A. Ground 1: Claims 1, 4-6, 8, 9, 11-14, 16, and 18-20 are Rendered Obvious in View of Lu-577 and Tian.....	29
1. Combination of Lu-577 and Tian	29
2. Claim 1	41
3. Claim 4.....	73

4.	Claim 5	78
5.	Claim 6	84
6.	Claim 8	86
7.	Claim 9	89
8.	Claim 11	99
9.	Claim 12	101
10.	Claim 13	103
11.	Claim 14	104
12.	Claim 16	105
13.	Claim 18	112
14.	Claim 19	113
15.	Claim 20	114
B.	Ground 2: Claims 1, 9, and 16 are Rendered Obvious in View of Nielsen-372 and Steinberg	115
1.	Combination of Nielsen-372 and Steinberg	115
2.	Claim 1	121
3.	Claim 9	157
4.	Claim 16	165
C.	Ground 3: Claims 1, 4-6, 8, 9, 11-14, 16, and 18-20 are Rendered Obvious in View of Nielsen-372, Steinberg, and Tian	172
1.	Combination of Nielsen-372, Steinberg, and Tian	172
2.	Claim 1	179
3.	Claim 4	183
4.	Claim 5	186
5.	Claim 6	191
6.	Claim 8	194
7.	Claim 9	197
8.	Claim 11	201
9.	Claim 12	202
10.	Claim 13	204

11. Claim 14.....	205
12. Claim 16.....	206
13. Claim 18.....	210
14. Claim 19.....	211
15. Claim 20.....	212
VII. BACKGROUND AND QUALIFICATIONS.....	213
A. Compensation.....	217
B. Materials and Other Information Considered	217
VIII. UNDERSTANDING OF THE LAW	217
IX. RESERVATION OF RIGHTS.....	217
APPENDIX 1: CURRICULUM VITAE	
APPENDIX 2: MATERIALS CONSIDERED	
APPENDIX 3: CHALLENGED CLAIMS	
APPENDIX 4: UNDERSTANDING OF THE LAW	

I, David Scott Doermann, declare as follows:

1. My name is David Scott Doermann. I am currently a Professor of Empire Innovation in the Department of Computer Science and Engineering at the University at Buffalo. I am also the inaugural director of the University at Buffalo Institute for Artificial Intelligence and Data Science. I received my B.Sc. degree in Computer Science and Mathematics from the Bloomsburg University of Pennsylvania in 1987, and I received my M.Sc. degree in computer science from the University of Maryland College Park, Maryland, in 1989. I continued my studies in the Computer Vision Laboratory at the University of Maryland, where I earned a Ph.D. degree in Computer Science in 1993. My Ph.D. thesis focused on applying image processing and computer vision techniques.

2. Over the past decade, my research has expanded to include the general field of artificial intelligence, applied primarily to computer vision. I have published extensively on image quality estimation, text detection in the wild, machine learning applications for front-end and edge computing, and cutting-edge applications in evolutionary computing.

3. I have over 65 refereed journal publications and over 225 refereed conference papers. In 2002, I received an honorary Doctor of Technology Sciences from the University of Oulu in Finland for my contributions to digital media processing and document image analysis research. I am a Fellow of the

Institute of Electrical and Electronics Engineers (IEEE) and the International Association of Pattern Recognition (IAPR).

4. I have prepared this report as an expert witness retained by TVision Insights, Inc. In this report I give my opinions as to whether certain claims of U.S. Patent No. 11,470,243 (“the ’243 patent”) are invalid. I provide technical bases for these opinions as appropriate.

5. This report contains statements of my opinions formed to date and the bases and reasons for those opinions. I may offer additional opinions based on further review of materials in this case, including opinions and/or testimony of other expert witnesses. I make this declaration based upon my own personal knowledge and, if called upon to testify, would testify competently to the matters contained herein.

I. OVERVIEW OF THE TECHNOLOGY

A. Television Audience Measurement

6. Television audience measurement services have been widely used since the 1950s for gathering quantitative information to support, for example, advertising, programming, and financial forecasting. Ex-1011. Furthermore, automated systems to provide such services were well-known before 2011. Ex-1016; Ex-1004; Ex-1006. Typical components of such a service included obtaining information about the audience, such as the number of members, their

locations, the directions they are facing, and their identities, and identifying a television program broadcasted at the time by analyzing media data (e.g., audio data) associated with the program.

7. In fact, Patent Owner or its affiliate has developed automated systems for television audience measurement at least by 1992, almost 20 years before the earliest possible priority date of the '243 patent. In 1992, the A.C. Nielsen Company demonstrated to a reporter its “passive meter” that “can identify viewers and even gives a rough indication of whether their heads are turned toward the TV.” Ex-1016, 000003. The passive meter “scans a room every two seconds, looking for signs of life” by “compar[ing] each fresh image in a given location with its predecessor to produce a digital representation of movement, the hall mark of humanness.” Ex-1016, 000009. To identify individuals in the audience, the passive meter obtains multiple images of a known person and matches a new image to the obtained images. Ex-1016, 000009 (“The machine and I had already been introduced - that is, it had scanned me and then stored four images of my face: three head-on portraits and one quarter-profile. It was now ready to try matching the real me to these digital mug shots.”). If the identification is successful, the passive meter “display[s] on a nearby video screen [the identified person’s] name and a number indicating how confident it was about the identification.” Ex-1016, 000009. The passive meter also distinguishes those who

are watching the television from those involved in other activities. Ex-1016, 000009-10 (“Its reliance on head-on or nearly head-on views, moreover, means that it will not count people who are present but involved in other activities[.]”).

8. The PO or its affiliates continued to develop television audience measurement systems that use audio and video monitoring equipment to identify television programs and the audience and disclosed such systems in patent applications well before the ’243 patent. *See, e.g.*, Ex-1004, Abstract (“A television audience measurement system measures viewing of a television program viewed on digital television located in a Statistically Selected Site”); Ex-1006, ¶0036 (“A block diagram of an example local metering system 100 capable of providing viewing and metering information for program content presented via an example home entertainment system 102 is illustrated in FIG. 1.”), FIG. 1.

9. As explained below, all of the technology enablers for automated television audience measurement, including audio content recognition and face detection analysis and recognition, were all widely available before 2011 and used in this domain.

B. Audio Content Recognition

10. Audio content recognition is an approach to identifying live or recorded media, typically through the use of acoustic features. Early technology from the late 1990s provided solutions that extracted features from acoustic signals

and used temporal templates to match the resulting signatures. *See, e.g.*, Ex-1017, Abstract. Solutions were available since at least the early 2000s that use audio “fingerprinting” technology based on audio spectrograms to identify the source media from a set of predefined targets. Ex-1018, Abstract. This technology was robust and was able to identify music, and audio in general in noisy environments and in a speaker independent manner. Ex-1026, 000003-04.

11. Audio content recognition was applied in a variety of applications, and in particular was applied in audience measurement systems. Ex-1004, Abstract (“A television audience measurement system measures viewing of a television program viewed on digital television located in a statistically selected site by ... (iii) extracting an audio signature from the television program in order to identify the television program[.]”); Ex-1006, Abstract (“Display device ON/OFF detection methods and apparatus are disclosed. An example method disclosed herein to determine whether a presentation device is ON or OFF comprises determining an audio signature based on an audio signal corresponding to an output of the presentation device, comparing the audio signature to a set of reference signatures corresponding to a set of known media content....”).

C. Image Processing and Facial Analysis

12. By 2011, techniques and commercial products were widely available for facial analysis, including face detection, face or head pose estimation, and face

recognition. Ex-1019, 000002; Ex-1020, 000003; Ex-1021, 000002-03. All three problems were well-defined, benchmarks existed, and vendor tests were commonplace. The solutions that emerged were advanced enough to be applied to real-world problems such as entertainment, where face recognition can be used to control video games or play an important role in human robot interaction. Facial analysis techniques were also applied to information security, providing access controls to sensitive data or computer systems, or law enforcements for video surveillance, suspect tracking or criminal investigation. Ex-1021, 000002-03.

13. By 2011, there were numerous conferences that focused on the systematic evaluation of various face detection techniques and survey papers that summarized the status of face detection. *See, e.g.*, Ex-1019; Ex-1023. Databases also emerged along with benchmarks for face detection to quantitatively evaluate problems which were focused on face recognition in general unconstrained environments. *See, e.g.*, Ex-1025. Head pose estimation techniques were also advanced as described in a state-of-the-art survey in 2009. Ex-1020. The paper was based on an analysis of over 90 papers published in the field and demonstrated that the state-of-the-art was advanced enough to actively apply to various real-world applications. Starting in 2000, the National Institute of Standards and Technology (“NIST”) began running vendor tests for face recognition systems.

Ex-1024. At the time the technology was widely available and had matured into commercial products that could address large-scale and real-world applications.

14. One of the challenges leading up to this time was the ability to perform analysis efficiently to save computation resources while maintaining the quality of the analysis. It was well-known that savings of computation resources can be achieved by performing facial analysis on lower-resolution images. Certain facial analysis methods tolerate low-resolution images better than others. For example, Tian showed that face detection was somewhat agnostic to resolution and able to operate equally as well on faces of resolution 288x384 as on faces of resolution 36x48. Ex-1005, 000005. This enabled acquisition systems to have fewer constraints on the size of the face in an image but also suggested that processing could be done on lower-resolution images to conserve precious computational resources. On the other hand, it was known that facial recognition suffered from degraded performances at lower resolution. Ex-1027, 000005, Fig. 7:

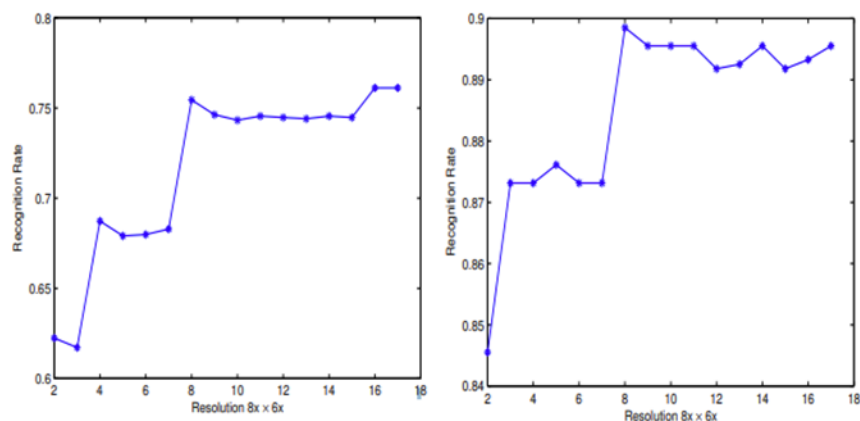


Fig. 7. Performance curve with the resolution changing from 16×12 to 136×102 based on PCA (the left figure) and LDA (the right figure)

15. By 2011, techniques for obtaining reduced-resolution images were widely known. Such techniques ranged from simply subsampling the image and dropping pixels to various forms of interpolation. These techniques were applied for a variety of reasons, including, for example, the ability to save computation by working on less data, providing representations that could be used to find objects and different scales, and binning pixels that compensate for low-light environments.

II. THE '243 PATENT

A. Overview

16. The '243 patent relates to “[m]ethods and apparatus to obtain exposure data for media exposure environment(s)[,]” which “determine[s] audience identification information and content identifying data.” Ex-1001,

Abstract. According to the '243 patent, “[an] audience measurement system ... collect[s] content identifying information to identify media content being presented ... in [a] room.” Ex-1001, 2:25-40. The system “recognize(s) objects as humans in a series of captured images of the room” to maintain a count of the humans and “recognize[s] an identity of one or more of the detected humans by, for example, performing a facial recognition process on image data[.]” Ex-1001, 2:25-56.

17. The '243 patent does not purport to invent media content identification or audience recognition, explaining in its “Background” section that these functionalities were known in the prior art:

Audience measurement of media content (e.g., broadcast television and/or radio, stored audio and/or video content played back from a memory such as a digital video recorder or a digital video disc, audio and/or video content played via the Internet, video games, etc.) often involves collection of content identifying data (e.g., signature(s), fingerprint(s), embedded code(s), channel information, time of consumption information, etc.) and people data (e.g., identifiers, demographic data associated with audience members, etc.). The content identifying data and the people data can be combined to generate, for example, media exposure data indicative of amount(s) and/or type(s) of people that were exposed to specific piece(s) of media content.

Ex-1001, 1:43-55. The '243 patent further explains that “[prior art] audience measurement systems identify people in the images via one or more identification techniques such as, for example, facial recognition.” Ex-1001, 1:56-2:3.

18. The '243 patent suggests that “the higher the resolution of a frame of image data, the more likely identification made via facial recognition will be accurate.” Ex-1001, 2:42-56. However, high-resolution images are not required for functionalities such as counting people. Ex-1001, 3:41-44. Illumination is typically used in prior art systems to capture high-resolution images. Ex-1001, 2:57-3:3 (“In previous systems, each time the high-resolution image sensor captures a frame of image data, the illumination source illuminates a surrounding area. The resulting illumination provides lighting conditions favorable to capturing high-resolution image data.”). The '243 patent suggests several disadvantages of frequent use of illumination, such as draining power, shortening the lifetime of an illumination source, requiring heatsinking devices and techniques, and annoying people exposed to the illumination source. Ex-1001, 3:4-30.

19. The '243 patent purports to address the disadvantages of frequently using illumination by introducing an approach that “reserve[s] use of an illumination source of a high-resolution image sensor for frames of image data designated for processing that requires high-resolution image data.” Ex-1001, 3:31-35; *see also* 3:63-4:1:

Accordingly, example methods, apparatus, and articles of manufacture disclosed herein selectively activate an illumination source associated with an image sensor for certain frames. In contrast, previous systems activate the illumination source for each frame captured by a corresponding image sensor.

Images captured without illumination are processed to enhance the lighting while reducing the resolution. 3:44-62. The reduced-resolution images are analyzed for purposes such as people counting. 4:1-4.

20. While the purported new approach of the '243 patent is directed to selective activation of illumination, the concept of illumination is only recited in claims 24 and 25. Ex-1001, 18:43-50. None of the challenged claims include any limitation directed to the purported new approach. Ex-1001, 15:13-17:61.

B. Prosecution History

21. The application that led to the '243 patent was allowed as originally filed. No claim amendments were made. In the Notice of Allowance, the Examiner referred to U.S. Patent Application Publication No. 2011/0164188 ("Karaoguz") as relevant art. The Examiner cited to paragraph 0107 of Karaoguz and noted its teachings of "captur[ing] images of the viewing environment of view" and "process[ing] such images to determine a current estimated location and/or head orientation of viewer." Ex-1002, 000293. The Examiner noted that Karaoguz teaches "compar[ing] such images to one or more reference images in

order to determine a current estimated location and/or head orientation of viewer” and that “[t]he reference images may comprise, for example, images of adaptable screen assembly or other objects or points of interest.” Ex-1002, 000293-94. The Examiner did not cite to any teachings of Karaoguz about face recognition and noted “[t]he prior art of record, either alone or in combination, fails to teach or suggest the determination of audience identification information based on a match of the head to a known person associated with the media exposure environment wherein the media exposure environment consists of media content presented on the television screen.” Ex-1002, 000294.

C. Priority Date

22. The application for the ’243 patent was filed on May 23, 2022, as a continuation application in a family of patent applications. The earliest application in the family was filed on December 15, 2011. I have been asked to assume that the ’243 patent is entitled to claim priority to December 15, 2011.

D. Challenged Claims

23. I understand that Petitioner is challenging the validity of claims 1, 4-6, 8, 9, 11-14, 16, and 18-20 of the ’243 patent in the Petition for Inter Partes Review to which this declaration will be attached. Those claims are reproduced in Appendix 3. While this Petition and declaration are directed to the challenged

claims, I have considered all claims 1-29 of the '243 patent, as well as portions of the '243 patent prosecution history in forming my opinions.

III. SUMMARY OF THE PRIOR ART

24. There are a number of patents and publications that constitute prior art to the '243 patent. I have reviewed and considered the prior art discussed in this section, along with the materials listed in Appendix 2.

A. Invalidity Grounds

25. Based on my review and analysis of the materials cited herein, my opinions regarding the understanding of a POSITA in the 2011 timeframe, and my training and experience, it is my opinion that the challenged claims of the '243 patent are invalid based on the following grounds:

Grounds	Claims	Statutory Basis	Prior Art
1	1, 4-6, 8, 9, 11-14, 16, and 18-20	§ 103	Lu-577 in view of Tian
2	1, 9, and 16	§ 103	Nielsen-372 in view of Steinberg
3	1, 4-6, 8, 9, 11-14, 16, and 18-20	§ 103	Nielsen-372 in view of Steinberg and Tian

26. I understand that none of the above-referenced art was cited or considered during prosecution of the '243 patent. First, none of these references or any related patents are listed on the face of the '243 patent. In addition, I reviewed

the file history for the '243 patent and am not aware of any of the above references being discussed in the prosecution history.

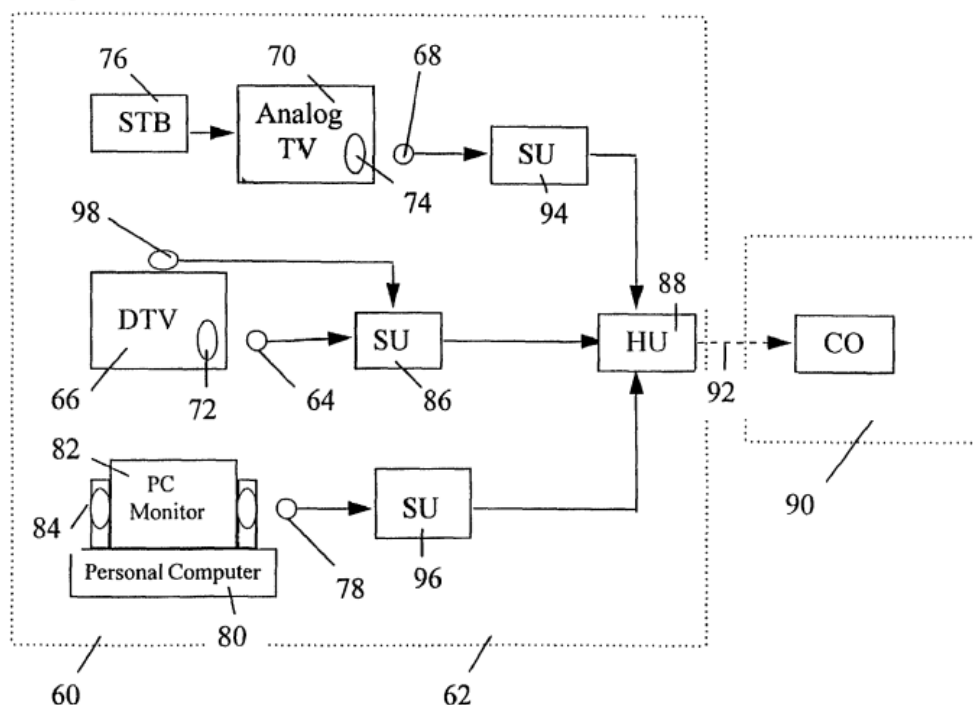
B. U.S. Patent Application Publication No. 2002/0059577 (“Lu-577”) (Exhibit 1004)

27. Lu-577 was filed July 19, 2001 and published May 16, 2002.

Therefore, Lu-577 is prior art under pre-AIA §102(b) because it was published over one year before December 15, 2011.

28. Lu-577 teaches “[a] television audience measurement system [that] measures viewing of a television program viewed on digital television located in a statistically selected site.” Ex-1004, Abstract. Lu-577 teaches that the television audience measurement system includes detectors (64, 68, 78) that detect audio portions of a presented television program and feed data to site units (86, 94, 96) or a home unit (88), which analyze the data to identify a television program. Ex-1004, ¶¶0035-0038, FIG. 2:

Figure 2



One approach to identify the television program is by extracting an audio signature, for which Lu-577 incorporates U.S. Patent No. 4,677,466 (“Lert”) by reference. Ex-1004, Abstract, ¶¶0003, ¶¶0017, ¶¶0035, ¶¶0040. Lu-577 teaches identifying the television program by comparing the audio signature with reference signatures. Ex-1004, ¶¶0035, ¶¶0040, ¶¶0041.

29. Lu-577 further teaches that its television audience measurement system includes a person identifier “to identify the persons watching television programming” “in the statistically selected monitoring site.” Ex-1004, ¶¶0039. Lu-577 teaches using “known head location and face recognition software” as taught by U.S. Patent No. 4,858,000 (“Lu-000”) “for the identification of the viewing persons and for the collection of other demographic data.” Ex-1004, ¶¶0039.

30. Lu-000 teaches “[a]n image recognition method and system ... for identifying predetermined individual members of a viewing audience in a monitored area.” Ex-1008, Abstract. A video camera is used to capture images for audience members in the monitored area. Ex-1008, Abstract, 2:20-38, 3:27-60, 3:61-4:8. Lu-000 teaches analyzing images to detect a head appearing in an image and to determine the location of the head. Ex-1008, 7:39-47, 9:47-52. Lu-000 further teaches storing a pattern image signature corresponding to each predetermined individual member of the viewing audience, extracting a pattern image signature from a captured image, and identifying an audience member by comparing a newly extracted pattern image signature with the stored pattern image signatures. Ex-1008, Abstract, 2:10-47, 7:39-8:6, FIGs. 6-7.

31. Lu-577 further teaches communicating television audience data to an intermediate data collector or a data collection central office. Ex-1004, ¶0050, ¶0037.

32. Lu-577 is in the same field of endeavor as the ’243 patent, which is media analysis. *See infra* §IV. Lu-577 “relates to the measurement of audiences of digitally broadcast television programming and to the measurement of usage of other information services provided to the audiences.” Ex-1004, ¶0001. Lu-577 teaches analyzing audio data to identify a television program and analyzing images to identify members of the audience. *See, e.g., infra* §VI.A.2.

C. *Evaluation of Face Resolution for Expression Analysis* (“Tian”) (Exhibit 1005)

33. Exhibit 1005 is a copy of Ying-li Tian, *Evaluation of Face Resolution for Expression Analysis*, 2004 Conference on Computer Vision and Pattern Recognition Workshop, Washington, DC, USA, 2004, pp. 82-82, doi: 10.1109/CVPR.2004.334. I understand that Tian is prior art under pre-AIA §102(b) because it was published no later than July 2, 2004, over one year before December 15, 2011. Ex. 1028.

34. Tian teaches obtaining a sequence of images of human subjects using a camera and analyzing the images to detect heads appearing in the images using face detection techniques.. Ex-1005, 000002, 000005. Tian further teaches performing pose estimation to determine orientations of the heads in the images. Ex-1005, 000001-03.

35. Tian creates lower-resolution images by down-sampling from original images and analyzes multiple different resolutions and performs face detection and pose estimation based on reduced-resolution images. Ex-1005, Abstract, 000005. According to Tian, the accuracy of face detection and pose estimation can be maintained at relatively low resolutions. Ex-1005, 000005. Tian further teaches determining a face/head location in a first image and only tracking changes to face features in subsequent images. Ex-1005, 000001, 000003.

36. Tian is in the same field of endeavor as the '243 patent, which is media analysis. *See infra* §IV. Tian “explores the effects of different image resolutions for each step of facial expression analysis.” Ex-1005, Abstract. Tian teaches analyzing images for head detection and pose estimation. *See, e.g., infra* §VI.A.2.

D. U.S. Patent Application Publication No. 2010/0274372 (“Nielsen-372”) (Exhibit 1006)

37. Nielsen-372 was filed July 7, 2010 and published October 28, 2010. Therefore, Nielsen-372 is prior art under pre-AIA §102(b) because it was published over one year before December 15, 2011.

38. Nielsen-372 is directed to “audience measurement.” Ex-1006, ¶0002. Nielsen-372 teaches a local metering system that is placed in the same environment as a home entertainment system, which presents media content and includes a display device such as a television. Ex-1006, ¶¶0036-0037, ¶0040.

39. The local metering system includes a sensor to receive audio signals corresponding to a displayed program and a metering unit and/or display device ON/OFF detector that identify the displayed program based on the audio signals. Ex-1006, ¶0036, ¶0044, ¶¶0049-0050, ¶0057, FIG. 1:

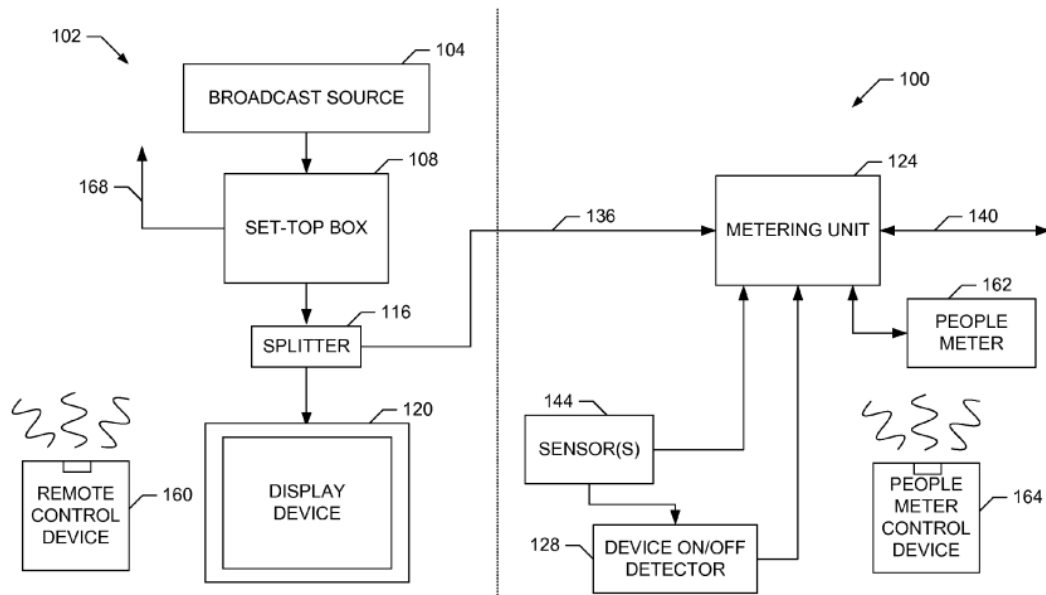


FIG. 1

The metering unit and/or display device ON/OFF detector generates an audio signature corresponding to a program being presented Ex-1006, ¶¶0044, ¶¶0051, ¶¶0055-0057, ¶¶0095-0100, FIGs. 2-3, FIGs. 14-15. Then, identifying information of the presented program is obtained by comparing the audio signature with a set of reference signatures. Ex-1006, ¶¶0044, ¶¶0057, ¶¶0096, FIG. 14. The local metering system also includes a people meter 162 that detects the identities of persons viewing programs on a television. Ex-1006, ¶¶0003, ¶¶0048.

40. The local metering system outputs viewing records to a central office or data processing facility over a network connection. Ex-1006, ¶¶0036, ¶¶0041, FIG. 1.

41. Nielsen-372 is in the same field of endeavor as the '243 patent, which is media analysis. *See infra* §IV. Nielsen-372 teaches analyzing audio data,

including “determining an audio signature based on an audio signal corresponding to an output of the presentation device [and] comparing the audio signature to a set of reference signatures corresponding to a set of known media content[.]” Ex-1006, Abstract; *see, e.g., infra* §VI.B.2.

E. U.S. Patent Application Publication No. 2010/0066822 (“Steinberg”) (Exhibit 1007)

42. Steinberg was filed September 4, 2009 and published March 18, 2010. Therefore, Steinberg is prior art under pre-AIA §102(b) because it was published over one year before December 15, 2011.

43. Steinberg is directed to grouping and classification of images based on the occurrence of faces in the images and the identification and classification of such faces. Ex-1007, ¶0003. Steinberg teaches an image capture device (e.g., a digital camera) that captures images and handles certain image processing operations. Ex-1007, ¶¶0128-0129, FIG. 1(a)-(c).

44. Steinberg teaches performing facial detection on acquired images to detect faces in the images and their corresponding locations. Ex-1007, Abstract, ¶0005, ¶0057, ¶0092, ¶¶0117-0118, ¶0128, ¶0130, ¶0151. Next, Steinberg detects the orientations of the faces to normalize them for further processing. Ex-1007, ¶0092, ¶0120, ¶0130, ¶0136, ¶0166, ¶0168, ¶0176, FIGs. 1(a)-(c), FIG. 5(a)-(b). Steinberg further teaches performing face recognition on the detected faces by generating a faceprint and comparing it with faceprints stored in a database. Ex-

1007, ¶¶0092, ¶0096, ¶¶0126-0127, ¶0130, FIGs. 1(a)-(c). As the database stores faceprints in association with identity information of known persons, identity information is determined based a match of faceprints. Ex-1007, ¶¶0099, ¶0135, ¶¶0156-0157, ¶0178, FIGs. 4(a)-(b), FIG. 9(a).

45. Steinberg is in the same field of endeavor as the '243 patent, which is media analysis. *See infra* §IV. Steinberg “relates to digital image processing, particularly to the field of automatic or semiautomatic grouping and classification of images in a database or image collection and based on the occurrence of faces in the images and the identification and classification of such faces.” Ex-1007, ¶0003. Steinberg teaches analyzing images to identify known persons. *See, e.g., infra* §VI.B.2.

IV. LEVEL OF ORDINARY SKILL IN THE ART

46. I have analyzed the '243 patent and determined that the field of the patent is media analysis, including analysis of audio signals (e.g., audio signals of a television program) and image signals (e.g., images of individuals in a room).

47. In determining the characteristics of a hypothetical person of ordinary skill in the art (“POSITA”) of the '243 patent at the time of the claimed invention, I considered several things, including various prior art techniques relating to media analysis, the type of problems that such techniques gave rise to, and the rapidity with which innovations were made. I also considered the sophistication of the

technologies involved, and the educational background and experience of those actively working in the field at the time. I also considered the level of education that would be necessary to understand the '243 patent. Finally, I placed myself back in the relevant period of time and considered the engineers and programmers that I have worked with and managed in the field of media analysis.

48. I came to the conclusion that a person of ordinary skill in the field of art 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. A person with less education but more relevant practical experience may also meet this standard.

49. In the year 2011 timeframe, a POSITA would have known and had the skills necessary to:

- Design or implement systems that collect audio and image data from an environment for analysis;
- Understand and implement known techniques for extracting characteristic features from audio content and identifying the audio content based on the extracted features;

- Understand and implement known techniques of image processing, including adjusting parameters of images, such as resolution, illumination, and orientation;
- Understand and implement known face or head analysis techniques on images, such as face or head detection, face or head pose estimation, and identity determination based on face or head in images.

V. CLAIM CONSTRUCTION

50. For purposes of this *inter partes* review, I have considered the claim language, specification, and portions of the prosecution history, to determine the meaning of the claim language as it would have been understood by a person of ordinary skill in the art at the time of the invention. I have applied the *Phillips* standard in my analysis. Unless otherwise stated, I have applied the plain and ordinary meaning to claim terms.

A. Media Exposure Environment

51. Challenged claims 1, 4, 9, 11, 16, and 18 recite “media exposure environment.” Ex-1001, 15:13-17:61. The claims recite “a television within the media exposure environment,” “images of the media exposure environment,” and “a known person associated with the media exposure environment.” *Id.* The specification of the ’243 patent provides examples of a “media exposure environment,” including “a television room, a family room, a living room, a bar, a

restaurant.” *Id.*, 1:56-2:3, 4:15-51. The ’243 patent further explains that the “media exposure environment” can be “a room of a house in which a television is located.” *Id.*, 2:25-41. In light of Figure 1, the specification states that “the media exposure environment 100 is a room of a household that has been statistically selected to develop television ratings data for a population/demographic of interest.” *Id.*, 4:52-5:8, FIG. 1:

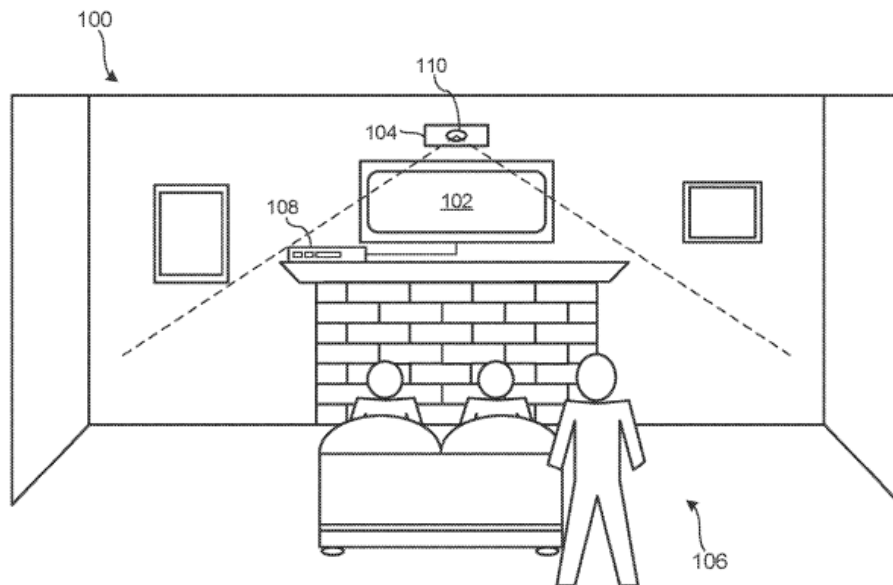


FIG. 1

The ’243 patent states that Figure 1 illustrates an “example” media exposure environment, which includes a media presentation device and an audience measurement device. *Id.* This indicates that the meaning of “media exposure environment” is not limited to that illustrated in Figure 1.

52. In light of the claim language and the specification, it is my opinion that a POSITA would have understood the meaning of “media exposure environment” in the context of the ’243 patent to at least encompass a space in which a television is located.

VI. INVALIDITY

53. Based on my review and analysis of the materials cited herein, my opinions regarding the understanding of a POSITA in the 2011 timeframe, and my training and experience, it is my opinion that:

- Claims 1, 4-6, 8, 9, 11-14, 16, and 18-20 of the ’243 patent are obvious in view of Lu-577 and Tian.
- Claims 1, 9, and 16 of the ’243 patent are obvious in view of Nielsen-372 and Steinberg.
- Claims 1, 4-6, 8, 9, 11-14, 16, and 18-20 of the ’243 patent are obvious in view of Nielsen-372, Steinberg, and Tian.

The reasons for my conclusions are explained more fully below.

A. Ground 1: Claims 1, 4-6, 8, 9, 11-14, 16, and 18-20 are Rendered Obvious in View of Lu-577 and Tian

54. It is my opinion that a POSITA would have found Claims 1, 4-6, 8, 9, 11-14, 16, and 18-20 obvious based on the teachings of Lu-577 and Tian.

1. Combination of Lu-577 and Tian

55. Lu-577 teaches a television audience measurement system that measures viewing of a television program on television. Ex-1004, Abstract. The television audience measurement system “extract[s] an audio signature from the television program in order to identify the television program[.]” Ex-1004, Abstract, ¶¶0017, ¶¶0035, ¶¶0040, FIG. 2. The television audience measurement system further uses “known head location and face recognition software ... for the identification of the viewing persons[.]” Ex-1004, ¶¶0039.

56. Lu-577 incorporates Lu-000 by reference as teaching already-known exemplary location and face recognition software. Ex-1004, ¶¶0013 (“Notable among teachings in this area is that by Lu in U.S. Pat. No. 4,858,000. The teaching of this patent is herein incorporated by reference.”), ¶¶0039 (“known head location and face recognition software (e.g., as taught by Lu in U.S. Pat. No. 4,858,000)”). Lu-000 teaches “[a]n image recognition method and system ... for identifying predetermined individual members of a viewing audience in a monitored area.” Ex-1008, Abstract. Lu-000 captures video images for individual audience members, analyzes the images to locate the heads or faces of the audience members, extracts pattern image signatures from the images, and compares the extracted pattern images signatures with stored signatures to determine the identities of the audience members. Ex-1008, Abstract, 7:39-9:52.

57. Similar to Lu-000, Tian teaches techniques for image processing and facial analysis. Tian teaches processing image sequences of human subjects to perform facial analysis that includes, for example, face detection, head pose estimation, and expression recognition. Ex-1005, 000002, 000005. Tian teaches processing reduced-resolution images and studies the effects of the reduction of resolution on the performance of different facial analysis techniques, including face detection and head pose estimation. Ex-1005, Abstract, 000005.

58. A POSITA would have found it obvious to combine the teachings of Lu-000 and Tian because they are directed to similar functionalities (e.g., image processing and facial analysis) and address similar software-based tasks (e.g., identifying a head in an image). As explained below, Tian provides multiple techniques that provide performance improvements for facial analysis methods, such as those taught by Lu-000. A POSITA would have understood that Tian's techniques are applicable to Lu-000's teachings and would have been motivated to use these techniques to improve Lu-000.

59. Lu-577 teaches employing known head location and face recognition software to identify audience members. Ex-1004, ¶0039. Lu-577 refers to Lu-000 as one example of such software solutions, clearly indicating that it contemplates using other software solutions with relevant functions in the context of television audience measurement. Ex-1004, ¶0039. A POSITA would have realized that

Tian's solutions of facial analysis are applicable to the television audience measurement solution of Lu-577 because Lu-577 provides the capability of capturing images of audience members (Ex-1004, ¶0039) and the captured images provide the basis of Tian's analysis. A POSITA would have found it obvious to employ the known head location and face analysis software solutions of Tian in Lu-577's system for audience measurement based on Lu-577's explicit invitation of employing such known solutions.

60. Tian teaches performing facial analysis on sequences of images of subjects. *See, e.g.*, Ex-1005, 000005 ("The database contains 704 image sequences from 97 subjects. ... The length of the image sequences are varying from 9 to 47 frames."). A POSITA would have found it obvious to apply Tian's teaching to Lu-577's teachings (including Lu-000) such that the television audience measurement system analyzes sequences of images to identify audience members. Lu-577 and Lu-000 teaches identifying members of an audience by analyzing images of the audience. *See, e.g.*, Ex-1004, ¶0039; Ex-1008, 3:27-60. Neither Lu-577 nor Lu-000 limits the number of images for an audience member that can be analyzed. Furthermore, Lu-577 and Lu-000 teach using a video camera to obtain audience images. *See, e.g.*, Ex-1004, ¶0039 ("The person identifier 98 may be video camera, an IR camera, or the like."); Ex-1008, 3:61-4:8 ("[T]he audience scanner 12 includes a video camera 28 for capturing a video image[.]") A POSITA would

have understood that a video camera is capable of taking a sequence of images over time. It would have been obvious to implement Tian's teachings of analyzing a sequence of images as obtained by a video camera.

61. A POSITA would have been motivated to apply Tian's teaching of performing facial analysis on a sequence of images because of the benefits provided by such an approach. First, a POSITA would have known that facial analysis usages always strive for higher accuracy levels, and that operations such as head detection and face recognition generally correspond to certain levels of accuracy. As contemporaneous writings recognized, the accuracy of identifying a person can be improved by analyzing multiple images including the same person and aggregating the analysis results from the multiple images. *See, e.g.*, Ex-1022, 000007. This would have motivated a POSITA to analyze multiple images in order to achieve a higher confidence level for the identification than a single image. Second, a POSITA would have understood that a single image captures a person's facial features at a particular point in time. The person's facial expression and orientation, as captured by that image, may or may not be suitable for face recognition. For example, the person may close their eyes or face a direction other than the camera, and in those cases, face recognition with high accuracy may not be feasible. Ex-1010, ¶0003 (“[A] typical face detector can only accurately detect faces in a semi-frontal pose.”). Obtaining and analyzing a sequence of images

allows the selection of images in which the facial features are suitable for face recognition and selective performance of face recognition on suitable images. A POSITA would have understood that this would have improved the accuracy of the audience identification.

62. Tian teaches determining a face/head location in a first image and only tracking changes to face features in subsequent images. *See, e.g.*, Ex-1005, 000001 (“Face acquisition is a processing stage to automatically find the face region for the input images or sequences. It can be a face detector to ... just detect face in the first frame and then track the face in the remainder of the video sequence.”), 000003 (“Given an image sequence, the region of the face and approximate location of individual face features are detected automatically in the initial frame. ... Then, all face feature changes are automatically tracked in the image sequence.”). Under this approach, if no change is detected, the face is determined to be at the same location in subsequent images.

63. A POSITA would have found it obvious to apply Tian’s teaching to Lu-577 such that when the television audience measurement system processes a sequence of images, it determines a face or head location in a first image and use the determined location for subsequent images. For example, the television audience measure system analyzes a portion in a subsequent image corresponding to the location for face recognition.

64. A POSITA would have understood that this approach is applicable when a sequence of images is taken for a person by a video camera. Video cameras were able to take multiple images in a relatively short period of time, in which a person's head or face is unlikely to move significantly. Therefore, the region corresponding to the person's head or face likely remains the same across the sequence of images. A POSITA would have found it obvious to apply this approach since this approach is taught in contemporaneous writings. Ex-1010, ¶0003 ("Once a face is detected, its location is recorded and a localized region around that face is scanned by a face detector in the next frame. Thus, once a face is initially detected, it can be accurately tracked from frame to frame without a need to run a face detector across the entire image."). A POSITA would have been motivated to apply the teachings of Tian because this approach has the benefit of saving processing resources and time by reducing the number of times face detection needs to be applied.

65. Tian teaches performing head pose estimation on a subject based on an image. *See, e.g.*, Ex-1005, 000002 ("[T]he head pose was estimated based on the detected head."). Tian teaches categorizing the subject into one of several possible poses such as frontal or new frontal view, side view or profile, and others. *See, e.g.*, Ex-1005, 000003. A POSITA would have found it obvious to apply Tian's teaching of determining head poses to Lu-577 such that the television

audience measurement system determines an orientation of the head with respect to the camera before performing further steps to identify the corresponding person. A POSITA would have understood that a person's head pose refers to an orientation of the person's head with respect to the camera. The visibility of the person's facial features differ for different poses. Ex-1005, 000002 ("In the frontal or near frontal view, both eyes and lip corners are visible. In side view or profile, at least one eye or one corner of the mouth becomes self-occluded because of the head.").

66. A POSITA would have understood that certain poses are not suitable for face recognition. For example, 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. Ex-1010, ¶0003 ("[A] typical face detector can only accurately detect faces in a semi-frontal pose."). In fact, Tian teaches performing facial expression analysis only for frontal or near frontal view faces. Ex-1005, 000002 ("The expression analysis process is applied only to the frontal and near frontal view faces."). A POSITA would have found it obvious to apply this approach to Lu-577 such that 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. In this manner, it can be ascertained that face recognition is only performed on images in which a person's head pose is suitable for such analysis. This improves the accuracy of identification of audience members and addresses the problem that pose variations

adversely affect recognition performance. Ex-1021, 000042. A POSITA would have been motivated to apply Tian's teachings to Lu-577 in order to achieve this benefit.

67. Furthermore, 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 the television. Using an individual's head pose to determine the individual's watching activity was known in the 2011 timeframe. Ex-1012, ¶0039 ("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."). In light of Lu-577's function of measuring viewing of a television program, a POSITA would have found it beneficial to apply the teachings of Tian such that the television audience measurement system better determines whether the detected individuals are watching the television program.

68. Tian teaches reducing the resolution of an image to obtain reduced-resolution images and performing analysis on the reduced-resolution images such as face detection and head pose estimation. *See, e.g.*, Ex-1005, Abstract, 000002, 000005. Tian teaches that the accuracy of face detection and head pose estimation

can be maintained for reduced-resolution images. *See, e.g.*, Ex-1005, 000005 (“The results show that both face detectors [23, 29] can detect all faces for head region in resolution of 288x384, 144x192, 72x96, and 36x48, but failed for resolution of 18x24. Compared to face detection, head pose estimation can detect 98% faces for all the resolutions.”).

69. A POSITA would have found it obvious to apply Tian’s teaching to Lu-577 such that the television audience measurement system reduces the resolution of an image and performs face detection or head pose estimation on the reduced-resolution image. Tasks such as face detection can be resource-consuming. Ex-1013, 000001 (“[V]iew-based [face detection] methods ... are reported to have a good detection performance but the drawback to be computationally rather expensive and at present are not the appropriate choice for real-time applications.”). Some approaches of face detection may require analyzing all pixels of an image to search for a face. The processing resources and time can be reduced if the resolution of the image is reduced. Ex-1014, 2:43-53 (teaching that “calculations are avoided of a complete highest resolution integral image for every acquired image in an image stream” in face tracking, thereby “reduces processing overhead for face detection[.]”); Ex-1015, ¶0022 (“By reducing the resolution of the saturation of the image, computational requirements are substantially reduced and a relatively simple method may be used.”).

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 according to Tian's teachings. Ex-1005, 000005. A POSITA would have been motivated to apply Tian's teachings of analyzing reduced-resolution images to achieve the benefit of savings in processing resources and time.

70. It was known that the performance of face recognition may be adversely affected by the reduction of image resolution. Ex-1027, 000004-05. Therefore, a POSITA would have been motivated to perform certain facial analysis techniques (e.g., face detection, head pose estimation) on reduced-resolutions images, while other techniques (e.g., face recognition) on full-quality images.

71. A POSITA would have found it obvious and been motivated to apply the teachings of Tian to Lu-577 because this is an application of known techniques to a known system and method ready for improvement. Lu-577 teaches a system and method for measuring viewing of a television program, in which Lu-577 explicitly describes the use of known head location and face recognition software. Tian provides known techniques for determining the location and orientation of faces or heads in images. Lu-577 contemplates the application of known techniques, such as those taught by Tian, to be applied to its system and method. Furthermore, as explained above, applying Tian's techniques to Lu-577 would

have provided improvements such as improved accuracy of face recognition (e.g., by processing a sequence of images and determining the orientation of the head) and savings in processing resources and time (e.g., by analyzing reduced-resolution images and tracking the location of the head in a sequence of images). Applying Tian's teachings to Lu-577 would have yielded predictable results at least because Tian provides experimental assessments of its techniques, such as the accuracy of face detection and head pose estimation for images of different resolutions.

72. A POSITA would have had a reasonable expectation of success when combining the teachings of Lu-577 and Tian, given the overlap between Lu-000 and Tian and that Tian's image processing techniques were well known. As explained above, the combination simply uses Tian's teachings about image processing and facial analysis, in the manner taught by Tian, to implement functionalities in Lu-577 where Lu-577 expressly refers to techniques that were known in the art. A POSITA would have been more than capable of implementing the image processing techniques taught by Tian, such as creating reduced-resolution images by down-sampling an original image, face detection, head pose estimation, and face tracking. *See supra* §I. Furthermore, a POSITA would have been motivated and would have had a reasonable expectation of success in combining the references, as explained above, because the combination would not have changed the principle of operation for any of the teachings of the references

relied upon in the combination. Here, Tian's techniques are used in a conventional manner and applied to Lu-577 in a manner contemplated by Lu-577.

2. Claim 1

73. It is my opinion that a POSITA would have found claim 1 obvious based on the combined teachings of Lu-577 and Tian.

1[p]. An audience measurement system to obtain exposure data for a media exposure environment,

74. I understand that a preamble may or may not state a limitation of a claim. However, to the extent the preamble limits claim 1, it is my opinion that Lu-577 teaches the preamble.

75. Lu-577 teaches **an audience measurement system** (e.g., a television audience measurement system). Ex-1004, Abstract (“A television audience measurement system measures viewing of a television program viewed on digital television located in a statistically selected site”); ¶¶0014-0017.

76. Lu-577 teaches that its television audience measurement system **obtains exposure data** (e.g., data identifying a television program(s), data identifying the audience) **for a media exposure environment** (e.g., a statistically selected site). Ex-1004, Abstract, ¶¶0017, ¶¶0035, ¶¶0039-0040, FIG. 2. For example, Lu-577 teaches that its television audience measurement system “measures viewing of a television program viewed on digital television located in a statistically selected site[.]” Ex-1004, Abstract, ¶¶0017, ¶¶0035. The television

audience measurement system of Lu-577 “extract[s] an audio signature from the television program in order to identify the television program[.]” Ex-1004, Abstract, ¶¶0017, ¶¶0035, ¶¶0040, FIG. 2. Lu-577 further teaches that its television audience measurement system includes a person identifier “to identify the persons watching television programming” “in the statistically selected monitoring site.” Ex-1004, ¶¶0039.

77. Lu-577 teaches using “known head location and face recognition software” as taught by Lu-000 “for the identification of the viewing persons and for the collection of other demographic data.” Ex-1004, ¶¶0039. Lu-000 teaches “[a]n image recognition method and system ... for identifying predetermined individual members of a viewing audience in a monitored area.” Ex-1008, Abstract. Lu-577 teaches obtaining information regarding the television program and the persons watching the television program as “television audience data.” Ex-1004, ¶¶0050.¹

¹ Lu-577 includes a first embodiment corresponding to Figure 2 and a second embodiment corresponding to Figure 3 both directed to its television audience measurement system. Lu-577 states that “this second embodiment 100 may include some or all of the features of the first embodiment 60.” Ex-1004, ¶¶0044. In other words, Lu-577 contemplates the use of all the features of the two embodiments together. In light of this teaching, a POSITA would have found it obvious, for example, for the software agents of the second embodiment to implement some or all of the functionalities of the first embodiment. Ex-1004, ¶¶0046. The first and second embodiments are compatible with each other and intended to be applied together.

78. The television audience data in Lu-577 constitute “exposure data,” as the term is used in the ’243 patent. The ’243 patent explains that exposure data include media content identification(s) and personal identification(s). Ex-1001, 5:9-28 (“Identification(s) of media content to which the audience 106 is exposed are correlated with the people tallies and/or the personal identifications to generate exposure data for the media content.”), 1:43-2:3. The meaning of “exposure data” in the ’243 patent is consistent with Lu-577’s teaching of collecting data identifying a television program and individuals watching the program.

79. The statistically selected site in Lu-577 and the monitored area in Lu-000 both refer to a “media exposure environment” as they are spaces in which a television is located. *Supra* §V.A. Lu-577 explains that the statistically selected site is one in which a television is located, a television program is viewed, and the television viewing behavior is monitored. Ex-1004, ¶0017 (“[A] television audience measurement system measures viewing of a television program viewed on digital television located in a statistically selected site[.]”), ¶0089 (“[A] statistically selected monitoring site ... includes certain viewing equipment as well as certain monitoring equipment.”). Lu-000 describes that the monitored area is a space in which an audience watches television. Ex-1008, Abstract (“An image recognition method and system are provided for identifying predetermined individual members of a viewing audience in a monitored area.”), 1:15-19 (“The

present invention relates generally to image recognition systems for use with television audience measurement and marketing data collection systems, and more particularly to image recognition systems and methods for identifying predetermined individual members of a viewing audience.”). The ’243 patent’s description that the media exposure environment is statistically selected is further consistent with Lu-577’s teaching of the statistically selected site. Ex-1001, 4:56-5:8. Therefore, Lu-577 teaches an audience measurement system to obtain exposure data for a media exposure environment.

80. Furthermore, the ’243 patent admits that the preamble was known in the prior art by discussing it in the “Background” section. Ex-1001, 1:56-2:3 (“[S]ome [audience] measurement systems capture a series of images of a media exposure environment ... and analyze the images to determine how many people appear in the images at a particular date and time. The calculated amount of people in the media exposure environment can be correlated with media content being presented at the particular date and time to provide exposure data (e.g., ratings data) for that media content.”).

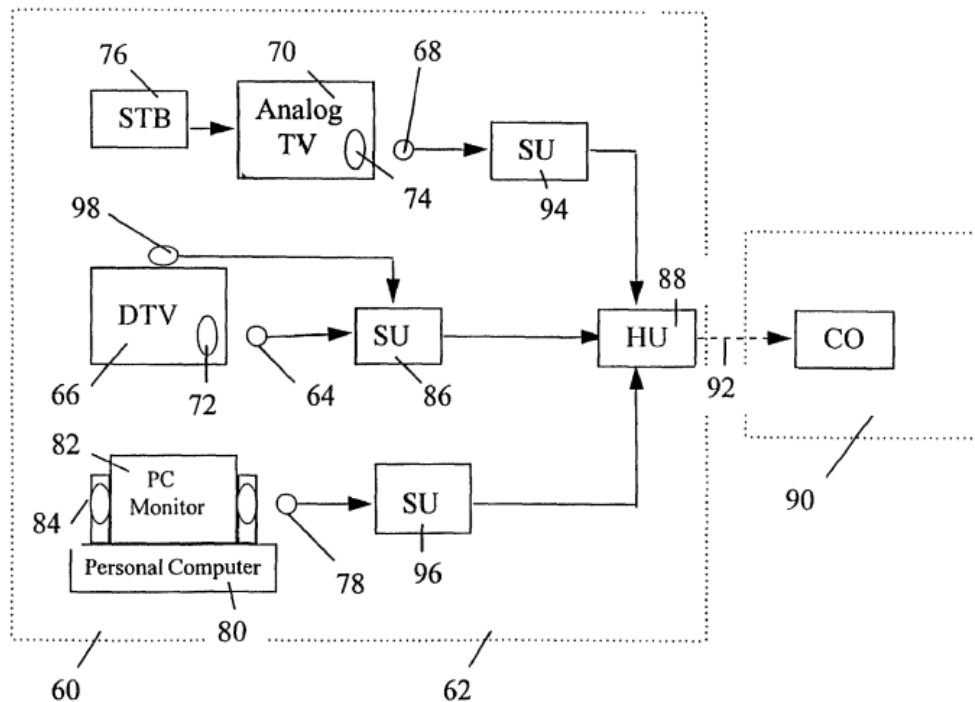
[1a] the audience measurement system comprising: memory; machine readable instructions; and processor circuitry to execute the machine readable instructions to:
--

81. Lu-577 teaches this limitation. Lu-577 teaches that its television audience measurement system comprises **memory, machine readable**

instructions, and processor circuitry to execute the machine readable

instructions. Lu-577 teaches that the television audience measurement system includes multiple components, including those shown in Figure 2:

Figure 2

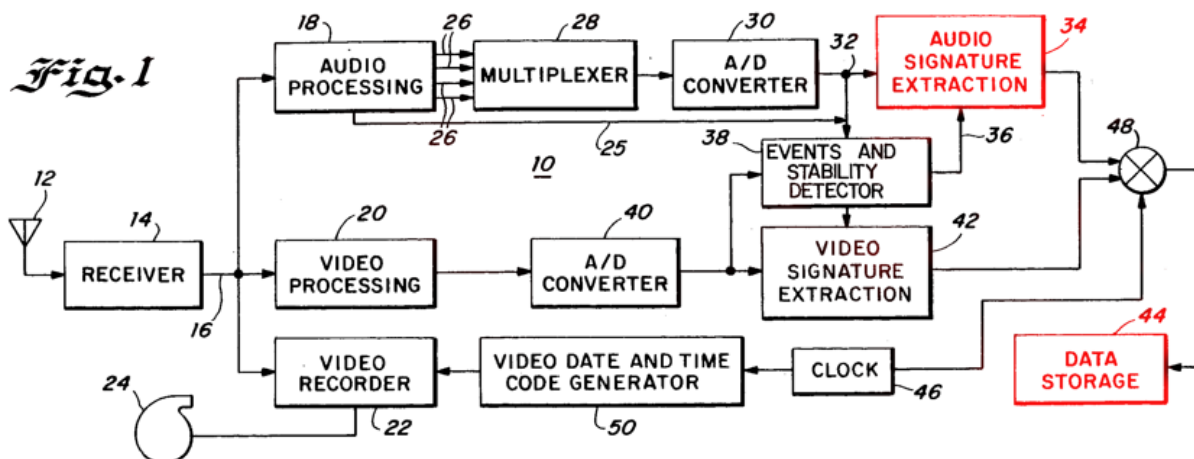


Ex-1004, FIG. 2. Lu-577 explains that the statistically selected site (62) includes detectors (64, 68, 78) that detect audio portions of a presented program. Ex-1004, ¶¶0035-0036. The detectors (64, 68, 78) feed data to site units (86, 94, 96) or a home unit (88), which analyze the data to identify a television program. Ex-1004, ¶¶0037-0038. One approach to identify the television program is by extracting an audio signature, for which Lu-577 incorporates Lert by reference. Ex-1004, ¶0003 (“This approach [of extracting a characteristic feature signature] is taught by Lert

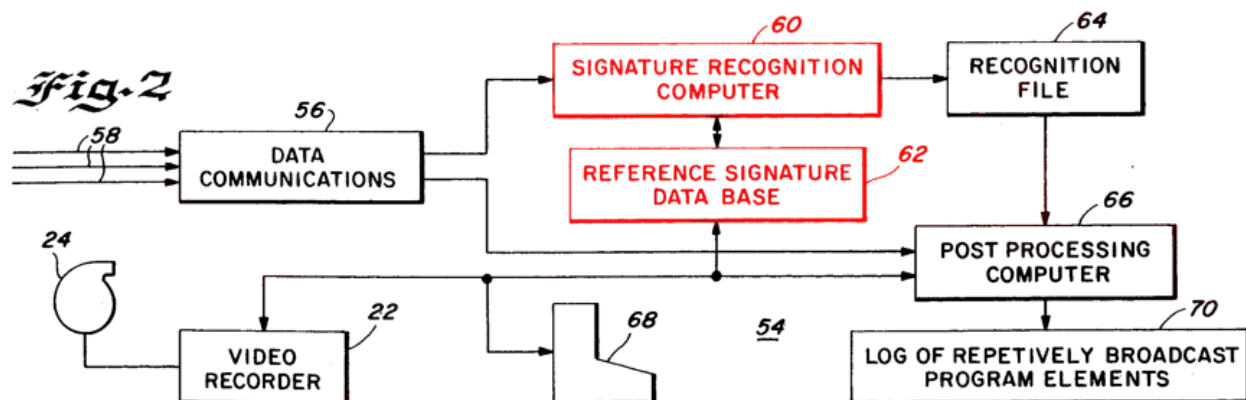
and Lu in U.S. Pat. No. 4,677,466. The teaching of U.S. Pat. No. 4,677,466 is herein incorporated by reference.”), ¶0040.

82. Lu-577 further teaches that the statistically selected site (60) includes a person identifier (98) that “may be video camera, an IR camera, or the like.” Ex-1004, ¶0039. The person identifier (98) feeds data to the site unit (86), which employs “known head location and face recognition software” (e.g., as taught by Lu-000) to identify viewing persons in the audience. Ex-1004, ¶0039. The detectors, person identifier, site units, and home unit in Lu-577 are components of the television audience measurement system as they collectively measure exposure to media content. Ex-1004, ¶¶0035-0040.

83. Lert teaches using **processor circuitry** (e.g., audio processing circuitry, signature recognition computer) and **memory** to perform its methods. Lert teaches an audio processing circuitry for extracting an audio signature and a memory for storing the audio signature. Ex-1009, 3:57-4:12 (“The audio signature extraction circuitry 34 extracts an audio signature from the digitized audio signals 32”), 4:13-42 (“The signature thus generated is stored in a data storage device or memory 44”), FIG. 1 (emphasis added):

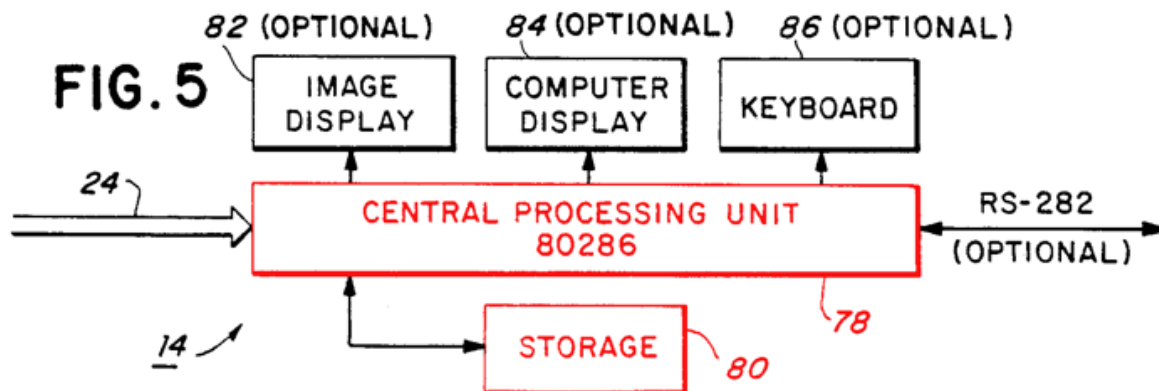


Lert further teaches a signature recognition computer that compares the extracted signature “to a plurality of reference signatures stored in a memory[.]” Ex-1009, 4:61-5:19; FIG. 2 (emphasis added):



84. Lu-000 teaches using **processor circuitry** (e.g., audience locating circuitry, control command processor subsystem, audience recognition subsystem) and **memory** to perform its methods. Lu-000 teaches “[an] image recognition system 10 [that] includes an audience scanner 12 ... and a control command processor subsystem 14[.]” Ex-1008, 3:27-60. The audience scanner “includes

audience locating circuitry for locating individual audience members in the monitored area.” Ex-1008, 2:20-38. The control command processor subsystem includes a central processing unit 78 and a memory device 80. Ex-1008, 5:61-6:4, FIG. 5 (emphasis added):



The image recognition system further includes “[a]n audience recognition subsystem 26 for learning and for recognizing feature image signatures of the images of the audience members.” Ex-1008, 3:27-60. The image recognition system extracts pattern image signatures from captured images and compares them to stored pattern image signatures to identify audience members. 7:39-8:6. Face pattern image signatures are stored in the memory. Ex-1008, 7:25-38.

85. By incorporating Lert and Lu-000 by reference, Lu-577 teaches performing steps of identifying a television program and audience members, which are further discussed with respect to elements [1b] – [1g], using computer hardware such as memory and processor circuitry. Lu-577 teaches that the

processor circuitry performs the steps by **executing machine readable instructions** (e.g., head location and face recognition software, software program for identifying television program based on audio data). Ex-1004, ¶¶0039-0040. By 2011, computing systems that include basic components such as memories, processor circuitries, and machine readable instructions executable by the processor circuitries, such as personal computers, were well developed and commonly used. In fact, Lu-577 teaches that its digital television equipment comprises a processor circuitry (e.g., microprocessor), a memory, and machine readable instructions (e.g., operating system). Ex-1004, cl. 14. A POSITA would have found it obvious to implement Lu-577's television audience measurement system, including its components such as the site unit and home unit, using this ordinary architecture that includes memory, machine readable instructions, and processor circuitry, such that the processor circuitry executes the machine readable instructions to perform operations directed to identifying a television program and audience members. Therefore, Lu-577 teaches limitation [1a].

[1b] generate an audio signature of media content presented by a television within the media exposure environment;
--

86. Lu-577 teaches this limitation. Lu-577 teaches that its television audience measurement system **generates an audio signature of media content presented by a television** (e.g., a television program viewed on digital television)

within the media exposure environment (e.g., statistically selected site). Ex-1004, Abstract (“A television audience measurement system measures viewing of a television program viewed on digital television located in a statistically selected site by ... (iii) extracting an audio signature from the television program in order to identify the television program....”); ¶0017 (“[A] television audience measurement system measures viewing of a television program viewed on digital television located in a statistically selected site and comprises ... extracting means ... The extracting means extracts an audio signature from the television program in order to identify the television program.”). A POSITA would have understood that a television program is media content presented by a television. Also, Lu-577’s statistically selected site is a media exposure environment, as explained above. *Supra* §VI.A.2[1p].

87. Lu-577 teaches operations performed by its television audience measurement system to “measure[] tuning to digital television programming received in a statistically selected monitoring site[.]” Ex-1004, ¶0035. Lu-577 sets forth the extraction of an audio signature as one of the techniques for identifying a television program. Ex-1004, ¶0035 (“The first embodiment 60 measures such tuning by detecting program identification codes and/or **extracting signatures from the audio portion of the television signal to which a receiver is tuned.**” (emphasis added)), ¶0040 (“[C]haracteristic signatures can be extracted from these

representations for subsequent comparison....”), ¶0041 (“[T]he embodiments of the present invention described herein can determine the programs being viewed, when the tapes are played, by extracting characteristic signatures from these programs....”).

88. The ’243 patent does not limit the method used to generate the audio signature. A POSITA would have understood that the terms “extracting” and “generating” are often used interchangeably to refer to the creation (e.g. generation) of an audio signature based on audio content. Extracting a characteristic signature of a piece of audio content is at least one way to generate an audio signature. Therefore, Lu-577 teaches an audience measurement system generating an audio signature of media content presented by a television within the media exposure environment.

89. Furthermore, the ’243 patent admits that this limitation was known in the prior art by discussing it in the “Background” section. Ex-1001, 1:43-55 (“Audience measurement of media content ... often involves collection of content identifying data (e.g., **signature(s)**, fingerprint(s), embedded code(s), channel information, time of consumption information, etc.)....” (emphasis added)).

[1c] obtain content identifying data corresponding to the presented media content, the content identifying data based on the audio signature of the media content presented by the television within the media exposure environment;
--

90. Lu-577 teaches this limitation. Lu-577 teaches that its television audience measurement system **obtains content identifying data** (e.g., data identifying a television program(s)) **corresponding to the presented media content** (e.g., television program), **the content identifying data based on the audio signature of the media content presented by the television within the media exposure environment** (e.g., statistically selected site). Ex-1004, Abstract (“A television audience measurement system measures viewing of a television program viewed on digital television located in a statistically selected site by ... (iii) extracting an audio signature from the television program in order to identify the television program[.]”); ¶0017 (“The extracting means extracts an audio signature from the television program in order to identify the television program.”).

91. As explained with respect to limitation [1b], Lu-577 teaches its television audience measurement system generating an audio signature of media content presented by a television within the media exposure environment. *Supra* §VI.A.2[1b]. Lu-577 further teaches identifying the media content (e.g., television program) based on the audio signature. Ex-1004, ¶0035 (“A first embodiment 60 of the present invention, which measures tuning to digital television programming

received in a statistically selected monitoring site 62, is depicted schematically in FIG. 2. The first embodiment 60 measures such tuning by ... extracting signatures from the audio portion of the television signal to which a receiver is tuned.”), ¶0040 (“[C]haracteristic signatures can be extracted from these representations for subsequent comparison, either in the statistically selected monitoring site 62 or in the central office 90, with previously collected reference signatures in order to identify the television programs to which the digital television set 66, the analog television set 70, and the personal computer 80 are tuned.”), ¶0041 (“[T]he embodiments of the present invention described herein can determine the programs being viewed, when the tapes are played, by extracting characteristic signatures from these programs and by matching the extracted signatures to reference signatures stored in a reference signature database.”).

92. Lu-577 teaches representing information identifying a television program as “tuning data.” Ex-1004, ¶0044 (“The second embodiment 100 measures tuning to digital television programming in connection with a statistically selected monitoring site 102.”), ¶0045 (“According to this second embodiment 100, tuning data, and possibly demographic data, are collected from digital television reception equipment...”), ¶0050 (“The software agents 112, 118, and 122 may be telecommunication enabled so that television audience data from all the viewing sites in the statistically selected monitoring site 102 can be

communicated....”). A POSITA would have understood that such tuning data is “content identifying data” because the tuning data identifies the media content that is presented by a television. The process of identifying the media content is the process of obtaining the content identifying data. Therefore, Lu-577 teaches an audience measurement system for obtaining content identifying data corresponding to the presented media content, the content identifying data based on the audio signature of the media content presented by the television within the media exposure environment.

93. Furthermore, the ’243 patent admits that this limitation was known in the prior art by discussing it in the “Background” section. Ex-1001, 1:43-55 (“Audience measurement of media content ... often involves collection of **content identifying data** (e.g., signature(s), fingerprint(s), embedded code(s), channel information, time of consumption information, etc.)....” (emphasis added)).

[1d] analyze a sequence of images of the media exposure environment to detect a head appearing in one or more of the images, the sequence of images obtained by a camera while the media content corresponding to the content identifying data is presented by the television;
--

94. The combination of Lu-577 and Tian teaches this limitation. Lu-577 teaches that its television audience measurement system includes a person identifier for identifying persons watching the television program presented. Ex-1004, ¶0039 (“Moreover, a person identifier 98 may be provided in order to

identify the persons watching television programming on the digital television set 66. ... Similarly, person identifiers may be provided in order to identify the persons watching television programming on the analog television set 70 and on the personal computer 80.”). Lu-577 incorporates Lu-000’s teachings on using head location and face recognition software in identifying members of a viewing audience. Ex-1004, ¶0013 (“Additionally, it is well known in the audience measurement arts to use computer-based image recognition in order to identify members of a viewing audience. Notable among teachings in this area is that by Lu in U.S. Pat. No. 4,858,000. The teaching of this patent is herein incorporated by reference.”), ¶0039 (“[T]he site unit 86 may employ known head location and face recognition software (e.g., as taught by Lu in U.S. Pat. No. 4,858,000) for the identification of the viewing persons and for the collection of other demographic data.”).

95. The combination of Lu-577 (including Lu-000) and Tian teaches **obtaining a sequence of images by a camera while the media content (e.g., television program) corresponding to the content identifying data (e.g., data identifying the television program) is presented by the television.** Lu-577 teaches that the person identifier identifies “the persons watching television programming on the digital television set” or “the analog television set” based on image recognition. Ex-1004, ¶0039. Since the identified persons are “watching television

programming” on the television set, the images of the persons are captured while the media content (e.g., television program) is presented by the television. Ex-1004, ¶0039. Lu-577 further teaches that the images are captured by a camera. Ex-1004, ¶0039 (“The person identifier 98 may be video camera, an IR camera, or the like.”). As explained with respect to limitation [1c], Lu-577 teaches obtaining content identifying data corresponding to the presented media content. *Supra* §VI.A.2[1c]. Therefore, the images are obtained while the media content corresponding to the content identifying data is presented.

96. Lu-000 further teaches “[a]n image recognition method and system ... for identifying predetermined individual members of a viewing audience in a monitored area.” Ex-1008, Abstract. Lu-000 teaches obtaining (e.g., capturing) images for audience members in the monitored area using a video camera. *Id.*, 2:20-38 (“A video image is captured for each of the located individual audience members in the monitored area.”), 3:27-60 (“[T]he image recognition system 10 includes an audience scanner 12 for scanning and capturing an image of the viewing audience members within a monitored area ... The image recognition system 10 includes a scanner controller and signal processing circuit 18 ... for locating and capturing the images of the audience members within the monitored area[.]”), 3:61-4:8 (“Referring also to FIGS. 2 and 3, the audience scanner 12 includes a video camera 28 for capturing a video image[.]”), Abstract. Lu-000

teaches capturing multiple “images of the audience members.” Ex-1008, 3:27-60.

In light of Lu-000’s teachings of obtaining the images using a video camera, which a POSITA would have known to be capable of taking sequences of images over time, Lu-000 teaches or renders obvious obtaining **a sequence of images**. Because Lu-000 teaches a television audience (Ex-1008, 1:15-19), the images of audience members are obtained when media content is presented by a television.

97. Tian further teaches obtaining a sequence of images of human subjects using a camera. Ex-1005, 000005 (“The database contains 704 image sequences from 97 subjects. Subjects sat directly in front of the camera and performed a series of facial behaviors which were recorded in an observation room. Image sequences with in-plane and limited out-of-plane motion were included. The image sequences began with a neutral face and were digitized into 640x480 pixel arrays with either 8-bit gray-scale values. The length of the image sequences are varying from 9 to 47 frames.”). To the extent it is argued that Lu-577 does not teach “a **sequence** of images,” a POSITA would have found it obvious and been motivated to apply Tian’s teachings to Lu-577 such that a sequence of images is obtained for each person and analyzed. *Supra* §VI.A.1 (explaining why it would have been obvious to obtain and analyze a sequence of images; that analysis is incorporated here).

98. The combination of Lu-577 (including Lu-000) and Tian further teaches **analyzing a sequence of images of the media exposure environment to detect a head appearing in one or more of the images**. As explained above, the combination of Lu-577 and Tian teaches obtaining a sequence of images of the media exposure environment. Lu-577 teaches using head location software for identification of audience members. Ex-1004, ¶0039 (“[T]he site unit 86 may employ known *head location* and face recognition software (e.g., as taught by Lu in U.S. Pat. No. 4,858,000) for the identification of the viewing persons and for the collection of other demographic data.). Determining the location of a head in an image includes detecting a head appearing in the image. Lu-000 teaches analyzing images to detect a head appearing in an image and to determine the location of the head based on a “search raw picture for heads subroutine” and a “search head(s) subroutine.” Ex-1008, 7:39-47 (“The digital representation of the infrared image signal from the image acquisition block 88 corresponding to an identified direction of an audience member by the processor subsystem 14 is applied to a zooming and head location block 104. A gray-level subimage output G-Sub of the zooming and head location block 104 is applied to a normalization block 106.”), 9:47-52 (“FIG. 8H is a flow chart illustrating an operational function or mode 4 logical steps performed to locate the faces. The sequential steps begin with a search raw picture

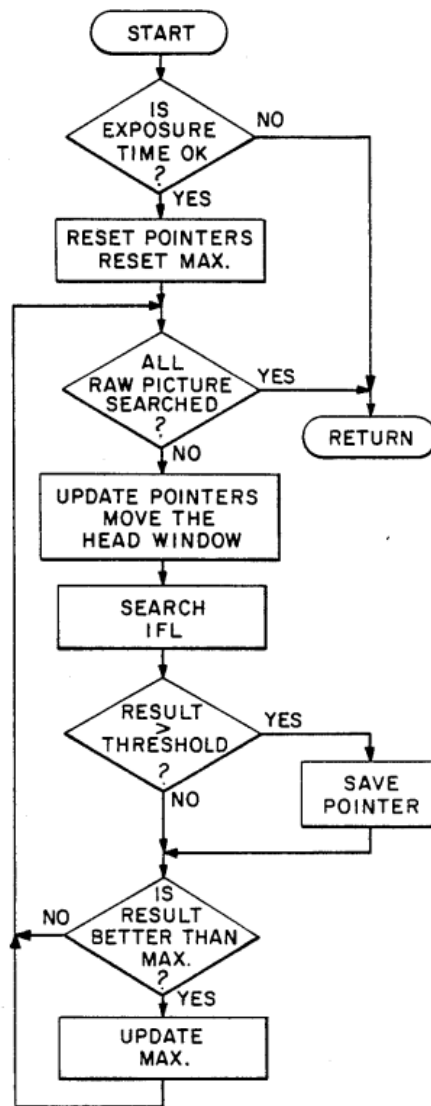
for heads subroutine illustrated in FIG. 8I. Next a locate and search head(s) subroutine illustrated in FIG. 8J is performed.”).

99. The search raw picture for head(s) subroutine searches for heads based on pattern image signatures and saves search area pointers that identify areas corresponding to heads:

The first search area matrix is identified and sequentially compared with each of the individual pattern image signatures (IFM block 112 of FIG. 7 and FIG. 8B). The result is compared with the set predetermined correlation threshold MAX value, and if the result is greater than the initial MAX value, then that search area pointer is saved and the MAX value is updated. Otherwise, the search area pointer is not saved and the MAX value is not changed. Then the search area pointer value is updated and the next search area matrix is identified and the sequential steps are repeated until the total raw picture has been searched.

Ex-1008, 9:53-10:3, FIG. 8I:

FIG. 8I

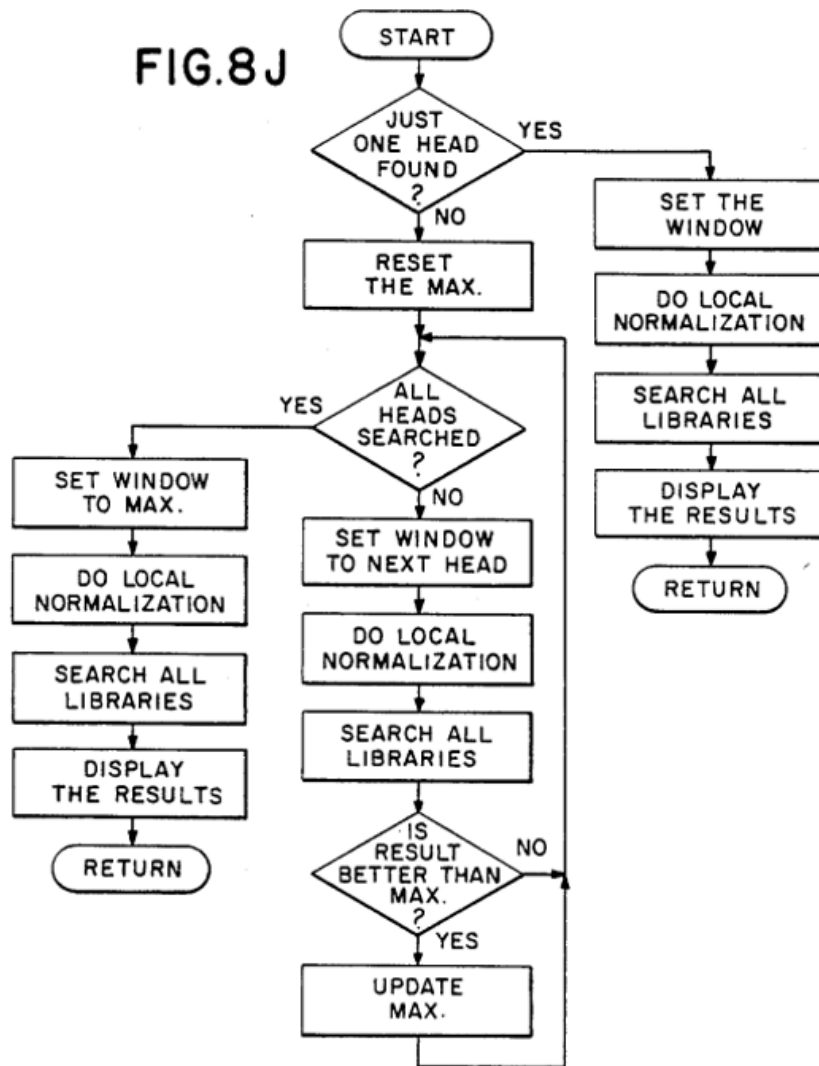


100. The locate and search head(s) subroutine sets a search area window based on the saved pointer:

FIG. 8J illustrates the locate and search head(s) subroutine performed by the control command processor subsystem 14 in the mode 4. If one search area pointer is stored in the subroutine of FIG. 8I, then the search area window is set to the identified search area matrix by the saved pointer value which corresponds to the head image portion. The

exposure time is adjusted and the search all libraries subroutine of FIG. 8E is performed and the results are displayed.

Ex-1008, 10:4-12, FIG. 8J:



101. Tian also teaches analyzing a sequence of images to detect a head appearing in one or more of the images using face detection techniques. Ex-1005, 000002 (“Many face detection methods have been developed to detect faces in an arbitrary scene....”), 000005 (“All images in 704 sequences from 97 subjects are

used for face detection and head pose estimation. The results show that both face detectors [23, 29] can detect all faces for head region in resolution of 288x384, 144x192, 72x96, and 36x48, but failed for resolution of 18x24.”). Therefore, the combination of Lu-577 and Tian teaches an audience measurement system analyze a sequence of images of the media exposure environment to detect a head appearing in one or more of the images, the sequence of images obtained by a camera while the media content corresponding to the content identifying data is presented by the television.

102. Furthermore, the '243 patent admits the prior art included approaches for identifying a number of people in a media exposure environment by discussing it in the “Background” section. Ex-1001, 1:56-2:3 (“In some audience measurement systems, the collected people data includes an amount of people being exposed to media content. To calculate the amount of people being exposed to the media content, some measurement systems capture a series of images of a media exposure environment (e.g., a television room, a family room, a living room, a bar, a restaurant, etc.) and analyze the images to determine how many people appear in the images at a particular date and time.”). A POSITA would have known that a common approach of identifying people was based on their heads.

[1e] determine an orientation of the head with respect to the camera; and

103. The combination of Lu-577 and Tian teaches this limitation. Tian teaches **determining an orientation of the head** (e.g., pose estimation) **with respect to the camera**. Ex-1005, 000001 (“In order to handle large head motion, head finding, head tracking and *pose estimation* can be applied to a facial expression analysis system.”), 000002 (“To handle the full range of head motion, we detected the head instead of the face. Then the head pose was estimated based on the detected head.”), 000002 (“In order to handle the full range of head motion for expression analysis in real environments, we also evaluate head pose detection for different image resolutions.”). Tian explains example methods of head pose estimation, which determines the orientation of the head:

After the head is located, the head image is converted to gray scale, histogram equalized and resized to 32x32. Then a three-layer neural network (NN) is employed to estimate the head pose. The inputs to the network are the processed head image. The outputs are the 3 head poses: 1) frontal or near frontal view, 2) side view or profile, 3) others such as back of the head or occluded face (see Table 2).

Ex-1005, 000002. As Tian explains, the input for the head pose estimation is a head image and the output is a head pose that describes the view of the head in the head image. Ex-1005, 000002. A POSITA would have understood that Tian’s

head pose describes the orientation of the head with respect to the camera that obtained the head image.

104. Tian provides examples of head poses:

Table 2: The definitions and examples of the 3 head pose classes: 1) frontal or near frontal view, 2) side view or profile, 3) others such as back of the head or occluded faces. The expression analysis process is applied to only the frontal and near frontal view faces.

Poses	Frontal or near frontal	Side view or profile	Others
Definitions	Both eyes and lip corners are visible	One eye or one lip corner is occluded	Not enough facial features
Examples			

Ex-1005, 000003.

105. A POSITA would have found it obvious to apply Tian's teaching of determining head poses to Lu-577 such that the audience measurement system determines an orientation of the head with respect to the camera before performing further steps to identify the corresponding person. *Supra* §VI.A.1 (explaining why it would have been obvious to perform head pose estimation; that analysis is incorporated here). Therefore, the combination of Lu-577 and Tian teaches an

audience measurement system determining an orientation of the head with respect to the camera.

[1f] determine audience identification information based on a match of the head to a known person associated with the media exposure environment; and
--

106. Lu-577 teaches this limitation. Lu-577 teaches that its television audience measurement system includes a person identifier for identifying persons watching the television program presented. Ex-1004, ¶0039 (“Moreover, a person identifier 98 may be provided in order to identify the persons watching television programming on the digital television set 66. ... Similarly, person identifiers may be provided in order to identify the persons watching television programming on the analog television set 70 and on the personal computer 80.”). Lu-577 incorporates Lu-000’s teachings on using head location and face recognition software in identifying members of a viewing audience. Ex-1004, ¶0013 (“Additionally, it is well known in the audience measurement arts to use computer-based image recognition in order to identify members of a viewing audience. Notable among teachings in this area is that by Lu in U.S. Pat. No. 4,858,000. The teaching of this patent is herein incorporated by reference.”), ¶0039 (“[T]he site unit 86 may employ known head location and face recognition software (e.g., as taught by Lu in U.S. Pat. No. 4,858,000) for the identification of the viewing persons and for the collection of other demographic data.”).

107. Lu-000 teaches **determining audience identification information based on a match of the head to a known person** (e.g., predetermined individual) **associated with the media exposure environment** (e.g., monitored area). Lu-000 teaches “identifying predetermined individual members of a viewing audience in a monitored area.” Ex-1008, Abstract, 2:10-14, 2:20-38. Lu-000 teaches storing a pattern image signature corresponding to each predetermined individual member of the viewing audience. Ex-1008, Abstract (“A pattern image signature is stored corresponding to each predetermined individual member of the viewing audience to be identified.”), 2:20-47 (“In accordance with a feature of the invention, each of the plurality of feature image signatures are stored in a distinct memory space of a predetermined capacity. Each of the feature image signatures is generated by processing a plurality of video images of the face of the individual to be identified. A signature from each of the processed video images is extracted and stored in the corresponding predetermined memory space for the particular feature image signature.”). As explained above, Lu-000’s monitored area is a media exposure environment. *Supra* §VI.A.2.1[p]. Lu-000 thus teaches storing pattern image signatures of known person associated with the media exposure environment.

108. Lu-000 next teaches identifying an audience member by comparing a newly extracted pattern image signature with the stored pattern image signatures:

A video image is captured for each of the located individual audience members in the monitored area. A pattern image signature is extracted from the captured image. The extracted pattern image signature is compared with each of the stored pattern image signatures to identify a particular one of the predetermined audience members. These steps are repeated to identify all of the located individual audience members in the monitored area.

Ex-1008, Abstract, 2:20-38.

109. Lu-000 teaches that the newly extracted pattern image signature is generated based on a “B-Sub image” that corresponds to the head region of a person:

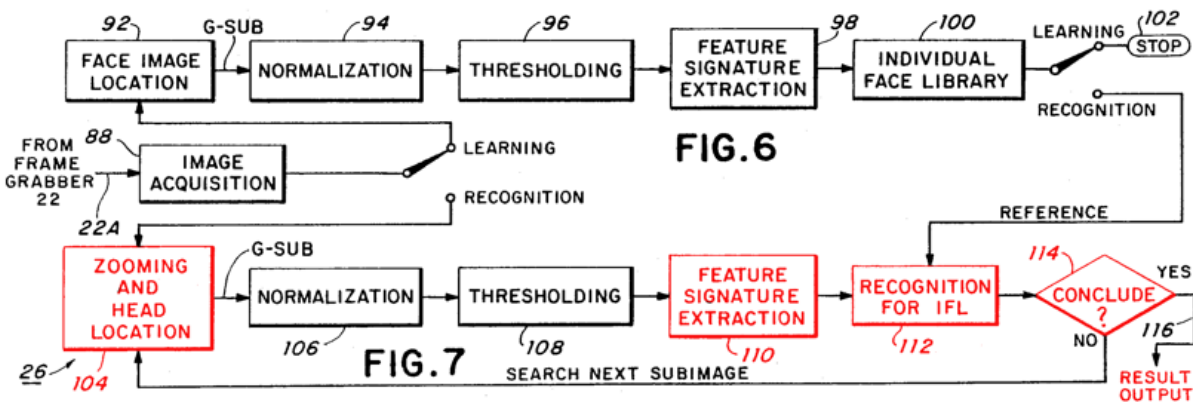
A gray-level subimage output G-Sub of the zooming and **head location** block 104 is applied to a normalization block 106. The normalized output of block 106 is applied to a thresholding block 108 to provide a thresholding, binary level image output B-Sub. A feature signature is extracted from the B-Sub image at a feature signature extraction block 110.

Ex-1008, 7:39-8:6 (emphasis added). The extracted pattern image signature is then compared with stored pattern image signatures to identify an audience member:

The extracted B-Sub image feature signature is compared with each of the pattern image signatures stored in the individual face library as illustrated at a block 112 to identify a match. A particular audience member is identified when the compared signatures exceed a predetermined correlation threshold value corresponding to the best

matching rate or highest correlation with an individual face pattern image signature in the individual face library is identified at a conclusion block 114. ... The satisfactory conclusions include both “identified” and “unidentified” audience members. The logic steps performed for recognizing the particular individual audience member are described with respect to FIG. 8B. An output signal at a line 116 is then stored corresponding to the particular identified individual member of the viewing audience. The thus identified individual viewing member data can be stored together with other parameter data of a television data collection system, such as channel reception of a monitored receiver.

Id., FIGs. 6-7 (emphasis added):



Therefore, Lu-000 teaches that the identification of the audience member is based on a match of the head to a known person.

110. Lu-000 teaches comparing an extracted image feature signature with an image feature signature stored in an individual face storage library (IFL) and determining a match based on an N-found value that reflects the similarity between

the signatures. Ex-1008, 8:46-9:2, FIG. 8B. Lu-000 further teaches searching through the libraries to identify the stored image feature signature having the maximum N-found value.

[A]n initial MAX value is set for the predetermined N-found value. A first library is searched (block 112 of FIG. 7 and FIG. 8B) and if the result N-found value is greater than the initial MAX value, then the MAX value is updated. Otherwise the MAX value is not changed. Then a next library is searched and the result is compared to the resulting MAX value and adjusted, until all the libraries have been searched.

Ex-1008, 9:18-31, FIG. 8E. After searching through the libraries, a result is displayed and is verifiable by a person. Ex-1008, 9:32-39 (“Next the search all libraries subroutine illustrated, in FIG. 8E is performed. The results are displayed, verified by a user operator of the system 10, then added to an identified correct library.”), 9:40-46, 9:3-10, 10:4-32, FIG. 8C, 8F-8G, 8J. A POSITA would have understood or found it obvious that the displayed result indicates the image feature signature that has the maximum N-found value or information about the corresponding individual. Such a result includes identification information because it identifies a predetermined individual member of the viewing audience, as explained above. Therefore, Lu-577 teaches an audience measurement system determining audience identification information based on a match of the head to a known person associated with the media exposure environment.

111. Furthermore, the '243 patent admits that audience identification based on facial recognition was known in the prior art by discussing it in the “Background” section. Ex-1001, 1:56-2:3 (“Additionally, some audience measurement systems identify people in the images via one or more identification techniques such as, for example, facial recognition.”).

[1g] network interface circuitry to output a signal indicative of the content identifying data and the audience identification information to a data collection facility.

112. The combination of Lu-577 and Tian teaches **network interface circuitry** (e.g., interface and communication apparatus) **to output a signal indicative of the content identifying data and the audience identification information to a data collection facility** (e.g., intermediate data collector, data collection central office). As explained with respect to limitations [1b]-[1f], the combination of Lu-577 and Tian teaches obtaining content identifying data and determining audience identification information. *Supra* §VI.A.2[1b-1f]. Lu-577 teaches the television audience measurement system analyzing data to determine content identifying data and audience identification information locally in the statistically selected site. Ex-1004, ¶0040 (“[C]haracteristic signatures can be extracted from these representations for subsequent comparison ... in the statistically selected monitoring site 62 ... with previously collected reference signatures in order to identify the television programs to which the digital

television set 66, the analog television set 70, and the personal computer 80 are tuned.”); ¶0039 (“When such equipment is available in the statistically selected monitoring site 62, the site unit 86 may employ known head location and face recognition software ... for the identification of the viewing persons and for the collection of other demographic data.”).

113. Lu-577 further teaches communicating television audience data to an intermediate data collector or a data collection central office. Ex-1004, ¶0050 (“The software agents 112, 118, and 122 may be telecommunication enabled so that television audience data from all the viewing sites in the statistically selected monitoring site 102 can be communicated, via the Internet, a public telephone system, or the like, to a locally located or remotely located intermediate data collector 128 and then to a remotely located central office 130 through a communication channel 170.”), ¶0037 (“The remotely located data collection central office 90 is schematically depicted as a computer based data collection central office that interchanges data with the home unit 88 over a communication network 92, such as a public switched telephone network, the Internet, or the like.”). A POSITA would have understood that the television audience data in Lu-577 includes television content identifying data and audience identification information, which are communicated to the intermediate data collector or data collection central office. A POSITA would have understood that data commonly

transmit over networks in the form of electromagnetic signals, therefore Lu-577 teaches outputting a signal to a data collection facility.

114. The '243 patent does not include any limitation on the meaning of a “data collection facility.” A POSITA would have understood that the intermediate data collector and the data collection central office in Lu-577 are data collection facilities because they collect and store data. Lu-577 teaches that the television audience measurement system includes “an interface and communication apparatus.” Ex-1004, cl. 13, ¶0015 (“The interface and communication apparatus is adapted to transmit the identification datum to a remotely located central office.”). The interface and communication apparatus includes network communication features and functionalities such as “a serial port,” “a parallel port,” “a universal serial bus,” “a firewire,” or the ability to send data “to an internet service provider via the internet.” Ex-1004, cls. 29-33. A POSITA would have understood that Lu-577’s interface and communication apparatus comprises a network interface circuitry or that it is at least obvious to implement the interface and communication apparatus using a circuitry. A POSITA would have further understood that the interface and communication apparatus is suitable and obvious to be used for outputting the content identifying data and audience identification information to the intermediate data collector or the data collection central office. Therefore, the combination of Lu-577 and Tian teaches an audience measurement

system comprising a network interface circuitry to output a signal indicative of the content identifying data and the audience identification information to a data collection facility.

3. Claim 4

115. It is my opinion that a POSITA would have found claim 4 obvious based on the combined teachings of Lu-577 and Tian.

4[p]. The audience measurement system of claim 1, wherein the processor circuitry is to:
--

116. The combination of Lu-577 and Tian teaches the audience measurement system of claim 1. *Supra* §VI.A.2. As explained below, the combination of Lu-577 and Tian further teaches that the audience measurement system performs the further operations recited in claim 4. *Infra* §VI.A.3[4a-4b]. Furthermore, a POSITA would have found it obvious to implement Lu-577's television audience measurement system such that the processor circuitry executes the machine readable instructions to perform operations directed to identifying a television program and audience members. *Supra* §VI.A.2[1a]. It would have been obvious to use the processor circuitry to perform the operations of limitations [4a] and [4b].

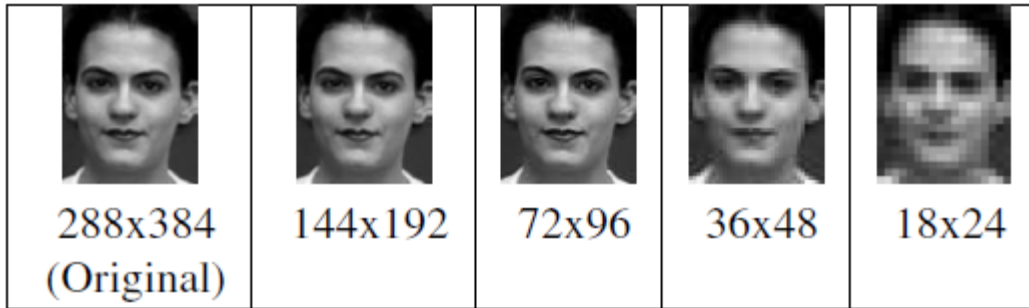
[4a] reduce a resolution of a first image of the one or more of the images of the media exposure environment to obtain a reduced-resolution image, and
--

117. The combination of Lu-577 and Tian teaches that the processor circuitry of the audience measurement system **reduces 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.**

118. Tian recognizes that “in real environments, the face image is often in lower resolution” and describes the effects of different image resolutions on face image analysis. Ex-1005, Abstract (“This paper explores the effects of different image resolutions for each step of facial expression analysis. The different approaches are compared for face detection, face data extraction and expression recognition.”); 000002 (“However, in real applications, the face region is often in lower resolutions.”).

119. Tian creates lower-resolution images by down-sampling from original images and analyzes multiple different resolutions. Ex-1005, Abstract (“A total of five different resolutions of the head region are studied (288x384, 144x192, 72x96, 36x48, and 18[]x24) based on a widely used public database [16]. The lower resolution images are down-sampled from the originals.”). Tian illustrates example images in reduced resolutions:

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.



Ex-1005, 000002 (“As shown in Table 1, total five different resolutions of the head region are studied (288x384, 144x192, 72x96, 36x48, and 18x24) based on Cohn-Kanade AU-Coded Face Expression Image Database [16]. The lower resolution images are down-sampled from the originals.”). Therefore, Tian teaches reducing the resolution of an image to obtain a reduced-resolution image.

120. As explained above, the combination of Lu-577 and Tian teaches obtaining a sequence of images of the media exposure environment, detecting a head in one or more of the images, and determining an orientation of the head. *Supra* §VI.A.2[1d-1e]. A POSITA would have found it obvious to apply Tian’s teaching to Lu-577 such that the television audience measurement system reduces the resolution of an image and performs face detection or head pose estimation on the reduced-resolution image. *Supra* §VI.A.1 (explaining why it would have been obvious to analyze reduced-resolution images; that analysis is incorporated here).

Therefore, the combination of Lu-577 and Tian teaches the television audience measurement system reducing 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.

[4b] determine the orientation of the head with respect to the camera based on the reduced-resolution image.
--

121. The combination of Lu-577 and Tian teaches that the processor circuitry of the audience measurement system **determines the orientation of the head with respect to the camera** (e.g., pose estimate) **based on the reduced-resolution image**. As explained above, Tian’s disclosure about head pose estimation teaches determining an orientation of the head with respect to the camera. *Supra* §VI.A.2[1e]. Tian teaches determining a pose estimation based on a reduced-resolution image. Ex-1005, 000005 (“[H]ead pose estimation can detect 98% faces for all the resolutions. Our previous results of head pose estimation showed that the head pose estimation can achieve 96% accuracy for 5 poses when head region resolution is about 8x8 pixels[.]”). Tian provides a summary of the accuracy level of head pose estimation for reduced-resolution images at different resolutions:

Table 3: Summary of the effects of faces at different resolutions for expression analysis. For face acquisition, "FD" indicates face detector. "HPE" indicates head pose estimation. For feature extraction, "G1" indicates geometric features extracted by feature tracking. "G2" indicates geometric features extracted by feature detection. "AP" indicates appearance features extracted by Gabor wavelets.

Face Process						
		288x384 (Original)	144x192	72x96	36x48	18x24
Face Acquisition	FD	100%	100%	100%	100%	0%
	HPE	98.5%	98%	98.2%	97.8%	98%
Feature Extraction	G1	Yes	Yes	Yes	No	No
	G2	Yes	Yes	Yes	Yes	No
	AP	Yes	Yes	Yes	Yes	Yes
FACS AUs	G1	90%	90.2%	89.8%	N/A	N/A
	G2	71%	70.8%	72%	54.3%	N/A
	AP	90.7%	90.2%	89.6%	72.6%	58.2%
	G1+AP	92.8%	93%	92.2%	N/A	N/A
	G2+AP	91.2%	90.8%	90%	87.7%	N/A
Basic Expressions	G1	92.5%	91.8%	91.6%	N/A	N/A
	G2	74%	73.8%	72.9%	61.3%	N/A
	AP	91.7%	92.2%	91.6%	77.6%	68.2%
	G1+AP	93.8%	94%	93.5%	N/A	N/A
	G2+AP	93.2%	93%	92.8%	89%	N/A

Ex-1005, 000005 (emphasis added).

122. As shown in the above table, the accuracy of head pose estimation remains high at relatively low resolutions (e.g., 98% at 18x24 compared with 98.5% at 288x384). A POSITA would have found it obvious to apply Tian's teachings of determining the orientation of the head with respect to the camera

based on the reduced-resolution image to the television audience measurement system of Lu-577. *Supra* §VI.A.1 (explaining why it would have been obvious to analyze reduced-resolution images; that analysis is incorporated here). A POSITA would have been motivated to do so because performing the head pose estimation on the reduced-resolution image would have saved processing resources and time without substantially reducing the accuracy of the determination. Therefore, the combination of Lu-577 and Tian teaches the audience measurement system determining the orientation of the head with respect to the camera based on the reduced-resolution image.

4. Claim 5

123. It is my opinion that a POSITA would have found claim 5 obvious based on the combined teachings of Lu-577 and Tian.

5. The audience measurement system of claim 4, wherein the processor circuitry is to determine the audience identification information by: (i) generating a facial signature from a region of a second image of the one or more of the images corresponding to a location of the head in the reduced-resolution image, and (ii) comparing the generated facial signature to a database of facial signatures.

124. The combination of Lu-577 and Tian teaches the audience measurement system of claim 4. *Supra* §VI.A.3. As explained below, the combination of Lu-577 and Tian further teaches that the audience measurement system performs the further operations recited in claim 5. Furthermore, a POSITA would have found it obvious to implement Lu-577's television audience

measurement system such that the processor circuitry executes the machine readable instructions to perform operations directed to identifying a television program and audience members. *Supra* §VI.A.2[1a]. It would have been obvious to use the processor circuitry to perform the operations of claim 5.

125. Lu-000 teaches **generating a facial signature** (e.g., pattern image signature) **from a region of an image corresponding to a location of the head.** *Supra* §VI.A.2[1f]. Lu-000 teaches extracting a pattern image signature from an image:

A video image is captured for each of the located individual audience members in the monitored area. A pattern image signature is extracted from the captured image. The extracted pattern image signature is compared with each of the stored pattern image signatures to identify a particular one of the predetermined audience members. These steps are repeated to identify all of the located individual audience members in the monitored area.

Ex-1008, Abstract, 2:20-38.

126. Lu-000's pattern image signature is extracted from a region of an image corresponding to a location of the head:

A gray-level subimage output G-Sub of the zooming and **head location** block 104 is applied to a normalization block 106. The normalized output of block 106 is applied to a thresholding block 108 to provide a thresholding, binary level image output B-Sub. A feature signature is

extracted from the B-Sub image at a feature signature extraction block 110.

Ex-1008, 7:39-8:6 (emphasis added). The pattern image signature is a facial signature because it represents the features of a person's face as it is extracted from the portion of the image corresponding to the person's head/face and is alternatively called a "face pattern image signature" in Lu-000. Ex-1008, 6:49-51 ("An extracted pattern image signature output of the feature signature extraction block 98 is stored in an individual face storage library (IFL) 100."), 7:25-27 ("A distinct memory space of a predetermined capacity is defined for each of the individual face pattern image signatures within the image face library IFL block 100."), 7:39-8:6 ("A particular audience member is identified when the compared signatures exceed a predetermined correlation threshold value corresponding to the best matching rate or highest correlation with an individual face pattern image signature in the individual face library is identified at a conclusion block 114.").

127. As the '243 patent does not limit the method used to generate the facial signature, a POSITA would have understood that the terms "extracting" and "generating" are often used interchangeably to refer to the creation of a facial signature based on an image. Extracting a pattern image signature based on a processed region of an image corresponding to the face of a person is at least one

way to generate a facial signature. Therefore, Lu-000 teaches generating a facial signature from a region of an image corresponding to a location of the head.

128. The combination of the teachings of Lu-000 and Tian teaches generating the facial signature from a region of **a second image of the one or more of the images** corresponding to a location of the head **in the reduced-resolution image**. Tian teaches generating a reduced-resolution image by reducing the resolution of a first image of the one or more of the images. *Supra* ¶VI.A.3[4a]. A POSITA would have been motivated to perform face detection (e.g., determining the location of the face or head) on a reduced-resolution image. *Supra* §VI.A.1 (explaining why it would have been obvious to analyze reduced-resolution images; that analysis is incorporated here). As Tian explains, face detection can be performed on a reduced-resolution image with high accuracy. Ex-1005, 000005 (“The results show that both face detectors [23, 29] can detect all faces for head region in resolution of 288x384, 144x192, 72x96, and 36x48....”). Therefore, a POSITA would have been motivated to detect a face/head location on a reduced-resolution image to save processing power before performing face recognition to identify a person. *Supra* §VI.A.1.

129. Furthermore, Tian teaches determining a face/head location in a first image and only tracking changes to face features in subsequent images. Ex-1005, 000001 (“Face acquisition is a processing stage to automatically find the face

region for the input images or sequences. It can be a face detector to detect a face in each frame or **just detect face in the first frame and then track the face in the remainder of the video sequence.**” (emphasis added)), 000003 (“Given an image sequence, the region of the face and approximate location of individual face features are detected automatically in the initial frame. ... Then, all face feature changes are automatically tracked in the image sequence.”). Under this approach, if no change is detected, the face is determined to be at the same location in subsequent images.

130. A POSITA would have found it obvious to apply Tian’s teaching to Lu-577 such that when the television audience measurement system processes a sequence of images, it determines a face or head location in a first image and use the determined location for subsequent images. *Supra* §VI.A.1 (explaining why it is obvious to determine a face or head location in a first image and use the location for subsequent images; that analysis is incorporated here). Further in light of Tian’s teachings of analyzing reduced-resolution images, a POSITA would have found it obvious to determine the location of a person’s face in a reduced-resolution version of a first image and use this location to perform facial recognition based on the sequence of higher-resolution images because doing so would have saved processing resources without sacrificing performance. *Supra* §VI.A.1 (explaining why it is obvious to perform face detection in a reduced-

resolution image; that analysis is incorporated here). In this manner, after a location of the head is determined in a reduced-resolution image generated from a first image, facial recognition is performed on a region corresponding to the location in one or more images of the sequence, including a second image.

131. Lu-000 further teaches **comparing the generated facial signature** (e.g., pattern image signature) **to a database of facial signatures** (e.g., individual face storage library (IFL)). *Supra* §VI.A.2[1f]. Lu-000 teaches comparing the extracted pattern image signature to the pattern image signatures stored in an individual face storage library (IFL). Ex-1008, 7:39-8:6 (“The extracted B-Sub image feature signature is compared with each of the pattern image signatures stored in the individual face library as illustrated at a block 112 to identify a match.”), FIG. 7; *see also* Ex-1008, 8:46-9:31, FIGs. 8B, 8E, 8G; *supra* §VI.A.2[1f]. A POSITA would have understood that Lu-000’s individual face storage library is a database because it includes a structured collection of data (e.g., pattern image signatures) stored in memory. Ex-1008, 7:25-38 (“A distinct memory space of a predetermined capacity is defined for each of the individual face pattern image signatures within the image face library IFL block 100.”).

132. Lu-577 (including Lu-000) further teaches **determining the audience identification information** by generating the facial signature and comparing the generated facial signature to a database of facial signatures. *Supra* §VI.A.2[1f];

Ex-1004, ¶0039 (“[A] person identifier 98 may be provided in order to identify the persons watching television programming on the digital television set 66.”); Ex-1008, Abstract (“A pattern image signature is extracted from the captured image. The extracted pattern image signature is compared with each of the stored pattern image signatures to identify a particular one of the predetermined audience members.”), 2:20-38. Therefore, the combination of Lu-577 and Tian teaches the limitations of claim 5.

5. Claim 6

133. It is my opinion that a POSITA would have found claim 6 obvious based on the combined teachings of Lu-577 and Tian.

6[p]. The audience measurement system of claim 4, wherein the processor circuitry is to:

134. The combination of Lu-577 and Tian teaches the audience measurement system of claim 4. *Supra* §VI.A.3. As explained below, the combination of Lu-577 and Tian further teaches that the audience measurement system performs the further operations recited in claim 6. Furthermore, a POSITA would have found it obvious to implement Lu-577’s television audience measurement system such that the processor circuitry executes the machine readable instructions to perform operations directed to identifying a television program and audience members. *Supra* §VI.A.2[1a]. It would have been obvious to use the processor circuitry to perform the operations of limitations [6a]-[6c].

[6a] identify a region corresponding to the head within the first image from which the reduced-resolution image was obtained;

135. The combined disclosure of Lu-000 and Tian teaches **identifying a region corresponding to the head within an image**. *Supra* §VI.A.2[1d]. Lu-000 teaches analyzing images to detect a region corresponding to the head. *Supra* §VI.A.2[1d]; Ex-1008, 7:39-47, 9:47-52, 9:53-10:3, 10:4-12, FIGs. 8I-8J. Furthermore, Tian teaches determining a face/head location in a first image and use the location for subsequent images unless a change is detected. *Supra* §VI.A.4; Ex-1005, 000001, 000003.

136. Tian further teaches **obtaining a reduced-resolution image from the first image**. *Supra* ¶VI.A.3[4a]. Therefore, the combined disclosure of Lu-000 and Tian teaches identifying a region corresponding to the head within the first image from which the reduced-resolution image was obtained.

[6b] analyze a portion corresponding to the identified region of a second image; and based on the analysis of the region of the second image, determine that the head matches the known person.

137. The combined disclosure of Lu-000 and Tian teaches **analyzing a portion corresponding to the identified region of a second image; and based on the analysis of the region of the second image, determine that the head matches the known person**.

138. As explained with respect to claim 5, a POSITA would have found it obvious to apply Tian's teaching to Lu-577 such that when the television audience measurement system processes a sequence of images, it determines a face or head location in a first image and use the determined location for subsequent images. *Supra* §VI.A.4.

139. Lu-000 teaches analyzing a portion corresponding to the identified region of the head or face to determine that the head matches the known person based on the analysis (e.g., by extracting a pattern image signature from the portion and determining that the extracted pattern image signature matches a stored pattern image signature of the known person). *Supra* §VI.A.2[1f], §VI.A.4; Ex-1008, 7:39-8:6, FIG. 7; *see also* Ex-1008, 8:46-9:31, FIGs. 8B, 8E, 8G.

140. Therefore, the combined disclosure of Lu-000 and Tian teaches analyzing a portion corresponding to the identified region of a second image and based on the analysis of the region of the second image, determine that the head matches the known person.

6. Claim 8

141. It is my opinion that a POSITA would have found claim 8 obvious based on the combined teachings of Lu-577 and Tian.

8. The audience measurement system of claim 6, wherein the audience identification information includes an identifier associated with the known person and the processor circuitry is to, responsive to a determination that the head matches the known person, cause the identifier associated with the person to be stored in the memory.

142. The combination of Lu-577 and Tian teaches the audience measurement system of claim 6. *Supra* §VI.A.5. As explained below, Lu-577 further teaches that the audience measurement system performs the further operations recited in claim 8. Furthermore, a POSITA would have found it obvious to implement Lu-577's television audience measurement system such that the processor circuitry executes the machine readable instructions to perform operations directed to identifying a television program and audience members. *Supra* §VI.A.2[1a]. It would have been obvious to use the processor circuitry to perform the operations of claim 8.

143. Lu-577 teaches determining audience identification information based on a match of the head to a known person. *Supra* §VI.A.2[1f]. Lu-000 teaches outputting the identification information as an output signal and storing "identified individual viewing member data." Ex-1008, 7:67-8:6 ("An output signal at a line 116 is then stored corresponding to the particular identified individual member of the viewing audience. The thus identified individual viewing member data can be stored together with other parameter data of a television data collection system, such as channel reception of a monitored receiver."). A POSITA would have

understood that the stored individual viewing member data identifies the known person and thus at least includes an identifier associated with the person.

Furthermore, a POSITA would have found it obvious for the audience identification information to include an identifier associated with the known person given the broad meaning of the word “identifier.” In fact, the ’243 patent admits that the concept of an identifier was known in the prior art. Ex-1001, 1:43-55 (“Audience measurement of media content ... often involves collection of ... people data (e.g., **identifiers**, demographic data associated with audience members, etc.).” (emphasis added)).

144. Lu-577 teaches that the audience measurement system includes a memory. *Supra* §VI.A.2[1a]. Lu-000 teaches storing data related to facial recognition using a memory. Ex-1008, 2:39-47 (“In accordance with a feature of the invention, each of the plurality of feature image signatures are stored in a distinct memory space of a predetermined capacity.”), 5:61-6:4 (“The control command processor subsystem 14 includes a central processing unit 78, such as, an Intel 80286 high performance 16-bit microprocessor with integrated memory management and adapted for multi-tasking systems and an optional associated memory device 80.”), 7:25-38 (“All of the resulting extracted pattern image signatures for each of the audience members are stored in the particular corresponding memory space of the IFL memory spaces.”). A POSITA would

have found it obvious to cause the identifier associated with the person to be stored in memory. Therefore, Lu-577 teaches that the audience identification information includes an identifier associated with the known person and, responsive to a determination that the head matches the known person, causing the identifier associated with the person to be stored in the memory.

7. Claim 9

145. It is my opinion that a POSITA would have found claim 9 obvious based on the combined teachings of Lu-577 and Tian.

9[p]. An audience measurement system, comprising:
--

146. As explained for claim 1, Lu-577 teaches an audience measurement system. *Supra* §VI.A.2.1[p].

[9a] an audience measurement device at a media exposure environment, the audience measurement device to:

147. As explained for claim 1, Lu-577 teaches that the audience measurement system includes components such as audio detectors, cameras, and site unit/home unit that analyzes audio and image data to identify a television program and members of the audience. *Supra* §VI.A.2[1a]. Lu-577 further teaches implementing these components of the television audience measurement system in the statistically selected site. *Supra* §VI.A.2[1a]; Ex-1004, ¶¶0035-0040, FIG. 2.

148. In light of their functionalities, a combination of the components in the statistically selected site constitutes **an audience measurement device**, as the term is used in the '243 patent. The '243 patent describes example implementations of the audience measurement device, but clarified that the audience measurement device is broad and not limited to the example implementations. Ex-1001, 8:7-43 (“While an example manner of implementing the audience measurement device 104 of FIG. 1 has been illustrated in FIG. 2, one or more of the elements, processes and/or devices illustrated in FIG. 2 may be combined, divided, re-arranged, omitted, eliminated and/or implemented in any other way.”). Because the components of the audience measurement system in the statistically selected site perform operations that measure exposure of an audience to media content, their combination constitutes an audience measurement device. Ex-1004, ¶¶0035-0040. To the extent it is required that the audience measurement device is integrated into a single body, a POSITA would have found it obvious to integrate the components of Lu-577 in such a manner. For example, a POSITA would have found it obvious to integrate the structure or functionality of an audio detector and a camera on a site unit to form a device that is implemented in a single body. Such a device is an audience measurement device. As explained for claim 1, the statistically selected site is a media exposure environment. *Supra*

§VI.A.2.1[p]. Therefore, Lu-577 teaches an audience measurement device *at a media exposure environment*.

149. As explained above, a POSITA would have found it obvious to implement Lu-577's television audience measurement system, including its components such as the site unit and home unit, using memory, machine readable instructions, and processor circuitry, such that the processor circuitry executes the machine readable instructions to perform operations directed to identifying a television program and audience members. *Supra* §VI.A.2[1a]. Because Lu-577 teaches that the audience measurement device includes these components of the television audience measurement system, a POSITA would have found it obvious to implement the audience measurement device using the structure of memory, machine readable instructions, and processor circuitry, such that the processor circuitry executes the machine readable instructions to perform operations directed to identifying a television program and audience members. As explained below, the combination of Lu-577 and Tian teaches the audience measurement device performing the operations recited in limitations [9b]-[9g]. *Infra* §VI.A.7[9b-9g].

[9b] generate an audio signature of media content presented by a television within the media exposure environment;

150. As explained for claim 1, Lu-577 teaches that the processor circuitry of the audience measurement system executes the machine readable instructions to

generate an audio signature of media content presented by a television within the media exposure environment. *Supra* §VI.A.2[1b]. The audience measurement device includes the processor circuitry and machine readable instructions. *Supra* §VI.A.7[9a]. Therefore, Lu-577 teaches that the audience measurement device generates an audio signature of media content presented by a television within the media exposure environment.

[9c] obtain content identifying data corresponding to the presented media content, the content identifying data based on the audio signature;

151. As explained for claim 1, Lu-577 teaches that the processor circuitry of the audience measurement system executes the machine readable instructions to obtain content identifying data corresponding to the presented media content, the content identifying data based on the audio signature. *Supra* §VI.A.2[1c]. The audience measurement device includes the processor circuitry and machine readable instructions. *Supra* §VI.A.7[9a]. Therefore, Lu-577 teaches that the audience measurement device obtains content identifying data corresponding to the presented media content, the content identifying data based on the audio signature.

[9d] while the media content corresponding to the content identifying data is presented by the television, collect a first image of the media exposure environment with a camera;

152. As explained for claim 1, the combination of Lu-577 and Tian teaches that the processor circuitry of the audience measurement system executes the

machine readable instructions to obtain a sequence of images of the media exposure environment with a camera while the media content corresponding to the content identifying data is presented by the television. *Supra* §VI.A.2[1d]. The combination of Lu-577 and Tian further teaches detecting a head appearing in one or more of the images. *Supra* §VI.A.2[1d]. A POSITA would have understood that “**collect**” is used interchangeably with “obtain” in the context of a camera capturing images. The one or more images in which a head is detected includes at least one image, which constitutes a **first image** that is collected. The audience measurement device includes the processor circuitry and machine readable instructions. *Supra* §VI.A.7[9a]. Therefore, the combination of Lu-577 and Tian teaches that, while the media content corresponding to the content identifying data is presented by the television, the audience measurement device collects a first image of the media exposure environment with a camera.

[9e] attempt to detect a head in the first image;

153. As explained for claim 1, the combination of Lu-577 and Tian teaches that the processor circuitry of the audience measurement system executes the machine readable instructions to detect a head in an image. *Supra* §VI.A.2[1d]. This approach is applied to a sequence of images such that a head is detected in one or more of the images. *Supra* §VI.A.2[1d]. Because this approach does not guarantee detecting a head in each image, it includes attempts to detect a head. As

the first image is satisfied by one of the one or more images (*supra* §VI.A.7[9d]), it follows that the combination of Lu-577 and Tian teaches **attempting to** detect the head in the first image. The audience measurement device includes the processor circuitry and machine readable instructions. *Supra* §VI.A.7[9a]. Therefore, the combination of Lu-577 and Tian teaches that the audience measurement device attempts to detect a head in the first image.

[9f] determine an orientation of the head with respect to the camera;

154. As explained for claim 1, the combination of Lu-577 and Tian teaches that the processor circuitry of the audience measurement system executes the machine readable instructions to determine an orientation of the head with respect to the camera. *Supra* §VI.A.2[1e]. The audience measurement device includes the processor circuitry and machine readable instructions. *Supra* §VI.A.7[9a]. A POSITA would have found it obvious for the audience measurement device to determine the head orientation of the audience, which is part of audience information. Therefore, the combination of Lu-577 and Tian teaches that the audience measurement device determines an orientation of the head with respect to the camera.

[9g] determine audience identification information based on an indication that the head matches a known person associated with the audience measurement system; and

155. As explained for claim 1, Lu-577 teaches that processor circuitry of the audience measurement system executes the machine readable instructions to determine audience identification information based on a match of the head to a known person associated with the media exposure environment. *Supra* §VI.A.2[1f]. Lu-000, which is incorporated by reference in Lu-577, teaches comparing an extracted pattern image signature with stored pattern image signatures to determine a match. Ex-1008, 7:39-8:6 (“The extracted B-Sub image feature signature is compared with each of the pattern image signatures stored in the individual face library as illustrated at a block 112 to identify a match. A particular audience member is identified when the compared signatures exceed a predetermined correlation threshold value corresponding to the best matching rate or highest correlation with an individual face pattern image signature in the individual face library is identified at a conclusion block 114.”), 9:18-31. Lu-000 also teaches outputting and storing a signal and displaying a result after a match is found between the signatures, which indicates that the head matches a known person. Ex-1008, 7:39-8:6 (“An output signal at a line 116 is then stored corresponding to the particular identified individual member of the viewing audience.”), 9:40-46 (“Next the search all libraries subroutine illustrated in FIG.

8E is performed. The results are displayed.”). These disclosure of Lu-000 teaches **an indication that the head matches a known person** (e.g., signal, result, or stored value indicating match), on which audience identification is based.

156. Lu-577 teaches that the audience measurement system is located in and obtains exposure data for a media exposure environment. *Supra* §VI.A.2.1[p]. Because the audience measurement system is set up to monitor the media exposure environment, known persons associated with the media exposure environment are associated with the audience measurement system. Because Lu-577 teaches determining a match with a known person associated with the media exposure environment, it teaches determining the match with **a known person associated with the audience measurement system**. The audience measurement device includes the processor circuitry and machine readable instructions. *Supra* §VI.A.7[9a]. Therefore, Lu-577 teaches that the audience measurement device determines audience identification information based on an indication that the head matches a known person associated with the audience measurement system.

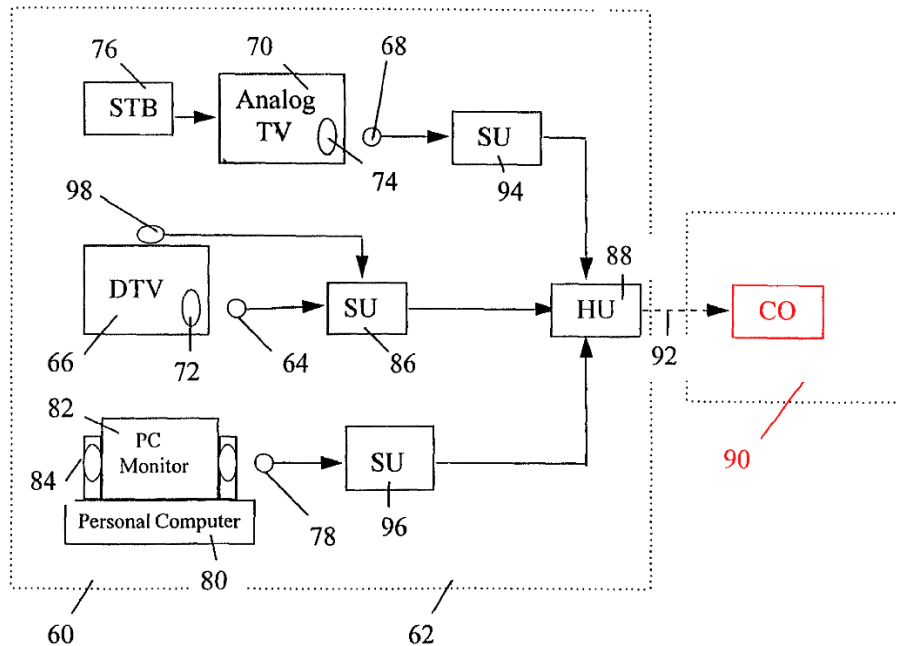
[9h] a data collection facility to obtain the content identifying data and the audience identification information from the audience measurement device.

157. As explained for claim 1, the combination of Lu-577 and Tian teaches a data collection facility to which a network interface circuitry outputs a signal indicative of the content identifying data and the audience identification

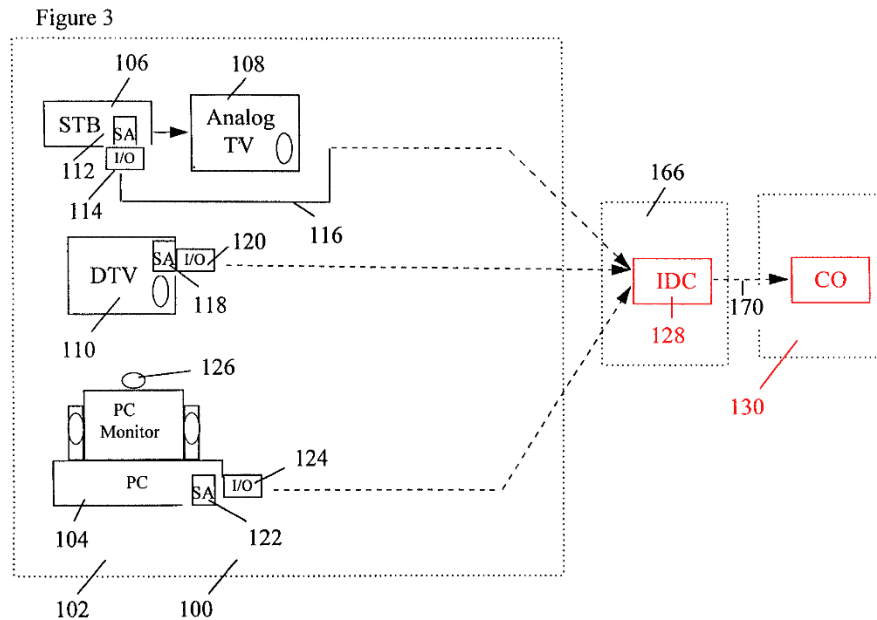
information. *Supra* §VI.A.2[1g]. Lu-577 teaches analyzing data to determine content identifying data and audience identification information locally in the statistically selected site and communicating such data to an intermediate data collector or a data collection central office, which teaches a data collection facility. *Supra* §VI.A.2[1g].

158. Lu-577 further teaches that its television audience measurement system includes the intermediate data collector and data collection central office. Ex-1004, ¶0037 (“The remotely located data collection central office 90 is schematically depicted as a computer based data collection central office that interchanges data with the home unit 88 over a communication network 92, such as a public switched telephone network, the Internet, or the like.”), FIG. 2 (emphasis added):

Figure 2



¶0050 (“The software agents 112, 118, and 122 may be telecommunication enabled so that television audience data from all the viewing sites in the statistically selected monitoring site 102 can be communicated, via the Internet, a public telephone system, or the like, to a locally located or remotely located intermediate data collector 128 and then to a remotely located central office 130 through a communication channel 170.”), FIG. 3 (emphasis added):



159. Therefore, the combination of Lu-577 and Tian teaches that the audience measurement system comprises a data collection facility to obtain the content identifying data and the audience identification information from the audience measurement device.

8. Claim 11

160. It is my opinion that a POSITA would have found claim 11 obvious based on the combined teachings of Lu-577 and Tian.

11[p]. The audience measurement system of claim 9, wherein the audience measurement device is to:

161. The combination of Lu-577 and Tian teaches the audience measurement system of claim 9. *Supra* §VI.A.7. As explained below, the combination of Lu-577 and Tian further teaches that the audience measurement

device performs the further operations recited in claim 11. *Infra* §VI.A.8[11a-11b].

[11a] reduce a resolution of the first image of the media exposure environment to obtain a reduced-resolution image, and
--

162. As explained for claim 4, the combination of Lu-577 and Tian teaches that the processor circuitry of the audience measurement system reduces a resolution of one of the one or more of the images of the media exposure environment to obtain a reduced-resolution image. *Supra* §VI.A.3[4a]. The “first image” in claim 9 is satisfied by one of the one or more images. *Supra* §VI.A.7[9d]. The audience measurement device includes the processor circuitry. *Supra* §VI.A.7[9a]. Therefore, the combination of Lu-577 and Tian teaches that the audience measurement device reduces a resolution of the first image of the media exposure environment to obtain a reduced-resolution image.

[11b] determine the orientation of the head with respect to the camera based on the reduced-resolution image.

163. As explained for claim 4, the combination of Lu-577 and Tian teaches that the processor circuitry of the audience measurement system determines the orientation of the head with respect to the camera based on the reduced-resolution image. *Supra* §VI.A.3[4b]. The audience measurement device includes the processor circuitry. *Supra* §VI.A.7[9a]. Therefore, the combination of Lu-577 and

Tian teaches that the audience measurement device determines the orientation of the head with respect to the camera based on the reduced-resolution image.

9. Claim 12

164. It is my opinion that a POSITA would have found claim 12 obvious based on the combined teachings of Lu-577 and Tian.

12[p]. The audience measurement system of claim 11, wherein the audience measurement device is to:

165. The combination of Lu-577 and Tian teaches the audience measurement system of claim 11. *Supra* §VI.A.8. As explained below, the combination of Lu-577 and Tian further teaches that the audience measurement device performs the further operations recited in claim 12. *Infra* §VI.A.9[12a-12b].

[12a] identify a region corresponding to the head within the first image from which the reduced-resolution image was obtained; and

166. As explained for claim 6, the combination of Lu-577 and Tian teaches that the processor circuitry of the audience measurement system identifies a region corresponding to the head within the first image from which the reduced-resolution image was obtained. *Supra* §VI.A.5[6a]. The audience measurement device includes the processor circuitry. *Supra* §VI.A.7[9a]. Therefore, the combination of Lu-577 and Tian teaches that the audience measurement device identifies a

region corresponding to the head within the first image from which the reduced-resolution image was obtained.

[12b] analyze a portion of a second image corresponding to the identified region to determine whether the head matches the known person.

167. As explained for claim 6, the combination of Lu-577 and Tian teaches that the processor circuitry of the audience measurement system analyzes a portion of a second image corresponding to the identified region, and based on the analysis of the region of the second image, determine that the head matches the known person. *Supra* §VI.A.5[6b]. Lu-000, which is incorporated in Lu-577 by reference, teaches that the known person is identified by comparing an extracted pattern image signature with stored pattern image signatures of different known individuals to determine whether there is a match with any of the individuals. Ex-1008, Abstract (“The extracted pattern image signature is compared with each of the stored pattern image signatures to identify a particular one of the predetermined audience members.”), 2:20-38, 7:39-8:6. Therefore, the combination of Lu-577 and Tian teaches determining **whether** the head matches the known person. The audience measurement device includes the processor circuitry. *Supra* §VI.A.7[9a]. Therefore, the combination of Lu-577 and Tian teaches that the audience measurement device analyzes a portion of a second image corresponding to the identified region to determine whether the head matches the known person.

10. Claim 13

168. It is my opinion that a POSITA would have found claim 13 obvious based on the combined teachings of Lu-577 and Tian.

13. The audience measurement system of claim 12, wherein the audience identification information includes an identifier associated with the known person and the audience measurement device is to, responsive to the indication that the head matches the known person, cause recordation of the identifier associated with the known person in a memory.

169. The combination of Lu-577 and Tian teaches the audience measurement system of claim 12. *Supra* §VI.A.9. As explained for claim 8, Lu-577 teaches that the audience identification information includes an identifier associated with the known person. *Supra* §VI.A.6. Lu-577 also teaches that the processor circuitry of the audience measurement system causes the identifier associated with the person to be stored in the memory responsive to a determination that the head matches the known person. *Supra* §VI.A.6. Because the audience measurement device includes the processor circuitry, the combination teaches that the audience measurement device performs this operation. *Supra* §VI.A.7[9a]. Lu-577 further teaches representing the match with an indication that the head matches the known person. *Supra* §VI.A.7[9g]. Furthermore, a POSITA would have understood that “storing” and “recordation” are interchangeable terms to describe saving of data in memory. Therefore, the combination of Lu-577 and Tian teaches that the audience measurement device causes **recordation** of the

identifier associated with the known person in a memory responsive to the **indication** that the head matches the known person.

11. Claim 14

170. It is my opinion that a POSITA would have found claim 14 obvious based on the combined teachings of Lu-577 and Tian.

14. The audience measurement system of claim 11, wherein the audience measurement device is to determine the audience identification information by: (i) generating a facial signature from a region of a second image corresponding to a location of the head in the reduced-resolution image, and (ii) comparing the generated facial signature to a database of facial signatures.

171. The combination of Lu-577 and Tian teaches the audience measurement system of claim 11. *Supra* §VI.A.8. As explained for claim 5, the combination of Lu-577 and Tian teaches that the processor circuitry of the audience measurement system determines the audience identification information by: (i) generating a facial signature from a region of a second image of the one or more of the images corresponding to a location of the head in the reduced-resolution image, and (ii) comparing the generated facial signature to a database of facial signatures. *Supra* §VI.A.4. The audience measurement device includes the processor circuitry. *Supra* §VI.A.7[9a]. Therefore, the combination of Lu-577 and Tian teaches that the audience measurement device determines the audience identification information by: (i) generating a facial signature from a region of a second image corresponding to a location of the head in the reduced-resolution

image, and (ii) comparing the generated facial signature to a database of facial signatures.

12. Claim 16

172. It is my opinion that a POSITA would have found claim 16 obvious based on the combined teachings of Lu-577 and Tian.

16[p]. A method for obtaining exposure data for a media exposure environment, the method comprising:

173. As explained for claim 1, the combination teaches an audience measurement system to obtain exposure data for a media exposure environment. *Supra* §VI.A.2.1[p]. The audience measurement system performs multiple steps. *Supra* §VI.A.2[1a-1g]. Both Lu-577 and Tian contemplate methods that include operations they teach. *See, e.g.*, Ex-1004, ¶0080 (“the audio signal methods employed by any of the other embodiments 60 ...”); Ex-1005, 000002-05 (“In Section 2 we describe the implementation details for the methods used in our evaluation in each step of facial expression analysis.”). In light of teachings of the combination of Lu-577 and Tian regarding operations performed by the audience measurement system, a POSITA would have understood that the combination teaches a **method** for obtaining exposure data for a media exposure environment. As explained below, the combination of Lu-577 and Tian further teaches that the method includes the operations recited in claim 16. *Infra* §VI.A.12[16a-16g].

[16a] generating, using a program detector, an audio signature of media content presented by a television within the media exposure environment;
--

174. As explained for claim 1, Lu-577 teaches that the audience measurement system generates an audio signature of media content presented by a television within the media exposure environment. *Supra* §VI.A.2[1b]. The audience measurement system comprises an audio detector that detects an audio portion of a presented program and a site unit or home unit that analyzes the audio data to identify a television program. *Supra* §VI.A.2[1a]; Ex-1004, ¶¶0037-0038, FIG.2. The audio detector and the site unit/home unit, executing machine readable instructions related to program detection, together form a program detector, as the term is used in the '243 patent. The '243 patent does not define “program detector” but states that an “example program detector ... detects presentation(s) of media content in the media exposure environment 100 and collects identification information associated with the detected presentation(s).” Ex-1001, 7:1-41. The functionality of the program detector in the '243 patent is consistent with that of the audio detector and site unit/home unit in Lu-577. A POSITA would have found it obvious to implement Lu-577's site unit/home unit using the architecture that includes memory, machine readable instructions, and processor circuitry. *Supra* §VI.A.2[1a]. A POSITA would have found it obvious to arrange the hardware and software of the audio detector and site unit/home unit to form a

program detector. Accordingly, Lu-577 teaches performing the operation of limitation [16a] using a program detector. Therefore, Lu-577 teaches generating, using a program detector, an audio signature of media content presented by a television within the media exposure environment.

[16b] obtaining, based on the audio signature, content identifying data corresponding to the presented media content;

175. As explained for claim 1, Lu-577 teaches that the audience measurement system obtains content identifying data corresponding to the presented media content, the content identifying data based on the audio signature. *Supra* §VI.A.2[1c]. Therefore, Lu-577 teaches obtaining, based on the audio signature, content identifying data corresponding to the presented media content.

[16c] while the media content corresponding to the content identifying data is presented by the television, capturing first and second images of the media exposure environment with a camera;

176. As explained for claim 1, the combination of Lu-577 and Tian teaches that the audience measurement system obtains a sequence of images of the media exposure environment with a camera while the media content corresponding to the content identifying data is presented by the television. *Supra* §VI.A.2[1d]. The combination of Lu-577 and Tian further teaches detecting a head appearing in one or more of the images. *Supra* §VI.A.2[1d]. The images are **captured** by a camera. Ex-1004, ¶0039 (“The person identifier 98 may be video camera, an IR

camera, or the like.”); Ex-1008, Abstract (“A video image is captured for each of the located individual audience members in the monitored area.”), 2:20-38, 3:27-60, 3:61-4:8. When there are more than one images in which a head is detected, the one or more images includes **first and second images** that are captured. Therefore, the combination of Lu-577 and Tian teaches, while the media content corresponding to the content identifying data is presented by the television, capturing first and second images of the media exposure environment with a camera.

[16d] analyzing, using an audience detector, the first image to attempt to detect a head in the first image;
--

177. As explained for claims 1 and 9, the combination of Lu-577 and Tian teaches that the audience measurement system analyzes the sequence of images including the first image and attempts to detect a head in the first image. *Supra* §VI.A.2[1d], §VI.A.7[9e].

178. The audience measurement system comprises a person identifier that is a camera and a site unit that employs head location and face recognition software to identify viewing persons in the audience. *Supra* §VI.A.2[1a]; Ex-1004, ¶0039, FIG.2. The person identifier and the site unit, executing machine readable instructions related to audience detection, together form an audience detector, as the term is used in the '243 patent. The '243 patent does not define “audience

detector” but states that “[t]he example audience detector 200 includes an image sensor 204, a people counter 206, a person identifier 207, a time stamper 208, and a memory 210” and “the audience detector 200 may have additional or alternative methods and/or components to identify people in the frames captured by the image sensor 204[.]” Ex-1001, 5:29-6:44. The functionality of the program detector in the ’243 patent is consistent with that of the person identifier and site unit in Lu-577. Furthermore, the ’243 patent broadly describes the “audience detector” as “may be implemented by hardware, software, firmware and/or any combination of hardware, software and/or firmware.” Ex-1001, 8:7-43. A POSITA would have found it obvious to implement Lu-577’s site unit using the architecture that includes memory, machine readable instructions, and processor circuitry. *Supra* §VI.A.2[1a]. A POSITA would have found it obvious to arrange the hardware and software of the person identifier and site unit to form an audience detector. Accordingly, Lu-577 teaches performing the operation of limitation [16d] using an audience detector.

179. Therefore, the combination of Lu-577 and Tian teaches analyzing, using an audience detector, the first image to attempt to detect a head in the first image.

[16e] determining, using the audience detector, an orientation of the head with respect to the camera;

180. As explained for claim 1, the combination of Lu-577 and Tian teaches that the audience measurement system determines an orientation of the head with respect to the camera. *Supra* §VI.A.2[1e]. Lu-577 teaches that the person identifier and the site unit, executing machine readable instructions related to audience detection, together form an audience detector. *Supra* §VI.A.12[16d]. Because the determination of head orientation is related to identifying audience members, it would have been obvious to use the audience detector to perform the operation of limitation [16e]. Therefore, the combination of Lu-577 and Tian teaches determining, using the audience detector, an orientation of the head with respect to the camera.

[16f] determining, using the audience detector, audience identification information based on an indication that the head matches a known person associated with the audience detector; and

181. As explained for claim 9, Lu-577 teaches that the audience measurement system determines audience identification information based on an indication that the head matches a known person associated with the audience measurement system. *Supra* §VI.A.7[9g].

182. Lu-577 teaches that the person identifier and the site unit, executing machine readable instructions related to audience detection, together form an

audience detector. *Supra* §VI.A.12[16d]. Because the determination of audience identification information is a functionality of the site unit, it would have been obvious to use the audience detector to perform the operation of limitation [16f].

183. As explained above, the audience detector is formed by components of the audience measurement system. *Supra* §VI.A.12[16d]. Therefore, known persons associated with the audience measurement system are also associated with audience detector. For example, the association exists because the person identifier is set up to capture images of the known persons. Because Lu-577 teaches determining a match with a known person associated with the audience measurement system, it teaches determining the match with **a known person associated with the audience detector**. Therefore, Lu-577 teaches determining, using the audience detector, audience identification information based on an indication that the head matches a known person associated with the audience detector.

[16g] providing the content identifying data and the audience identification information to a data collection facility.

184. As explained for claim 1, the combination of Lu-577 and Tian teaches that the audience measurement system outputs a signal indicative of the content identifying data and the audience identification information to a data collection facility. *Supra* §VI.A.2[1g]. A POSITA would have understood that one well-

known way to provide the content identifying data and the audience identification information is to output a signal thereof. Therefore, the combination of Lu-577 and Tian teaches providing the content identifying data and the audience identification information to a data collection facility.

13. Claim 18

185. It is my opinion that a POSITA would have found claim 18 obvious based on the combined teachings of Lu-577 and Tian.

18[p]. The method of claim 16, further comprising:

186. The combination of Lu-577 and Tian teaches the method of claim 16. *Supra* §VI.A.12. As explained below, the combination of Lu-577 and Tian further teaches that the method comprises the further operations recited in claim 18. *Infra* §VI.A.13[18a-18b].

[18a] reducing a resolution of the first image of the media exposure environment to obtain a reduced-resolution image, and

187. As explained for claim 11, the combination of Lu-577 and Tian teaches that the audience measurement system reduces a resolution of the first image of the media exposure environment to obtain a reduced-resolution image. *Supra* §VI.A.8[11a]. Therefore, the combination of Lu-577 and Tian teaches reducing a resolution of the first image of the media exposure environment to obtain a reduced-resolution image.

[18b] wherein the determining of the orientation of the head includes determining the orientation of the head with respect to the camera using the reduced-resolution image.

188. The combination of Lu-577 and Tian teaches determining an orientation of the head with respect to the camera. *Supra* §VI.A.12[16e]. As explained for claim 11, the combination of Lu-577 and Tian teaches carrying out the determination based on the reduced-resolution image. *Supra* §VI.A.8[11b]. Therefore, the combination of Lu-577 and Tian teaches that the determining of the orientation of the head includes determining the orientation of the head with respect to the camera using the reduced-resolution image.

14. Claim 19

189. It is my opinion that a POSITA would have found claim 19 obvious based on the combined teachings of Lu-577 and Tian.

19[p]. The method of claim 18, further comprising:

190. The combination of Lu-577 and Tian teaches the method of claim 18. *Supra* §VI.A.13. As explained below, the combination of Lu-577 and Tian further teaches that the method comprises the further operations recited in claim 19. *Infra* §VI.A.14[19a-19b].

[19a] identifying a region corresponding to the head within the first image from which the reduced-resolution image was obtained;

191. As explained for claim 12, the combination of Lu-577 and Tian teaches that the audience measurement system identifies a region corresponding to the head within the first image from which the reduced-resolution image was obtained. *Supra* §VI.A.9[12a]. Therefore, the combination of Lu-577 and Tian teaches identifying a region corresponding to the head within the first image from which the reduced-resolution image was obtained.

[19b] analyzing a portion of the second image corresponding to the identified region; and based on the analyzing of the region of the second image, determining whether the head matches the known person.

192. As explained for claim 12, the combination of Lu-577 and Tian teaches that the audience measurement system analyzes a portion of a second image corresponding to the identified region to determine whether the head matches the known person. *Supra* §VI.A.9[12b]. Therefore, the combination of Lu-577 and Tian teaches analyzing a portion of the second image corresponding to the identified region; and based on the analyzing of the region of the second image, determining whether the head matches the known person.

15. Claim 20

193. It is my opinion that a POSITA would have found claim 20 obvious based on the combined teachings of Lu-577 and Tian.

20. The method of claim 19, wherein the audience identification information includes an identifier associated with the known person, and further including recording the identifier associated with the known person in a memory.

194. The combination of Lu-577 and Tian teaches the method of claim 19.

Supra §VI.A.14. As explained for claim 13, Lu-577 teaches that the audience identification information includes an identifier associated with the known person.

Supra §VI.A.10. Furthermore, Lu-577 teaches that the audience measurement system causes recordation of the identifier associated with the known person in a memory. *Supra* §VI.A.10. Therefore, Lu-577 teaches that the audience identification information includes an identifier associated with the known person and that the method further includes recording the identifier associated with the known person in a memory.

B. Ground 2: Claims 1, 9, and 16 are Rendered Obvious in View of Nielsen-372 and Steinberg

195. It is my opinion that a POSITA would have found Claims 1, 9, and 16 obvious based on the teachings of Nielsen-372 and Steinberg.

1. Combination of Nielsen-372 and Steinberg

196. Nielsen-372 teaches a “local metering system 100 capable of providing viewing and metering information for program content presented via an example home entertainment system[.]” Ex-1006, ¶0036. The home entertainment system includes a television. Ex-1006, ¶0037, ¶0040. The local metering system

includes a sensor that receives audio signals corresponding to a presented program and a metering unit or display device ON/OFF detector that identifies the presented program based on the audio signals. Ex-1006, ¶0044, ¶0057. The local metering system also includes a people meter that captures information about the audience. Ex-1006, ¶0048. The people meter enables the detection of the identities and/or number of persons currently viewing programs displayed on the television. Ex-1006, ¶0003. Nielsen teaches that the people meter takes an approach of “prompt[ing] the audience members to indicate that they are present in the viewing audience by pressing the appropriate input key on [a] people meter control device[.]” Ex-1006, ¶0048.

197. Steinberg is directed to “[a] system, method and techniques ... for applying digital image processing using face detection and face recognition techniques to identify the persons contained within a digital image, which may be part of a personal or other image collection.” Ex-1007, ¶0090. Steinberg teaches performing face detection in images to detect one or more faces and the locations of the faces. Ex-1007, ¶0005, ¶0057, ¶0118, ¶0128, ¶0130. Steinberg teaches determining a pose or orientation of a face in an image in order to perform normalization. Ex-1007, ¶0120, ¶0166, ¶0168, ¶0176. Steinberg further teaches performing face recognition to determine identity information of a face in an image

based on a match with that of a known person. Ex-1007, ¶¶0005, ¶¶0057, ¶¶0092, ¶¶0096, ¶¶0126; ¶¶0127, ¶¶0130, ¶¶0157-0158.

198. A POSITA would have understood that Steinberg’s teachings regarding determining the identities of individuals appearing in images are applicable to identifying members of a television audience. In the 2011 timeframe, cameras were widely used. A POSITA would have found it obvious to apply Steinberg’s teachings to Nielsen-372—using a camera to capture one or more images of a television audience and applying Steinberg’s approach to identify the individuals in the audience by detecting and recognizing faces in the image(s).

199. A POSITA would have found it obvious to implement Steinberg’s approach on Nielsen-372’s people meter or other components of the local metering system because Nielsen-372’s system includes hardware needed to implement Steinberg’s approach. Ex-1006, ¶¶0050, ¶¶0135, FIG. 29. Steinberg describes its disclosures as “[a] processor-based system operating according to digitally embedded programming instructions....” Ex-1007, Abstract. Steinberg describes implementing its solution on a computer and an image acquisition device (e.g., camera). Ex-1007, ¶¶0092, ¶¶0112. According to Steinberg, modules for performing face detection, normalization, and recognition may be implemented on the computer (e.g., FIG. 1(a)), the image acquisition device (e.g., FIG. 1(c)), or both (e.g., FIG. 1(b)). Ex-1007, ¶¶0112 (“In particular, FIG. 1(a) describes an

external computer application, and FIG. 1(b) describes a system, wherein certain functions are implemented on the image acquisition device.”), ¶0129 (“FIG. 1(c) schematically illustrates the in-camera user interface (UI) modules 1500, the workflow module 1610 and the image acquisition engine 1600 each implemented entirely within a digital camera or other image capture device.”).

200. A POSITA would have realized that each of Steinberg’s alternative manners of implementation are applicable to Nielsen-372’s local metering system. As explained below, Nielsen-372 teaches implementing the local metering system using a computer that includes a processor and machine readable instructions or programming instructions. *Infra* §VI.B.2[1a]; *see, e.g.*, Ex-1006, FIG. 29, ¶¶0135-0136, ¶0141. Nielsen-372’s local metering system includes a camera. Ex-1006, ¶0050 (“The video sensor 208 may be, for example, a camera, a single output analog or digital light sensor, etc.”). A POSITA would have found it obvious to implement the face detection, normalization, and recognition teachings of Steinberg on either the computer or the camera of Nielsen-372 or distribute the functionality between the computer and the camera based on teachings of Steinberg. A POSITA would have been able to implement these functionalities using well known computer software and hardware structures as taught by Nielsen-372.

201. A POSITA would have been motivated to apply the teachings of Steinberg to Nielsen-372 because of the benefits provided by Steinberg's approach. Nielsen-372's approach of detecting individuals in the audience requires prompting the individuals and soliciting manual inputs from the individuals. Ex-1006, ¶0048. A POSITA would have understood that this approach may create a negative audience experience by interrupting an individual's viewing of a television program and requiring the individual to take an action unrelated to the entertainment experience. The interruption may potentially affect the individual's television viewing behavior. Furthermore, the accuracy of the data collected by this approach may be compromised if, for example, a member of the audience refuses to provide the input showing his or her presence or one individual incorrectly provides an input indicating the presence of another individual. Applying Steinberg's method to Nielsen-372 would allow images of the audience to be taken without interruptions to the entertainment experience and automatic identification of individuals in the audience. Furthermore, because the individuals are identified automatically based on images, the accuracy of the identification is not affected by the individual's input or lack thereof. A POSITA would have understood that these benefits would have furthered Nielsen-372's goal of collecting viewing information by facilitating a less intrusive manner that collects more accurate information.

202. A POSITA would have found it obvious and been motivated to apply the teachings of Steinberg to Nielsen-372 because this is a simple substitution of one known element for another to obtain predictable results. Nielsen-372 teaches an approach of identifying individuals in the audience by prompting them to provide input indicating their presence. Steinberg teaches a different approach of identifying individuals by performing face detection, normalization, and recognition on images of the individuals. The combination of simply substitutes Nielsen-372's approach for Steinberg's approach. Doing so would have yielded predictable results because the techniques of facial analysis were well known in the 2011 timeframe. *Supra* §I. A POSITA would have been more than capable of applying such techniques to the context to television audience measurement.

203. A POSITA would have had a reasonable expectation of success when combining the teachings of Nielsen-372 and Steinberg, given that Nielsen-372 provide the hardware architecture for performing Steinberg's face recognition method and that Steinberg's techniques were well known. In fact, even the '243 patent itself admits that facial recognition was known in the prior art. Ex-1001, 1:67-2:3 ("[S]ome audience measurement systems identify people in the images via one or more identification techniques such as, for example, facial recognition."). As explained above, the combination simply implements Steinberg's teachings about facial analysis, in the manner taught by Steinberg, in

Nielsen-372's system. A POSITA would have been more than capable of implementing the face analysis techniques taught by Steinberg, such as face detection, normalization, and recognition. *See supra* §I. Furthermore, a POSITA would have been motivated and would have had a reasonable expectation of success in combining the references, as explained above, because the combination would not have changed the principle of operation for any of the teachings of the references relied upon in the combination. Applying the teachings of Steinberg to Nielsen-372 would have been nothing more than the use of a known technique to improve similar devices in the same way.

2. Claim 1

204. It is my opinion that a POSITA would have found claim 1 obvious based on the combined teachings of Nielsen-372 and Steinberg.

1[p]. An audience measurement system to obtain exposure data for a media exposure environment,

205. I understand that a preamble may or may not state a limitation of a claim. However, to the extent the preamble limits claim 1, it is my opinion that the combination of Nielsen-372 and Steinberg teaches the preamble.

206. Nielsen-372 teaches **an audience measurement system** (e.g., including a local metering system).² Ex-1006, ¶0036 (“A block diagram of an example local metering system 100 capable of providing viewing and metering information for program content presented via an example home entertainment system 102 is illustrated in FIG. 1.”), FIG. 1:

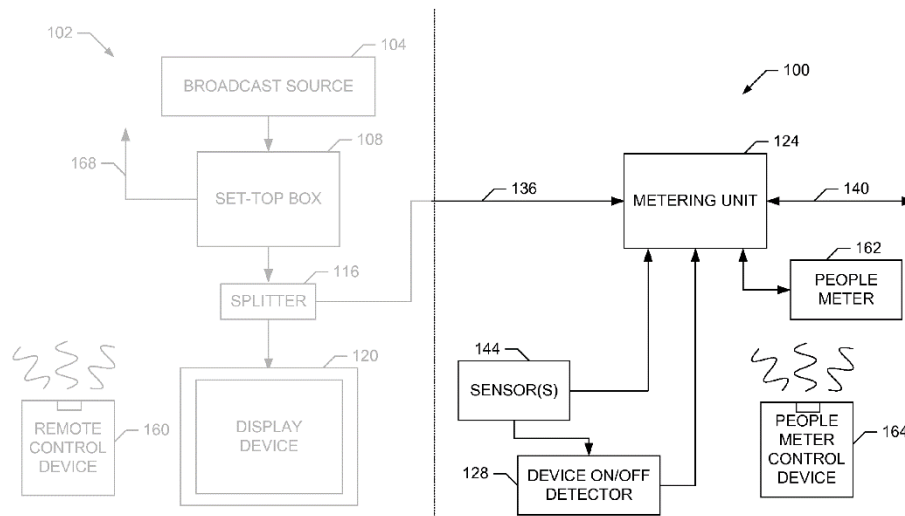


FIG. 1

207. Nielsen-372 teaches that its local metering system **obtains exposure data** (e.g., data identifying a displayed program, data identifying the audience) **for a media exposure environment** (e.g., environment in which a home entertainment system is placed). The local metering system “includes a metering unit 124 and a display device ON/OFF detector 128.” Ex-1006, ¶0036. Nielsen-372 teaches that

² Nielsen-372 teaches other systems or devices, such as a central (data processing) facility, that may be included as part of the audience measurement system. Ex-1006, ¶0036, ¶0041.

“the metering unit 124 and the display device ON/OFF detector 128 [are] integrated into a single unit.” Ex-1006, ¶0049. The local metering system further includes a sensor to receive audio signals corresponding to a displayed program. Ex-1006, ¶0044 (“For example, a sensor 144 may be implemented by a microphone placed in the proximity of the display device 120 to receive audio signals corresponding to the program being displayed.”). The metering unit or the display device ON/OFF detector identifies the displayed program based on the audio signals. Ex-1006, ¶0044 (“The metering unit 124 and/or display device ON/OFF detector 128 may then process the audio signals received from the microphone 144 to decode any embedded ancillary code(s) and/or generate one or more audio signatures corresponding to a program being displayed.”); ¶0057.

208. The local metering system also “includes a people meter 162 to capture information about the audience.” Ex-1006, ¶0048. Nielsen-372 explains that the people meter enables the detection of the identities and/or number of persons viewing programs on a television. Ex-1006, ¶0003 (“[A] people meter may be located in the viewing space of the television and in communication with the home unit, thereby enabling the home unit to detect the identities and/or number of the persons currently viewing programs displayed on the television.”). A POSITA would have found it obvious to apply Steinberg’s teachings to the

people meter such that the people meter identifies individuals in the audience based on images captured by a camera. *Supra* §VI.B.1.

209. The local metering system is placed in the same environment as a home entertainment system, such as a statistically selected household. Ex-1006, ¶0036. Nielsen-372 teaches that the home entertainment system includes a television for presenting media content. Ex-1006, ¶0037 (“The broadcast source 104 may provide analog and/or digital television signals to the home entertainment system 102, for example, over a coaxial cable or via a wireless connection.”), ¶0040 (“The display device 120 may be any type of video display device, such as a television.”). The environment of the local metering system is a “media exposure environment” because it is a space in which a television is located. *Supra* §V.A. The ’243 patent’s description that the media exposure environment is statistically selected is further consistent with Nielsen-372’s teaching of placing the local metering system in a statistically selected household. Ex-1001, 4:56-5:8. Therefore, the combination of Nielsen-372 and Steinberg teaches an audience measurement system to obtain exposure data for a media exposure environment.

210. Furthermore, the ’243 patent admits that the preamble was known in the prior art by discussing it in the “Background” section. *Supra* §VI.A.2.1[p]; Ex-1001, 1:56-2:3.

[1a] the audience measurement system comprising: memory; machine readable instructions; and processor circuitry to execute the machine readable instructions to:

211. The combination of Nielson-372 and Steinberg teaches this limitation. Nielsen-372 teaches “implementing the apparatus and methods” it discloses using a computer. Ex-1006, ¶0135, FIG. 29:

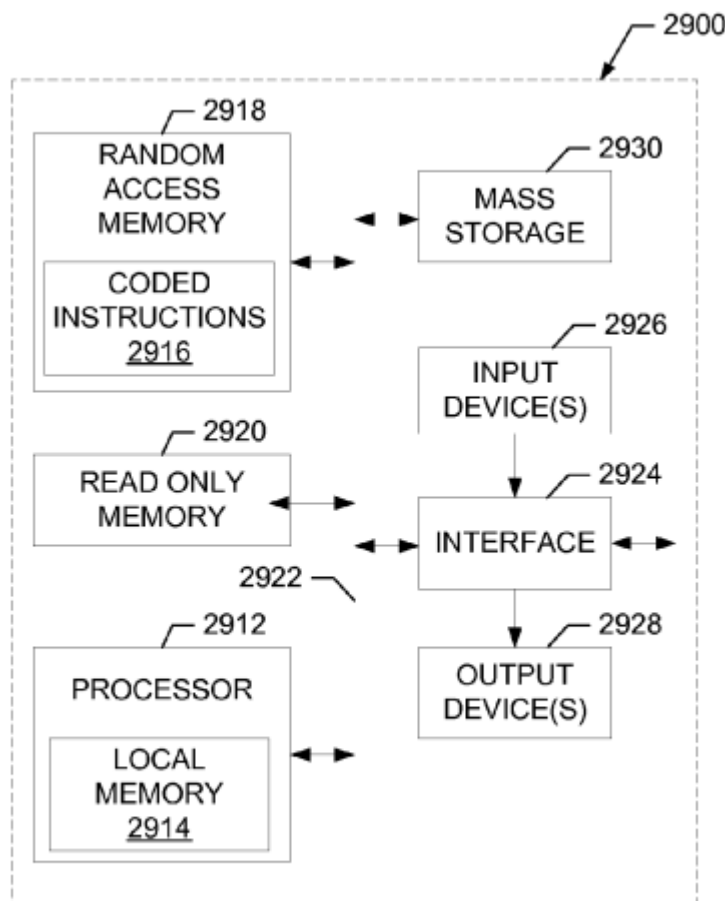


FIG. 29

Nielsen-372 teaches that the computer comprises **memory** (e.g., local memory, memory device, main memory, mass storage devices); **machine readable instructions** (e.g., coded instructions, software); and **processor circuitry** (e.g.,

processor) **to execute the machine readable instructions.** Ex-1006, ¶0135 (“The processor 2912 includes a local memory 2914, and executes coded instructions 2916 present in the local memory 2914 and/or in another memory device.”), ¶0136 (“The processor 2912 is in communication with a main memory including a volatile memory 2918 and a non-volatile memory 2920 via a bus 2922.”), ¶0141 (“The computer 2900 also includes one or more mass storage devices 2930 for storing software and data.”).

212. Nielsen-372 teaches that this architecture is applicable “to implement the [] display device ON/OFF detector of the local metering system.” Ex-1006, ¶0035, ¶0036 (“The example local metering system 100 includes a metering unit 124 and a display device ON/OFF detector 128.”). For example, Nielsen-372 teaches that “[t]he processor 2912 may execute ... the machine readable instructions represented in FIGS. 13 through 28 and/or implement any or all of the processors 612, 712, 812, 912, 1012, 1112, 1212, 1262 and/or 1292.” Ex-1006, ¶0135. Nielsen-372 further teaches that “[t]he mass storage device 2930 may be used [to] store any or all of the machine readable instructions 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700 and/or 2800.” Ex-1006, ¶0141.

213. Accordingly, this architecture is applicable to the audio processor systems of Figures 6 and 7 and the machine readable instructions of Figures 14 and

15, which are used to implement the audio signature processor that is part of display device ON/OFF detector. Ex-1006, FIGs. 6-7, FIGs. 14-15, ¶0075 (“A first example audio processor system 600 that may be used to implement ... the audio signature processor 316 ... is shown in FIG. 6.”), ¶0078 (“A second example audio processor system 700 that may be used to implement ... the audio signature processor 316 ... is shown in FIG. 7.”), ¶0095 (“First example machine readable instructions 1400 that may be executed to implement the audio signature processor 316 of FIG. 3 are shown in FIG. 14.”), ¶0097 (“Second example machine readable instructions 1500 that may be executed to implement the audio signature processor 316 of FIG. 3 are shown in FIG. 15.”), ¶0056 (“The example set of audio engines 228 of FIG. 3 includes ... an audio signature processor 316....”), ¶0051 (“The display device ON/OFF detector 200 includes one or more audio processors 228....”).

214. Nielsen-372 teaches that the metering unit of the local metering system “is a data logging and processing unit that may be used to generate viewing records and other viewing information useful for determining viewing and other metering information.” Ex-1006, ¶0041. A POSITA would have found it obvious to implement the metering unit using a memory for “data logging” functionalities and a processor circuitry and machine readable instructions for “processing” functionalities.

215. Furthermore, a POSITA would have found it obvious to apply Steinberg's teachings to the people meter such that the people meter identifies individuals in the audience based on images captured by a camera. *Supra* §VI.B.1. Because the computer architecture of memory, machine readable instructions, and processor circuitry was well known and widely used in the 2011 timeframe and it is applied to implement the local metering system, a POSITA would have found it obvious to apply this architecture to implement the people meter.

216. As explained below, the display device ON/OFF detector and metering unit perform operations for identifying media content presented by a television. *Infra* §VI.B.2[1b-1c]. The people meter, as modified based on Steinberg's teachings, performs operations for identifying members of the audience. *Infra* §VI.B.2[1d-1f]. Therefore, the combination of Nielsen-372 and Steinberg teaches that the processor circuitry executes the machine readable instructions to perform the operations of limitations [1b-1f].

[1b] generate an audio signature of media content presented by a television within the media exposure environment;
--

217. Nielsen-372 teaches this limitation. Nielsen-372 teaches that the audience measurement system (e.g., including the local metering system) **generates an audio signature of media content presented by a television** (e.g., a program displayed by the home entertainment system) **within the media exposure**

environment (e.g., environment in which the home entertainment system is placed). Ex-1006, Abstract (“An example method ... comprises determining an audio signature based on an audio signal corresponding to an output of the presentation device[.]”).

218. Nielsen-372 teaches that a sensor of the local metering system receives audio signals of a displayed program. Ex-1006, ¶0044 (“For example, a sensor 144 may be implemented by a microphone placed in the proximity of the display device 120 to receive audio signals corresponding to the program being displayed.”), ¶0050 (“The audio sensor 204 may be one or more microphones positioned to detect audio signals emitted by the display device 120 or corresponding information presenting device.”), FIGs. 1-2.

219. Nielsen-372 next teaches that “[t]he metering unit 124 and/or display device ON/OFF detector 128 ... process the audio signals received from the microphone 144 to ... generate one or more audio signatures corresponding to a program being displayed.” Ex-1006, ¶0044. The display device ON/OFF detector includes an audio signature processor. Ex-1006, ¶0051 (“The display device ON/OFF detector 200 includes one or more audio processors 228....”), ¶0056 (“The example set of audio engines 228 of FIG. 3 includes ... an audio signature processor 316....”), FIG. 3. The audio signature processor processes the audio signals and generates an audio signature of the displayed program. ¶0055 (“The

audio processors 228 process the input audio signal(s) 230 provided, for example, by the audio sensor(s) 204 of FIG. 2.”), ¶0057 (“The example audio signature processor 316 of FIG. 3 is configured to generate and process audio signatures corresponding to the input audio signal 230.”), FIGs. 2-3. Nielsen-372 teaches methods used by the audio signature processor to generate audio signatures. Ex-1006, ¶¶0095-0096 (“After convergence of the AGC algorithm at block 1404, control proceeds to block 1408 at which the audio signature processor 316 generates an audio signature from the received audio signal.”), FIG. 14 (emphasis added):

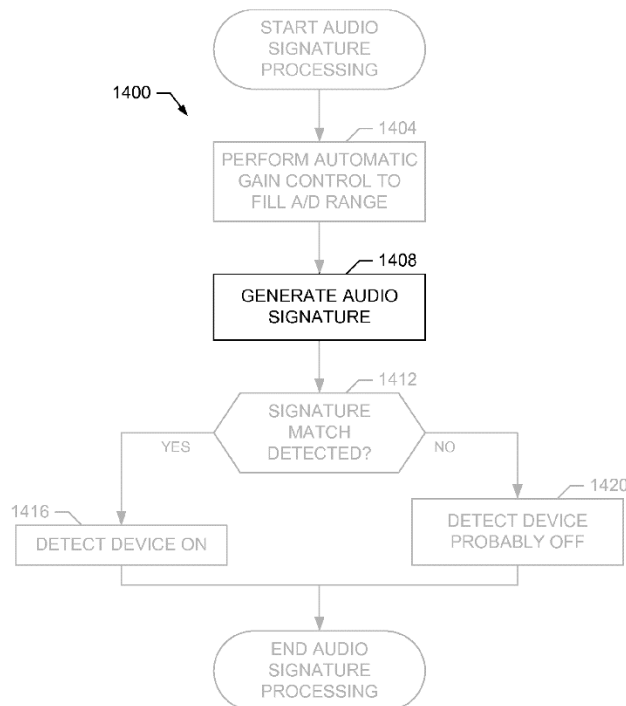


FIG. 14

See also ¶¶0097-0100, FIG. 15.

220. Nielsen-372 teaches that the displayed program is media content presented by a television. Ex-1006, ¶0036 (“The example home entertainment system 102 includes a broadcast source 104, a set-top box (STB) 108, a signal splitter 116 and a display device 120.”), ¶0037 (“The broadcast source 104 may be any broadcast media source, such as a cable television service provider, a satellite television service provider, a radio frequency (RF) television service provider, an internet streaming video/audio provider, etc.”), ¶0040 (“The display device 120 may be any type of video display device, such as a television.”).

221. As explained above, the environment in which the home entertainment system is placed is a media exposure environment. *Supra* §VI.B.2.1[p]. Thus, Nielsen-372 teaches media content presented by a television within the media exposure environment. Therefore, Nielsen-372 teaches an audience measurement system generating an audio signature of media content presented by a television within the media exposure environment.

222. Furthermore, the ’243 patent admits that this limitation was known in the prior art by discussing it in the “Background” section. *Supra* §VI.A.2[1b]; Ex-1001, 1:43-55.

[1c] obtain content identifying data corresponding to the presented media content, the content identifying data based on the audio signature of the media content presented by the television within the media exposure environment;
--

223. Nielsen-372 teaches this limitation. Nielsen-372 teaches that the audience measurement system (e.g., including the local metering system) **obtains content identifying data** (e.g., program identifying information) **corresponding to the presented media content** (e.g., television program). Nielsen teaches that the metering unit and the display device ON/OFF detector obtains program identifying information. NielsenEx-1006, ¶0044 (“To facilitate the determination of program identifying information and the generation of viewing records for the program content received and output by the STB 108 ..., the metering unit 124 and the display device ON/OFF detector 128 may be provided with one or more sensors 144.”). The program identifying information is part of viewing data that is collected and transmitted. Ex-1006, ¶0036 (“[T]he metering unit 124 may be implemented as a single home unit and one or more site units. ... Each site unit ... performs the functions of collecting viewing/metering data, processing such data (possibly in real-time) and sending the processed data to the single home unit for that home. The home unit receives and stores the data collected by the site units and subsequently forwards that collected data to the central facility.”), ¶0041 (“The metering unit 124 is a data logging and processing unit that may be used to

generate viewing records and other viewing information useful for determining viewing and other metering information.”), ¶0043 (“Thus, the display device ON/OFF detector 128 may be used to augment the viewing information and/or viewing records determined by the metering unit 124 to more accurately determine whether program content output by the STB 108 is actually presented by the display device 120.”). Because the program identifying information identifies the program presented by the home entertainment system, it is content identifying data. Ex-1006, ¶0044.

224. Nielsen further teaches that **the content identifying data is based on the audio signature of the media content presented by the television within the media exposure environment** (e.g., environment in which the home entertainment system is placed). As explained with respect to limitation [1b], Nielsen-327 teaches its local metering system generating an audio signature of media content presented by a television within the media exposure environment. *Supra* §VI.B.2[1b]. Nielsen-372 teaches identifying the presented media content by comparing the audio signature to a set of reference signatures and finding a match. Ex-1006, ¶0057 (“The signature information for the content being presented may be compared to a set of reference signatures corresponding to a known set of content. When a substantial match is found, the currently presented program content can be identified with a relatively high probability.”), ¶0096 (“If at block

1412 the audio signature processor 316 determines that the audio signature generates at block 1408 matches a known reference signature, control proceeds to block 1416 at which the audio signature processor 316 determines that the monitored information presenting device is ON.”), FIG. 14 (emphasis added):

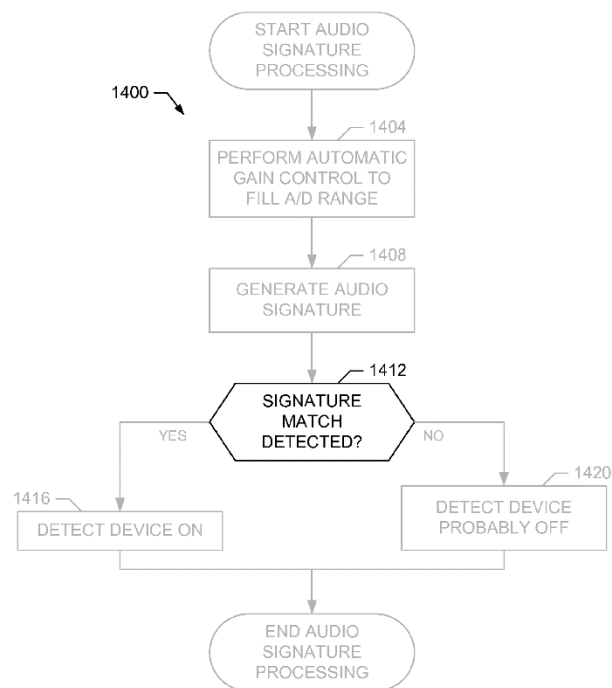


FIG. 14

Therefore, Nielsen-372 teaches that the content identifying data is based on the audio signature.

225. Nielsen-372 teaches that the operations of identifying the presented media content are performed by the audio signature processor (Ex-1006, ¶¶0057, ¶¶0095-0096), which is part of the display device ON/OFF detector (*supra* §VI.B.2[1a]; Ex-1006, ¶¶0051, ¶¶0056). Because content identifying data is

obtained by the metering unit and display device ON/OFF detector, Nielsen-372 teaches that the processor circuitry executes the machine readable instructions to perform the operations of limitation [1c]. *Supra* §VI.B.2[1a].

226. Furthermore, the '243 patent admits that this limitation was known in the prior art by discussing it in the “Background” section. *Supra* §VI.A.2[1c]; Ex-1001, 1:43-55.

[1d] analyze a sequence of images of the media exposure environment to detect a head appearing in one or more of the images, the sequence of images obtained by a camera while the media content corresponding to the content identifying data is presented by the television;
--

227. The combination of Nielsen-372 and Steinberg teaches this limitation. Steinberg teaches **obtaining a sequence of images by a camera** (e.g., digital camera, image capture device). Ex-1007, ¶0010 (“At this point we note that the present invention is presented primarily in the context of collections of consumer digital images which would be generated by a typical user of a digital camera. Such an image collection is in a constant state of growth as new sets of images are added every time the user off-loads pictures from the camera onto his computer.”). Steinberg teaches an image capture device (e.g., a digital camera) that captures images and handles certain image processing operations. Ex-1007, ¶0128 (“Similarly, the functionality of modules from the Image Classification Engine [1400] can be separated in various ways and implemented in part, or indeed in its

entirety within the Image Capture Device [1300].”), ¶0129 (“FIG. 1(c) schematically illustrates the in-camera user interface (UI) modules 1500, the workflow module 1610 and the image acquisition engine 1600 each implemented entirely within a digital camera or other image capture device.”), FIG. 1(b) (emphasis added):

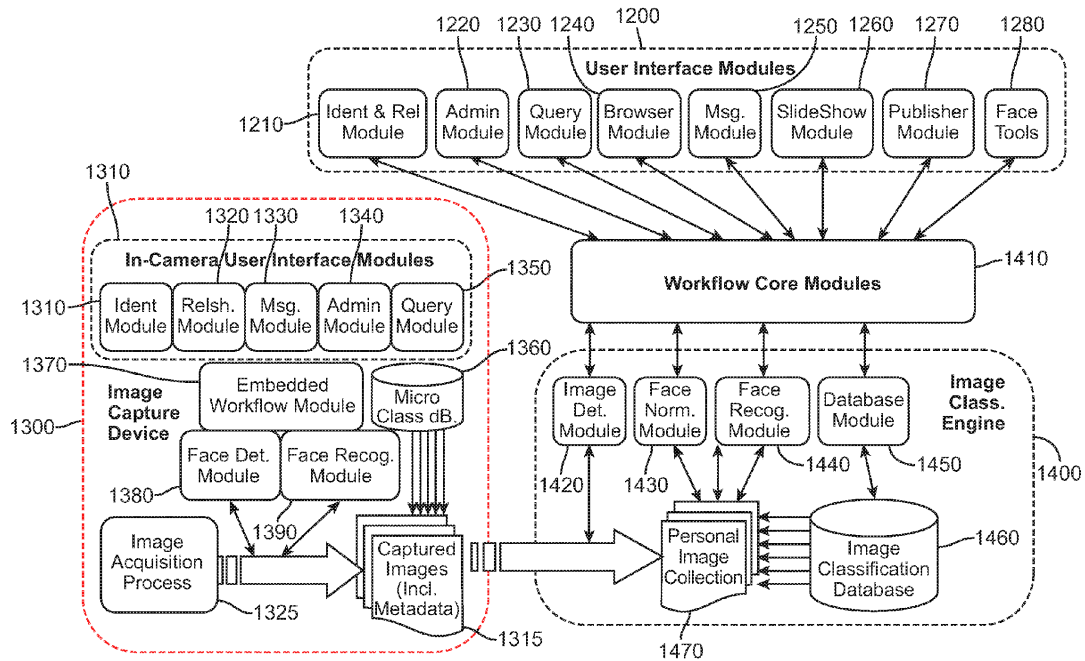
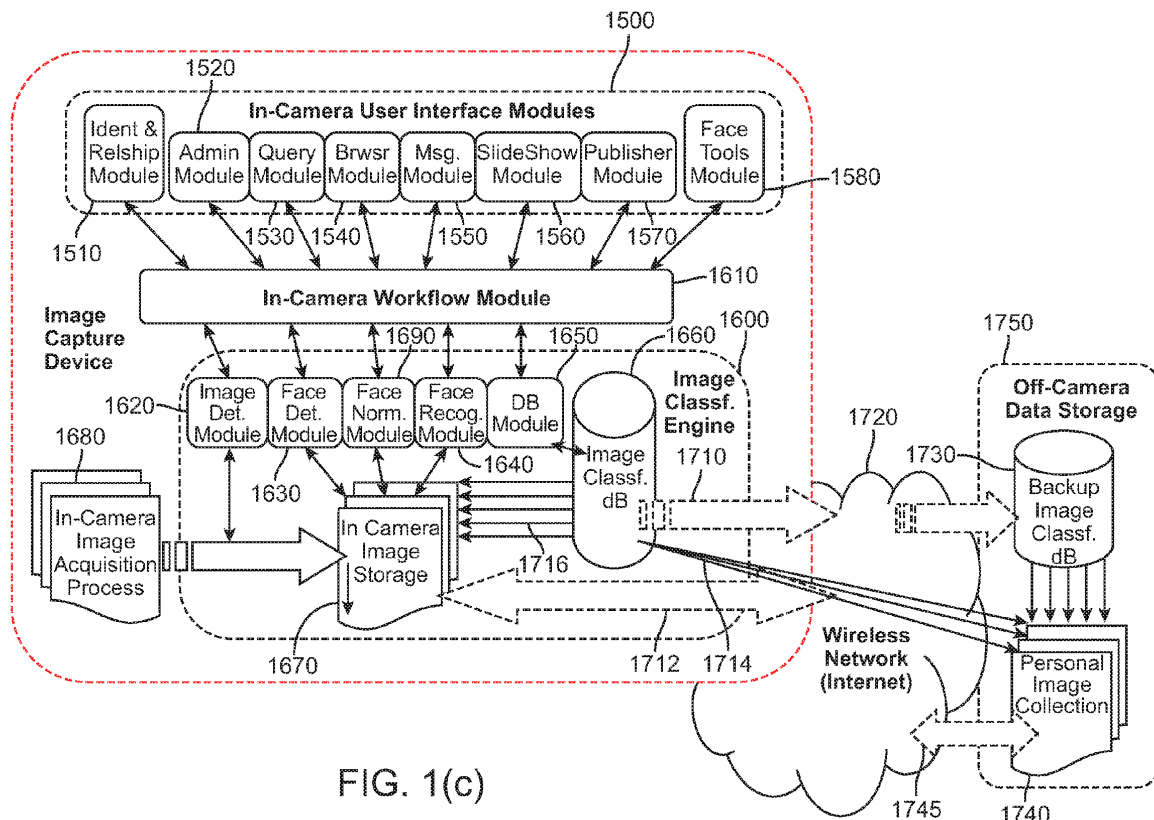


FIG. 1(b)

FIG. 1(c) (emphasis added):



228. Steinberg teaches adding sets of new images to an image collection.

Ex-1007, ¶0010. A POSITA would have understood that such a set of images encompasses consecutive images obtained by the camera within a certain period of time and that such a set of images form a sequence of images. Steinberg thus teaches the camera obtaining a sequence of images.

229. The combination of Nielsen-372 and Steinberg teaches obtaining the sequence of images **while the media content corresponding to the content identifying data is presented by the television**. A POSITA would have found it obvious to apply Steinberg's teachings to the people meter such that the people

meter identifies individuals in the audience based on images captured by a camera. *Supra* §VI.B.1. Nielsen-372 teaches that “a people meter may be located in the viewing space of the television and in communication with the home unit, thereby enabling the home unit to detect the identities and/or number of the persons currently viewing programs displayed on the television.” Ex-1006, ¶0003, ¶0048 (“The example local metering system 100 of FIG. 1 also includes a people meter 162 to capture information about the audience.”). To detect the identities of the persons “currently viewing programs displayed on the television” based on images, a POSITA would have found it obvious that the images are obtained in the media exposure environment, in which the persons view the programs, while the media content is presented by the television. As explained with respect to limitation [1c], Nielsen-372 teaches obtain content identifying data corresponding to the presented media content. *Supra* §VI.B.2[1c]. Therefore, when applying Steinberg’s teachings of obtaining a sequence of images to Nielsen-372’s teachings, a POSITA would have found it obvious to obtain the sequence of images while the media content corresponding to the content identifying data is presented by the television.

230. The combination of Nielsen-372 and Steinberg teaches **analyzing a sequence of images of the media exposure environment** (e.g., environment in which a home entertainment system is placed) **to detect a head appearing in one or more of the images**. As explained above, in light of Nielsen-372’s teachings, a

POSITA would have found it obvious to identify the persons in the audience based on **images of the media exposure environment**.

231. Steinberg teaches that, when new image(s) are acquired, a Face Detection module is alerted to further process the image(s). Ex-1007, ¶0092 (“An image detection module determines when new images are added to the image collection and alerts the main workflow module which then activates a face detection module which scans the image to determine if faces are present.”), ¶0117 (“The first of these modules is an Image Detection module [1120] alerts the system that at least one new image has been obtained from some external image source[.]”), ¶0128 (“If the Image Capture Device [1300] is set to operate in “recognition mode” then the Image Acquisition Process [1325] will alert the Face Detection module [1380] when a new image is successfully acquired.”), ¶0130 (“The image detection module 1620 works in conjunction with the in-camera image acquisition process to alert the workflow module that an original image has been captured and initially processed into in-camera image storage 1670. ... These stored images are then preferably processed by the face detection module 1630....”), ¶0151 (“The system sits in a wait state [3110] until the Image Detection Module [1120] informs the Face Detection module that an image is available for processing.”). Steinberg contemplates analyzing multiple images that are added to the image collection. *See, e.g.*, Ex-1007, ¶0092 (“new images”), ¶0117 (“at least

one new image”), ¶0130 (“stored images”). Therefore, Steinberg teaches analyzing a sequence of images.

232. Steinberg teaches face detection. Ex-1007, Abstract (teaching “identifying a group of pixels corresponding to a face region within digital image data acquired by an image acquisition device.”), ¶0005 (“Face detection has to do with the problem of locating regions within a digital image or video sequence which have a high probability of representing a human face.”), ¶0057 (“‘Face Detection’ involves the art of isolating and detecting faces in an image; Face Detection includes a process of determining whether a human face is present in an input image, and may include or is preferably used in combination with determining a position and/or other features, properties, parameters or values of parameters of the face within the input image[.]”). After the Face Detection module is alerted of new images, it detects a face in one or more of the images. Ex-1007, ¶0118 (“Once at least one new image is detected it is preferably then analyzed by the Face Detection module [1130]. This module employs a means to determine if facial candidate regions are present in the new image.”), ¶0128 (“If face candidate regions are present in this image the Face Detection module [1380] will alert the Embedded Workflow module [1370] which ... may pass the extracted face candidate regions on to the Face Recognition module [1390] to obtain the relevant Faceprint.”), ¶0130 (“These stored images are then preferably processed

by the face detection module 1630 ... when one or more faces are detected within the acquired images....”).

233. Steinberg teaches a workflow for face detection. Ex-1007, ¶0151,

FIG. 3:

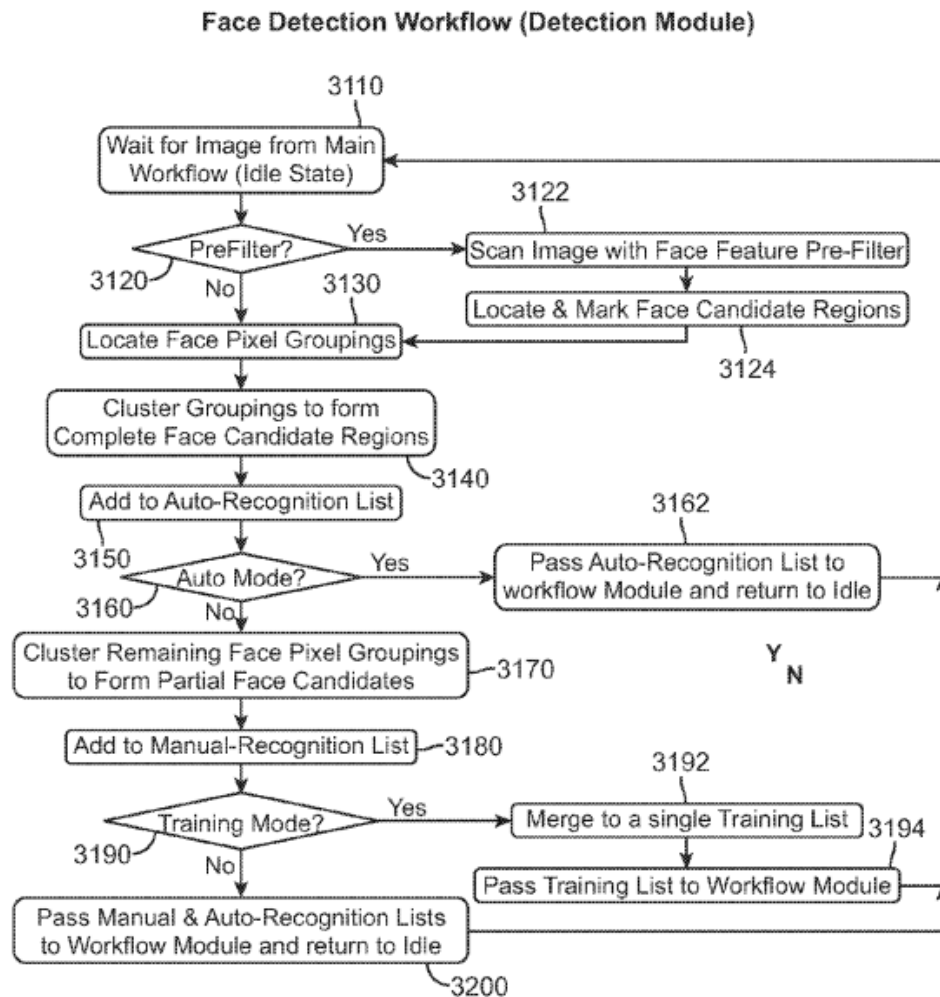


FIG. 3

A POSITA would have understood that, in general, the locations of faces and the locations of heads in images are the same. “Face detection” is often used interchangeably with “head detection” to refer to the process of identifying

individuals' faces or heads in images. Therefore, by teaching face detection, Steinberg teaches detecting a head appearing in one or more of the images.

234. Because a POSITA would have found it obvious to modify Nielsen-372's people meter based on Steinberg's teachings and Nielsen-372 teaches implementing the people meter using the architecture including memory, processor circuitry, and machine readable instructions (*supra* §VI.B.2[1a]), it is obvious that the processor circuitry executes the machine readable instructions to perform the operations of limitation [1d].

235. Furthermore, the '243 patent admits the prior art included approaches for identifying a number of people in a media exposure environment by discussing it in the "Background" section. *Supra* §VI.A.2[1d]; Ex-1001, 1:56-2:3. A POSITA would have known that a common approach of identifying people was based on their heads.

[1e] determine an orientation of the head with respect to the camera; and

236. The combination of Nielsen-372 and Steinberg teaches this limitation. Steinberg teaches **determining an orientation of the head with respect to the camera**. Steinberg teaches normalizing detected faces using a normalization module. Ex-1007, ¶0092 ("If faces are found, then they are preferably normalized and face prints are generated."), ¶0120 ("If however the image does contain face-candidate regions then further handling of the image is passed to ... the Face

Normalization module [1190][.]”), ¶0130 (“These stored images are then preferably processed by the ... face normalization module 1690[.]”), ¶0136 (“In a further embodiment, such normalizations may be advantageously achieved using a handheld imaging device.”), FIGs. 1(a)-(c). Normalizing includes determining a pose or orientation of the detected face. Ex-1007, ¶0120 (“These are then passed to the Face Normalization module [1190] which analyzes the lighting, pose, facial features and/or orientation of each face candidate region.”).

237. Steinberg teaches implementations of the normalization module that determines the pose or orientation of the face. Ex-1007, ¶0166 (“Now in order to further process this facial candidate region it is necessary to determine the facial pose which, in our preferred embodiment, is calculated based on the relative positions of the principle face features (eyes, nose, mouth and, occasionally, ears, hairline and chin).”), ¶0168 (“A number of approaches to pose determination are understood by those skilled in the art. In our preferred embodiment, we utilize a technique which is based on determining the 2-D projection of the 3-D eye-mouth triangle onto the face candidate region. The anthropometric data indicates that the region delineated by these face features forms an equilateral triangle in the plane of the face for the majority of persons. Thus, from the shape of the 2-D projection of this region, it is possible to determine the plane of the face, and thereby extrapolate the pose distortion [5030].”), FIG. 5(a) (emphasis added):

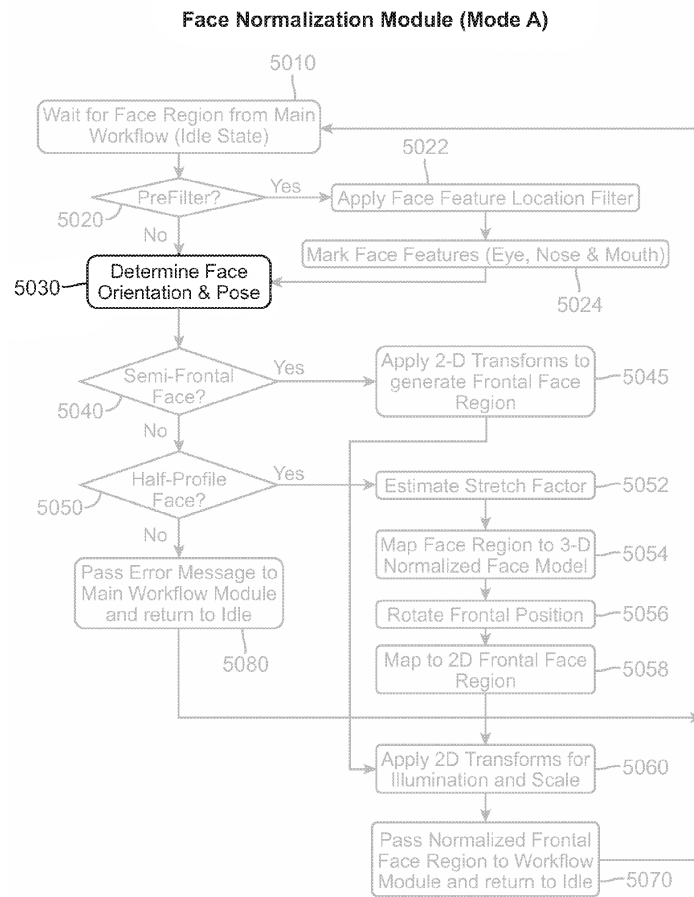


FIG. 5(a)

¶0176 (“Now in order to further process this facial candidate region it is necessary to determine the facial pose which, in our preferred embodiment, is calculated based on the relative positions of the principle face features (eyes, nose, mouth and, occasionally, ears, hairline and chin).”), FIG. 5(b) (emphasis added):

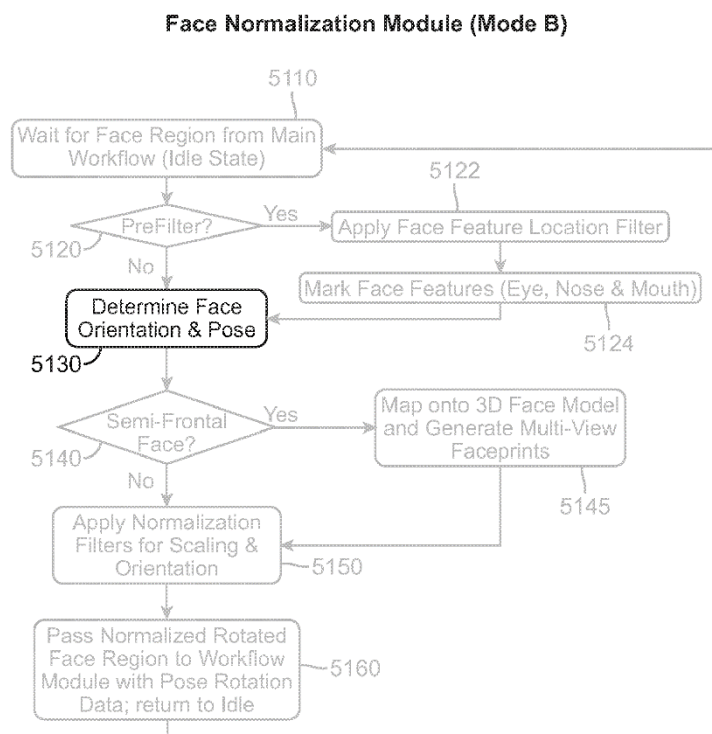


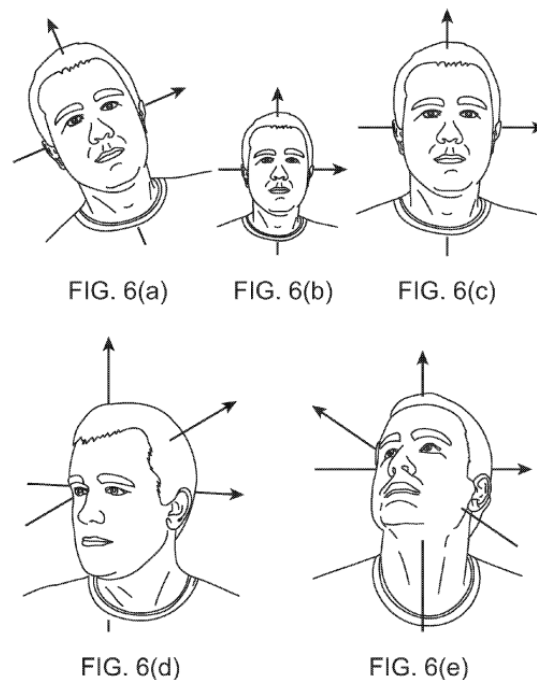
FIG. 5(b)

238. Steinberg explains that the determined facial pose or orientation is **an orientation of the head with respect to the camera** as it reflects different rotation angles of the head in an image with respect to the camera that captured the image (e.g., a frontal pose indicates that the face is directly facing the camera). Ex-1007, ¶0121 (“the pose or spatial orientation of a candidate face”), ¶0167 (“The most common of these, but also the more straight-forward of these to handle, as they relate to frontal, or semi-frontal face candidate regions, are in-plane rotation of the face, illustrated by FIG. 6(a) and scaling variations in the size of the face candidate region, illustrated in FIG. 6(b). The ideal face candidate region which is fully

frontal and is free of rotational or scaling distortions, is illustrated in FIG. 6(c).

The more difficult distortions of a face candidate region occur when the face is rotated to face to the left (or right) so that it appears in a semi-profile as illustrated in FIG. 6(d).... Equally difficult to correct is the case when the face candidate region is rotated so that the face looks up (or down) as illustrated in FIG. 6(e).”)

FIG. 6(a)-(e):



¶0173 (“In a simple case there would be face recognition datasets associated with left/right pitch rotations of 0, 15, 30, 45, 60, 70 and 90 degrees and for up/down yaw rotations of 0, 15, 30 and 45 degrees.”). While Steinberg discusses the orientations of the face, a POSITA would have understood that, because the face is a part of the head, the orientation of the face is controlled by that of the head. Steinberg thus teaches determining the orientation of the head.

239. Because a POSITA would have found it obvious to modify Nielsen-372's people meter based on Steinberg's teachings and Nielsen-372 teaches implementing the people meter using the architecture including memory, processor circuitry, and machine readable instructions (*supra* §VI.B.2[1a]), it is obvious that the processor circuitry executes the machine readable instructions to perform the operations of limitation [1e].

[1f] determine audience identification information based on a match of the head to a known person associated with the media exposure environment; and
--

240. The combination of Nielsen-372 and Steinberg teaches this limitation. Steinberg teaches **determining identification information** (e.g., identity region, identity grouping, face class, identity of person) **based on a match of the head to a known person**. Steinberg teaches face recognition based a match between a face region in an image and a known face. Ex-1007, ¶0005 (“Face recognition involves the analysis of such a ‘face region’ and its comparison with a database of known faces to determine if the unknown “face region” is sufficiently similar to any of the known faces to represent a high probability match.”), ¶0057 (““Face Recognition’ involves the art of matching an unknown facial region from an image with a set of ‘known’ facial regions.”).

241. Steinberg teaches performing face recognition by generating a faceprint and comparing it with faceprints stored in a database:

The role of the Face Recognition module [1140] is to generate a feature vector, or Faceprint, for a face candidate region identified in an image. This is returned to the Workflow Core module [1110] where the generated Faceprint is compared with existing identity regions and face classes. This is achieved by accessing the Image Classification Database [1160] via a Database Interface Module [1150]. If the Faceprint can be clearly associated with a single identity region then the face is considered to be positively identified and the Image Classification Database [1160] is updated accordingly.

Ex-1007, ¶¶0126-0127, ¶0130, FIGs. 1(a)-(c); *see also* ¶0092 (“Then, the images and face prints, together with the locations of the detected facial-regions and other face classifier parameter data, are made available for the face recognition module to extract, enhance and scan the facial regions and compares their faceprints with a database of ‘known faceprints’.”), ¶0096 (“When a faceprint is returned to the main workflow module from the face recognition module the “N” most probable matches are determined by searching the face recognition data tables in the image classification database and if the majority of these matches have the same “identity”, and the workflow module is running in “automatic mode”, then a successful “automatic match” is considered to have taken place.”).

242. The database stores faceprints in association with identity information—for example, a known person is associated with at least one identity region and multiple face classes each comprising faceprints. Ex-1007, ¶0099

(“Thus, each faceprint in the database module is preferably associated with a single identity, while each identity is associated with one or more face classes containing one or more faceprints.”), ¶0135 (“Each person may have multiple identities associated with him or her. Each identity may include one or more face classes.”), ¶0178 (“FIG. 9(a) shows a graphical representation of how multiple, distinct, face classes formed from collections of closely collocated faceprints can be used to define a unique region in Face Space which is associated with a particular Identity Region for a ‘known person’. In FIG. 9(a), a single Identity Region, ID1 [9000] is shown, which is composed of three separate Face Classes, FC1 [9010], FC2 [9020], and FC3 [9030]. FIG. 9(a) also illustrates three face prints which constitute the members of FC2 [9020], that is FP2A [9022], FP2B [9024] and FP2C [9026].”), FIG. 9(a):

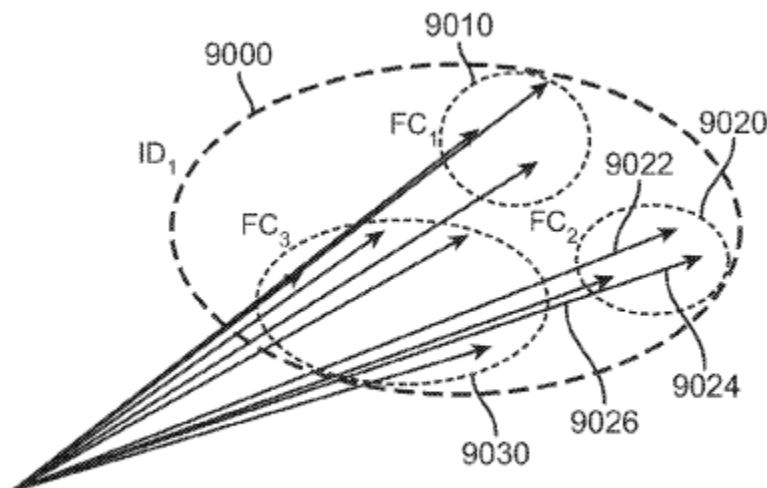


FIG. 9(a)

243. As faceprints are stored in association with identity information, a match of a faceprint in an image with a stored faceprint allows the determination of the identity information. Ex-1007, ¶0156 (“Firstly, a search of all the known Faceprints, or of a selected sub-set when working in semi-automatic mode, which have been recorded by the Image Classification Engine is made and the distance between the current Faceprint and each of these is calculated [4220].”), ¶0157:

After this distance calculation, the search algorithm first determines if the N closest matches lie within the same Face Class. If so, then the current Faceprint can be unambiguously associated with that particular Face Class [4230]. The system may be configured such that if the distance between the target faceprint and the nearest face print is much less than to any other face print, then that may be determined also to unambiguously identity a person. ... If it cannot be associated with a single Face Class, ... then the algorithm next determines if the N closest matches belong to the same Identity Grouping [4250], and if this is the case, an unambiguous match can still be achieved. ... Finally, if the current Faceprint cannot be unambiguously associated with a single Face Class, or a single Identity Grouping, it is still possible that the same “known person” may be associated with the N closest matches [4390]. If this is the same person then the new Faceprint can be unambiguously associated with that person [4395].

FIGs. 4(a)-(b) (emphasis added):

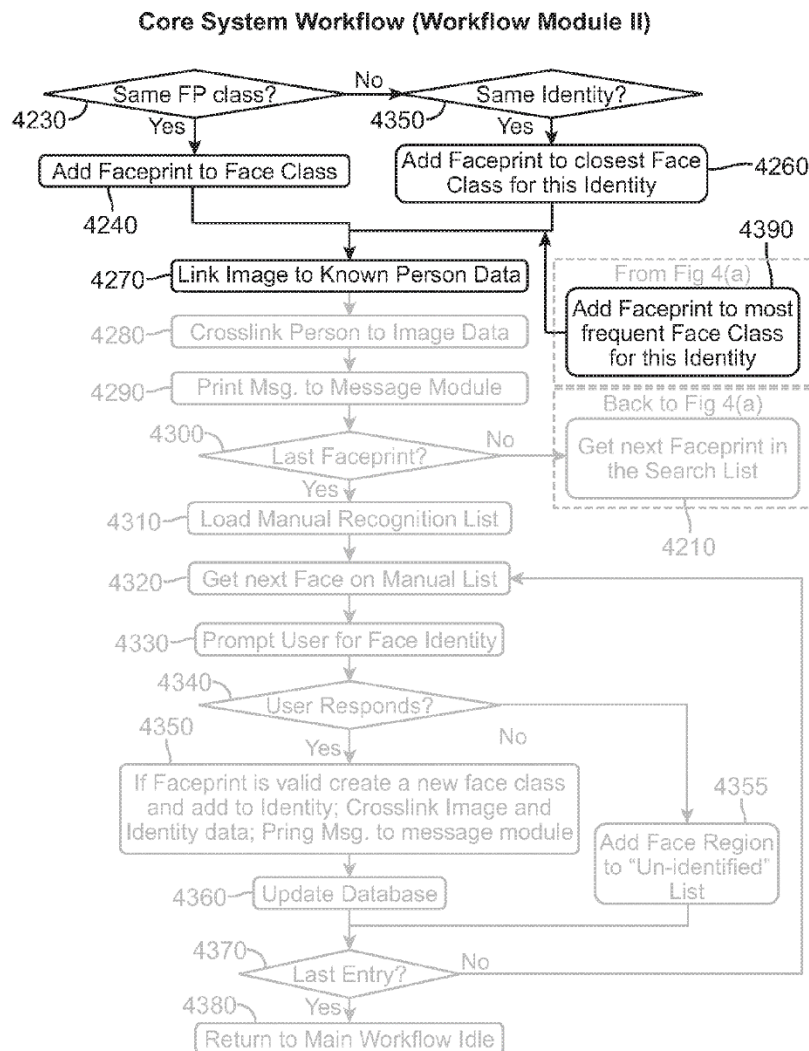


FIG. 4(b)

244. Because the face is an important part of the head, especially for image recognition purposes, a POSITA would have understood that Steinberg teaches finding a match of a head to a known person based on its face recognition teachings. The '243 patent discusses the recognition of heads and faces in an interchangeable manner consistent with this understanding. *See, e.g., Ex-1001,*

10:20-38 (“As described above, the frames stored in the frame cache 303 are used by the example hi-res trigger 302 to identify heads and/or faces in the environment 100 that, due to their orientation, may be identifiable by the person identifier 207.”).

245. The combination of Nielsen-372 and Steinberg teaches determine **audience** identification information based on a match of the head to a known person **associated with the media exposure environment**. A POSITA would have found it obvious to apply Steinberg’s teachings to the people meter such that the people meter identifies individuals in the audience based on images captured by a camera. *Supra* §VI.B.1. Nielsen-372 teaches that its people meter is located in a media exposure environment for identifying members of the audience. *Supra* VI.B.2[1d]; Ex-1006, ¶0003, ¶0048. When Steinberg’s method is applied in this scenario, a POSITA would have understood that the candidate individuals that can be identified are associated with the media exposure environment. It would have been obvious to compare a face or head in an image to those for persons associated with the media exposure environment and identify audience identification information based on a match. This can be achieved by, for example, filling the database with faceprints from persons associated with the media exposure environment or compare a new faceprint with a subset of the database corresponding to the media exposure environment, which a POSITA would have

recognized as obvious and readily doable. *See, e.g.*, Ex-1007, ¶0156 (“a search of all the known Faceprints, or of a selected sub-set...”).

246. Because a POSITA would have found it obvious to modify Nielsen-372’s people meter based on Steinberg’s teachings and Nielsen-372 teaches implementing the people meter using the architecture including memory, processor circuitry, and machine readable instructions (*supra* §VI.B.2[1a]), it is obvious that the processor circuitry executes the machine readable instructions to perform the operations of limitation [1f].

247. Furthermore, the ’243 patent admits that audience identification based on facial recognition was known in the prior art by discussing it in the “Background” section. *Supra* §VI.A.2[1f]; Ex-1001, 1:56-2:3.

[1g] network interface circuitry to output a signal indicative of the content identifying data and the audience identification information to a data collection facility.

248. The combination of Nielsen-372 and Steinberg teaches this limitation. Nielsen-372 teaches a **network interface circuitry** (e.g., interface circuit). Nielsen-372 teaches “implementing the apparatus and methods” it discloses using the computer as shown in Figure 29. Ex-1006, ¶0135, FIG. 29 (emphasis added):

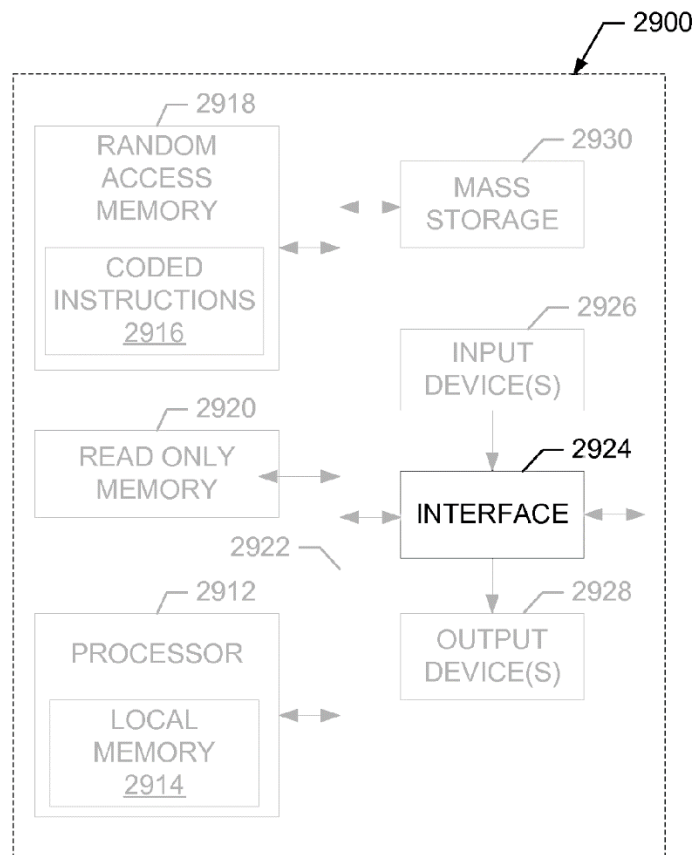


FIG. 29

The computer includes “includes a conventional interface circuit 2924.” Ex-1006, ¶0135. Nielsen-372 teaches the interface circuit being a network interface circuitry (e.g., when it is an Ethernet interface). Ex-1006, ¶0135 (“The interface circuit 2924 may be implemented by any type of well known interface standard, such as an Ethernet interface, a universal serial bus (USB), and/or a third generation input/output (3GIO) interface.”).

249. The combination teaches that the network interface circuitry **outputs a signal indicative of the content identifying data and the audience**

identification information to a data collection facility (e.g., central facility, central office, data processing facility, image collection database). Nielsen-372 teaches that the local metering system, which teaches at least part of the audience measurement system (*supra* §VI.B.2.1[p]), outputs viewing records to a central office or data processing facility over a network connection. Ex-1006, ¶0036 (“For example, in a statistically selected household having one or more home entertainment systems 102, the metering unit 124 may be implemented as a single home unit and one or more site units. In such a configuration, the single home unit may perform the functions of storing data and forwarding the stored data to a central facility for subsequent processing.”), ¶0041 (“The metering unit 124 typically collects a set of viewing records and transmits the collected viewing records over a connection 140 to a central office or data processing facility (not shown) for further processing or analysis. The connection 140 may be a telephone line, a return cable television connection, an RF or satellite connection, an internet connection or the like.”), FIG. 1 (emphasis added):

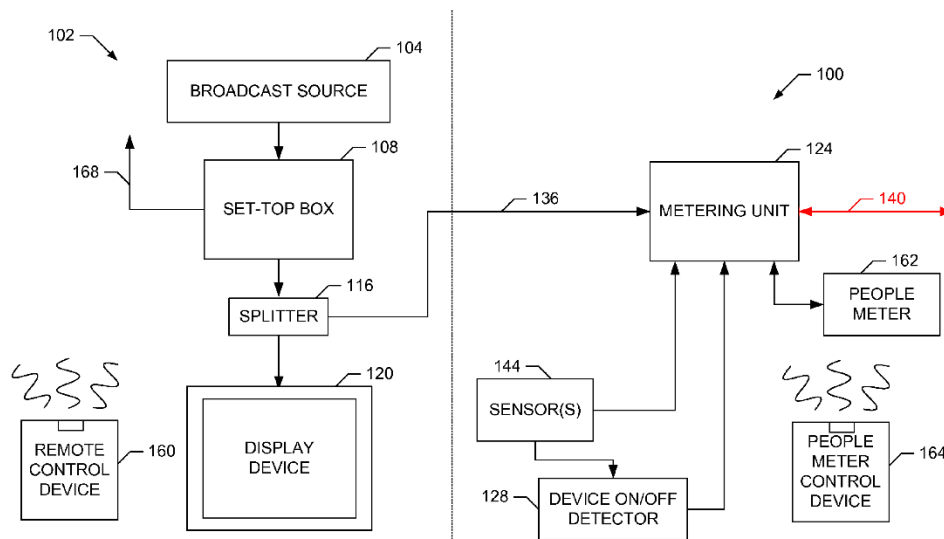


FIG. 1

250. Steinberg further teaches a local device (e.g., a camera) outputting data to an image classification database. Ex-1007, ¶0131 (“The camera shown is configured for preferably wirelessly networking over a network 1720, e.g., over the internet, to off-camera data storage 1750. The wireless network preferably communicates data from the camera to the off-camera storage by a connection 1745 for adding data to a backup image classification database 1730 and/or a personal image collection 1740.”).

251. The ’243 patent does not include any limitation on the meaning of a “data collection facility.” A POSITA would have understood that the central office or data processing facility in Nielsen-372 and the image classification database in Steinberg are data collection facilities because they collect and store data. A POSITA would have understood that data commonly transmit over networks in the

form of electromagnetic signals, therefore Nielsen-372 and Steinberg teach outputting a signal to a data collection facility.

252. The combination of Nielsen-372 and Steinberg teaches the audience measurement system obtaining content identifying data and determining audience identification information. *Supra* §VI.B.2[1b-1f]. In light of Nielsen-372's teachings of a network interface circuitry and Nielsen-372 and Steinberg's teachings of outputting data to a data collection facility, a POSITA would have found it obvious that the audience measurement system includes network interface circuitry to output a signal indicative of the content identifying data and the audience identification information to a data collection facility.

3. Claim 9

253. It is my opinion that a POSITA would have found claim 9 obvious based on the combined teachings of Nielsen-372 and Steinberg.

9[p]. An audience measurement system, comprising:
--

254. As explained for claim 1, the combination of Nielsen-372 and Steinberg teaches an audience measurement system. *Supra* §VI.B.2.1[p].

[9a] an audience measurement device at a media exposure environment, the audience measurement device to:

255. As explained for claim 1, Nielsen-372 teaches an audience measurement system at least based on its disclosure of a local metering system.

Supra §VI.B.2.1[p]. The local metering system includes components such as a metering unit, a display device ON/OFF detector, and a people meter. *Supra* §VI.B.2.1[p]. A POSITA would have found it obvious to apply Steinberg's teachings to the people meter such that the people meter identifies individuals in the audience based on images captured by a camera. *Supra* §VI.B.1. The local metering system is placed in the same environment as a home entertainment system. *Supra* §VI.B.2.1[p].

256. In light of their functionalities, a combination of the aforementioned components of the local metering system constitutes **an audience measurement device**, as the term is used in the '243 patent. The '243 patent describes example implementations of the audience measurement device, but clarified that the audience measurement device is broad and not limited to the example implementations. Ex-1001, 8:7-43; *supra* §VI.A.7[9a]. Because the components of the local metering system perform operations that measure exposure of an audience to media content, their combination constitutes an audience measurement device. *Supra* §VI.B.2.1[p]; Ex-1006, ¶0044, ¶0048. As explained for claim 1, the environment of the home entertainment system is a media exposure environment. *Supra* §VI.A.2.1[p]. Therefore, Nielsen-372 teaches an audience measurement device *at a media exposure environment*.

257. As explained above, a POSITA would have found it obvious to implement components of the local metering system using the architecture of memory, machine readable instructions, and processor circuitry, such that the processor circuitry executes the machine readable instructions to perform operations directed to identifying a television program and audience members. *Supra* §VI.B.2[1a]. Because Nielsen-372 teaches that the audience measurement device includes these components of the local metering system, a POSITA would have found it obvious to implement the audience measurement device using the structure of memory, machine readable instructions, and processor circuitry, such that the processor circuitry executes the machine readable instructions to perform operations directed to identifying a television program and audience members. As explained below, the combination of Nielsen-372 and Steinberg teaches the audience measurement device performing the operations recited in limitations [9b]-[9g]. *Infra* §VI.B.3[9b-9g].

[9b] generate an audio signature of media content presented by a television within the media exposure environment;

258. As explained for claim 1, Nielsen-372 teaches that the processor circuitry of the audience measurement system executes the machine readable instructions to generate an audio signature of media content presented by a television within the media exposure environment. *Supra* §VI.B.2[1b]. The

audience measurement device includes the processor circuitry and machine readable instructions. *Supra* §VI.B.3[9a]. Therefore, the combination of Nielsen-372 and Steinberg teaches that the audience measurement device generates an audio signature of media content presented by a television within the media exposure environment.

[9c] obtain content identifying data corresponding to the presented media content, the content identifying data based on the audio signature;

259. As explained for claim 1, Nielsen-372 teaches that the processor circuitry of the audience measurement system executes the machine readable instructions to obtain content identifying data corresponding to the presented media content, the content identifying data based on the audio signature. *Supra* §VI.B.2[1c]. The audience measurement device includes the processor circuitry and machine readable instructions. *Supra* §VI.B.3[9a]. Therefore, the combination of Nielsen-372 and Steinberg teaches that the audience measurement device obtains content identifying data corresponding to the presented media content, the content identifying data based on the audio signature.

[9d] while the media content corresponding to the content identifying data is presented by the television, collect a first image of the media exposure environment with a camera;

260. As explained for claim 1, the combination of Nielsen-372 and Steinberg teaches that the processor circuitry of the audience measurement system

executes the machine readable instructions to obtain a sequence of images of the media exposure environment with a camera while the media content corresponding to the content identifying data is presented by the television. *Supra* §VI.B.2[1d]. The combination of Nielsen-372 and Steinberg further teaches detecting a head appearing in one or more of the images. *Supra* §VI.B.2[1d]. A POSITA would have understood that “**collect**” is used interchangeably with “obtain” in the context of a camera capturing images. The one or more images in which a head is detected includes at least one image, which constitutes a **first image** that is collected. The audience measurement device includes the processor circuitry and machine readable instructions. *Supra* §VI.B.3[9a]. Therefore, the combination of Nielsen-372 and Steinberg teaches that, while the media content corresponding to the content identifying data is presented by the television, the audience measurement device collects a first image of the media exposure environment with a camera.

[9e] attempt to detect a head in the first image;

261. As explained for claim 1, the combination of Nielsen-372 and Steinberg teaches that the processor circuitry of the audience measurement system executes the machine readable instructions to detect a head in an image. *Supra* §VI.B.2[1d]. Steinberg teaches applying this approach to a sequence of images such that a head is detected in one or more of the images. *Supra* §VI.B.2[1d]. Because this approach determines if a head can be detected and does not guarantee

detecting a head in each image, it includes attempts to detect a head. *See, e.g.*, Ex-1007, ¶0118 (“This module employs a means to determine if facial candidate regions are present in the new image.”). As the first image is satisfied by one of the one or more images (*supra* §VI.B.3[9d]), it follows that the combination of Nielsen-372 and Steinberg teaches **attempting to** detect the head in the first image. The audience measurement device includes the processor circuitry and machine readable instructions. *Supra* §VI.B.3[9a]. Therefore, the combination of Nielsen-372 and Steinberg teaches that the audience measurement device attempts to detect a head in the first image.

[9f] determine an orientation of the head with respect to the camera;

262. As explained for claim 1, the combination of Nielsen-372 and Steinberg teaches that the processor circuitry of the audience measurement system executes the machine readable instructions to determine an orientation of the head with respect to the camera. *Supra* §VI.B.2[1e]. The audience measurement device includes the processor circuitry and machine readable instructions. *Supra* §VI.B.3[9a]. A POSITA would have found it obvious for the audience measurement device to determine the head orientation of the audience, which is part of audience information. Therefore, the combination of Nielsen-372 and Steinberg teaches that the audience measurement device determines an orientation of the head with respect to the camera.

[9g] determine audience identification information based on an indication that the head matches a known person associated with the audience measurement system; and

263. As explained for claim 1, the combination of Nielsen-372 and Steinberg teaches that the processor circuitry of the audience measurement system executes the machine readable instructions to determine audience identification information based on a match of the head to a known person associated with the media exposure environment. *Supra* §VI.B.2[1f]. Steinberg teaches comparing a faceprint of a detected face with faceprints stored in a database to determine a match. *Supra* §VI.B.2[1f]; Ex-1007, ¶0092, ¶0096, ¶¶0126-0127, ¶0130, FIGs. 1(a)-(c). Steinberg further teaches that the faceprints in the database are stored in association with identity information and associating the faceprint with identity information when a match is found. *Supra* §VI.B.2[1f]; Ex-1007, ¶0157. These disclosure of Steinberg teaches **an indication that the head matches a known person**, on which audience identification is based.

264. The combination of Nielsen-372 and Steinberg teaches that the audience measurement system is located in and obtains exposure data for a media exposure environment. *Supra* §VI.B.2.1[p]. Because the audience measurement system is set up to monitor the media exposure environment, known persons associated with the media exposure environment are associated with the audience measurement system. Because the combination of Nielsen-372 and Steinberg

teaches determining a match with a known person associated with the media exposure environment, they teach determining the match with **a known person associated with the audience measurement system**. The audience measurement device includes the processor circuitry and machine readable instructions. *Supra* §VI.B.3[9a]. Therefore, the combination of Nielsen-372 and Steinberg teaches that the audience measurement device determines audience identification information based on an indication that the head matches a known person associated with the audience measurement system.

[9h] a data collection facility to obtain the content identifying data and the audience identification information from the audience measurement device.
--

265. As explained for claim 1, the combination of Nielsen-372 and Steinberg teaches a data collection facility to which a network interface circuitry outputs a signal indicative of the content identifying data and the audience identification information. *Supra* §VI.B.2[1g]. Nielsen-372 teaches that the data collection facility (e.g., central facility, central office, data processing facility) further processes the obtained data and information regarding an audience's exposure to media content. *Supra* §VI.B.2[1g]; Ex-1006, ¶0036 (“[T]he single home unit may ... forward[] the stored data to a central facility for subsequent processing.”), ¶0041 (“The metering unit 124 typically ... transmits the collected viewing records ... to a central office or data processing facility (not shown) for

further processing or analysis.”). In light of Nielsen-372’s teachings, a POSITA would have found it obvious to structure the audience measurement system such that it includes the data collection facility in addition to the local metering system.

266. Therefore, the combination of Nielsen-372 and Steinberg teaches that the audience measurement system comprises a data collection facility to obtain the content identifying data and the audience identification information from the audience measurement device.

4. Claim 16

267. It is my opinion that a POSITA would have found claim 16 obvious based on the combined teachings of Nielsen-372 and Steinberg.

16[p]. A method for obtaining exposure data for a media exposure environment, the method comprising:

268. As explained for claim 1, the combination of Nielsen-372 and Steinberg teaches an audience measurement system to obtain exposure data for a media exposure environment. *Supra* §VI.B.2.1[p]. The audience measurement system performs multiple steps. *Supra* §VI.B.2[1a-1g]. Both Nielsen-372 and Steinberg contemplate methods that include operations they teach. *See, e.g.*, Ex-1006, Abstract (“Display device ON/OFF detection methods and apparatus are disclosed.”), ¶0134 (“FIG. 29 is a block diagram of an example computer 2900 capable of implementing the apparatus and methods disclosed herein.”), ¶0144 (“[T]his patent covers all methods, apparatus and articles of manufacture fairly

falling within the scope of the appended claims either literally or under the doctrine of equivalents.”); Ex-1007, Abstract (“A processor-based system operating according to digitally-embedded programming instructions performs a method....”), ¶0090 (“A system, method and techniques are now described for applying digital image processing using face detection and face recognition techniques to identify the persons contained within a digital image, which may be part of a personal or other image collection.”). In light of teachings of the combination of Nielsen-372 and Steinberg regarding operations performed by the audience measurement system, a POSITA would have understood that the combination teaches a **method** for obtaining exposure data for a media exposure environment. As explained below, the combination of Nielsen-372 and Steinberg further teaches that the method includes the operations recited in claim 16. *Infra* §VI.B.4[16a-16g].

[16a] generating, using a program detector, an audio signature of media content presented by a television within the media exposure environment;
--

269. As explained for claim 1, Nielsen-372 teaches that the audience measurement system generates an audio signature of media content presented by a television within the media exposure environment. *Supra* §VI.B.2[1b]. The audience measurement system comprises a sensor that receives audio signals of a displayed program and a metering unit and/or display device ON/OFF detector that

identifies a television program. *Supra* §VI.B.2[1b-1c]. The sensor and the metering unit and/or display device ON/OFF detector, executing machine readable instructions related to program detection, together form a program detector, as the term is used in the '243 patent. The '243 patent does not define “program detector” but states that an “example program detector ... detects presentation(s) of media content in the media exposure environment 100 and collects identification information associated with the detected presentation(s).” Ex-1001, 7:1-41. The functionality of the program detector in the '243 patent is consistent with that of the sensor and the metering unit and/or display device ON/OFF detector in Nielsen-372. A POSITA would have found it obvious to implement Nielsen-372's metering unit and/or display device ON/OFF detector using the architecture that includes memory, machine readable instructions, and processor circuitry. *Supra* §VI.B.2[1a]. A POSITA would have found it obvious to arrange the hardware and software of the sensor and the metering unit and/or display device ON/OFF detector to form a program detector. Accordingly, Nielsen-372 teaches performing the operation of limitation [16a] using a program detector. Therefore, the combination of Nielsen-372 and Steinberg teaches generating, using a program detector, an audio signature of media content presented by a television within the media exposure environment.

[16b] obtaining, based on the audio signature, content identifying data corresponding to the presented media content;

270. As explained for claim 1, Nielsen-372 teaches that the audience measurement system obtains content identifying data corresponding to the presented media content, the content identifying data based on the audio signature. *Supra* §VI.B.2[1c]. Therefore, Nielsen-372 teaches obtaining, based on the audio signature, content identifying data corresponding to the presented media content.

[16c] while the media content corresponding to the content identifying data is presented by the television, capturing first and second images of the media exposure environment with a camera;

271. As explained for claim 1, the combination of Nielsen-372 and Steinberg teaches that the audience measurement system obtains a sequence of images of the media exposure environment with a camera while the media content corresponding to the content identifying data is presented by the television. *Supra* §VI.B.2[1d]. The combination of Nielsen-372 and Steinberg further teaches detecting a head appearing in one or more of the images. *Supra* §VI.B.2[1d]. The images are **captured** by a camera. Ex-1007, ¶¶0010, ¶¶0128-0129. When there are more than one images in which a head is detected, the one or more images includes **first and second images** that are captured. Therefore, the combination of Nielsen-372 and Steinberg teaches, while the media content corresponding to the content

identifying data is presented by the television, capturing first and second images of the media exposure environment with a camera.

[16d] analyzing, using an audience detector, the first image to attempt to detect a head in the first image;

272. As explained for claims 1 and 9, the combination of Nielsen-372 and Steinberg teaches that the audience measurement system analyzes the sequence of images including the first image and attempts to detect a head in the first image. *Supra* §VI.B.2[1d], §VI.B.3[9e].

273. Nielsen-372 teaches that the audience measurement system includes a people meter. Ex-1006, ¶0048. A POSITA would have found it obvious to apply Steinberg's teachings to the people meter such that the people meter identifies individuals in the audience based on images captured by a camera. *Supra* §VI.B.1. The camera and the people meter, executing machine readable instructions related to audience detection, together form an audience detector, as the term is used in the '243 patent. The '243 patent does not define "audience detector" but states that "[t]he example audience detector 200 includes an image sensor 204, a people counter 206, a person identifier 207, a time stamper 208, and a memory 210" and "the audience detector 200 may have additional or alternative methods and/or components to identify people in the frames captured by the image sensor 204[.]" Ex-1001, 5:29-6:44. The functionality of the program detector in the '243 patent is

consistent with that of the camera and people meter as taught by the combination of Nielsen-372 and Steinberg. Furthermore, the '243 patent broadly describes the “audience detector” as “may be implemented by hardware, software, firmware and/or any combination of hardware, software and/or firmware.” Ex-1001, 8:7-43. A POSITA would have found it obvious to implement the people meter using the architecture that includes memory, machine readable instructions, and processor circuitry. *Supra* §VI.B.2[1a]. A POSITA would have found it obvious to arrange the hardware and software of the camera and people meter to form an audience detector. Therefore, the combination of Nielsen-372 and Steinberg teaches analyzing, using an audience detector, the first image to attempt to detect a head in the first image.

[16e] determining, using the audience detector, an orientation of the head with respect to the camera;
--

274. As explained for claim 1, the combination of Nielsen-372 and Steinberg teaches that the audience measurement system determines an orientation of the head with respect to the camera. *Supra* §VI.B.2[1e]. The combination teaches that the camera and the people meter, executing machine readable instructions related to audience detection, together form an audience detector. *Supra* §VI.B.4[16d]. Because the determination of head orientation is related to identifying audience members, it would have been obvious to use the audience

detector to perform the operation of limitation [16e]. Therefore, the combination of Nielsen-372 and Steinberg teaches determining, using the audience detector, an orientation of the head with respect to the camera.

[16f] determining, using the audience detector, audience identification information based on an indication that the head matches a known person associated with the audience detector; and
--

275. As explained for claim 9, the combination of Nielsen-372 and Steinberg teaches that the audience measurement system determines audience identification information based on an indication that the head matches a known person associated with the audience measurement system. *Supra* §VI.B.3[9g].

276. The combination teaches that the camera and the people meter, executing machine readable instructions related to audience detection, together form an audience detector. *Supra* §VI.B.4[16d]. Because the determination of audience identification information is a functionality of the people meter, it would have been obvious to use the audience detector to perform the operation of limitation [16f].

277. As explained above, the audience detector is formed by components of the audience measurement system. *Supra* §VI.B.4[16d]. Therefore, known persons associated with the audience measurement system are also associated with audience detector. Because the combination teaches determining a match with a known person associated with the audience measurement system, they teach

determining the match with **a known person associated with the audience detector**. Therefore, the combination of Nielsen-372 and Steinberg teaches determining, using the audience detector, audience identification information based on an indication that the head matches a known person associated with the audience detector.

[16g] providing the content identifying data and the audience identification information to a data collection facility.

278. As explained for claim 1, the combination of Nielsen-372 and Steinberg teaches that the audience measurement system outputs a signal indicative of the content identifying data and the audience identification information to a data collection facility. *Supra* §VI.B.2[1g]. A POSITA would have understood that one well-known way to provide the content identifying data and the audience identification information is to output a signal thereof. Therefore, the combination of Nielsen-372 and Steinberg teaches providing the content identifying data and the audience identification information to a data collection facility.

C. Ground 3: Claims 1, 4-6, 8, 9, 11-14, 16, and 18-20 are Rendered Obvious in View of Nielsen-372, Steinberg, and Tian

279. It is my opinion that a POSITA would have found Claims 1, 4-6, 8, 9, 11-14, 16, and 18-20 obvious based on the teachings of Nielsen-372, Steinberg, and Tian.

1. Combination of Nielsen-372, Steinberg, and Tian

280. As explained above, a POSITA would have found it obvious and been motivated to combine Nielsen-372 and Steinberg. *Supra* §VI.B.1. The combination of Nielsen-372 and Steinberg teaches an audience measurement system that determines media content presented by a television and identifies individuals in the audience. *Supra* §VI.B.2. The combination applies Steinberg's technique of identifying individuals based on face detection, normalization, and recognition to Nielsen-372's people meter. *Supra* §VI.B.1. Similar to Steinberg, Tian teaches techniques for image processing and facial analysis. Ex-1005, 000002, 000005. Tian teaches processing reduced-resolution images and studies the effects of the reduction of resolution on the performance of different facial analysis techniques, including face detection and head pose estimation. Ex-1005, Abstract, 000005.

281. A POSITA would have found it obvious to add Tian to the combination of Nielsen-372 and Steinberg because Steinberg and Tian are directed to similar functionalities (e.g., images process and facial analysis) and address similar software-based tasks (e.g., identifying a head in an image). As explained below, Tian provides multiple techniques that provide performance improvements for facial analysis methods, such as those taught by Steinberg. A POSITA would have understood that Tian's techniques are applicable to Steinberg's teachings and would have been motivated to use these techniques to improve Steinberg. A

POSITA would have been motivated to apply known facial analysis techniques, including the techniques taught by Tian, to the people meter of Nielsen-372.

282. Tian teaches performing facial analysis on sequences of images of subjects. *See, e.g.*, Ex-1005, 000005 (“The database contains 704 image sequences from 97 subjects. ... The length of the image sequences are varying from 9 to 47 frames.”). A POSITA would have found it obvious to apply Tian’s teaching to the teachings of Nielsen-372 and Steinberg such that the audience measurement system analyzes sequences of images to identify audience members. Steinberg teaches identifying members of an audience by analyzing one or multiple images. *See supra* §VI.B.2[1d-1f]. A POSITA would have understood that this is compatible with Tian’s teaching of analyzing a sequence of images and it would have been obvious to implement such an approach in the people meter.

283. A POSITA would have been motivated to apply Tian’s teaching of performing facial analysis on a sequence of images because of the benefits provided by such an approach. First, a POSITA would have known that facial analysis usages are always striving for higher levels of accuracy, and that operations such as head detection and face recognition generally correspond to certain levels of accuracy. As contemporaneous writings recognized, the accuracy of identifying a person can be improved by analyzing multiple images including the same person and aggregating the analysis results from the multiple images.

See, e.g., Ex-1022, 000007. This would have motivated a POSITA to analyze multiple images in order to achieve a higher confidence level for the identification than a single image. Second, a POSITA would have understood that a single image captures a person's facial features at a particular point in time. The person's facial expression and orientation, as captured by that image, may or may not be suitable for face recognition. For example, the person may close eyes or face a direction other than the camera, and in those cases, face recognition with a high accuracy may not be feasible. *Ex-1010, ¶0003* (“[A] typical face detector can only accurately detect faces in a semi-frontal pose.”). Obtaining and analyzing a sequence of images allows the selection of images in which the facial features are suitable for face recognition and selective performance of face recognition on the suitable images. A POSITA would have understood that this would have improved the accuracy of the audience identification.

284. Tian teaches determining a face/head location in a first image and only tracking changes to face features in subsequent images. *See, e.g., Ex-1005, 000001* (“Face acquisition is a processing stage to automatically find the face region for the input images or sequences. It can be a face detector to ... just detect face in the first frame and then track the face in the remainder of the video sequence.”), *000003* (“Given an image sequence, the region of the face and approximate location of individual face features are detected automatically in the

initial frame. ... Then, all face feature changes are automatically tracked in the image sequence.”). Under this approach, if no change is detected, the face is determined to be at the same location in subsequent images. A POSITA would have found it obvious to apply Tian’s teaching to the combination of Nielsen-372 and Steinberg such that when the audience measurement system processes a sequence of images, it determines a face or head location in a first image and use the determined location for subsequent images. For example, the audience measurement system analyzes a portion in a subsequent image corresponding to the location for face recognition.

285. A POSITA would have found it obvious to apply this approach since this approach is taught in contemporaneous writings. Ex-1010, ¶0003 (“Once a face is detected, its location is recorded and a localized region around that face is scanned by a face detector in the next frame. Thus, once a face is initially detected, it can be accurately tracked from frame to frame without a need to run a face detector across the entire image.”). A POSITA would have been motivated to apply the teachings of Tian because this approach has the benefit of saving processing resources and time as the process of face detection is performed once and not repeated.

286. Tian teaches reducing the resolution of an image to obtain reduced-resolution images and performing analysis on the reduced-resolution images such

as face detection and head pose estimation. *See, e.g.*, Ex-1005, Abstract, 000002, 000005. Tian teaches that the accuracy of face detection and head pose estimation can be maintained for reduced-resolution images. *See, e.g.*, Ex-1005, 000005 (“The results show that both face detectors [23, 29] can detect all faces for head region in resolution of 288x384, 144x192, 72x96, and 36x48, but failed for resolution of 18x24. Compared to face detection, head pose estimation can detect 98% faces for all the resolutions.”).

287. A POSITA would have found it obvious to apply Tian’s teaching to the combination of Nielsen-372 and Steinberg such that the audience measurement system reduces the resolution of an image and performs face detection or pose estimation on the reduced-resolution image. Tasks such as face detection can be resource-consuming. Some approaches of face detection may require analyzing all pixels of an image to search for a face. The processing resources and time can be reduced if the resolution of the image is reduced. Ex-1014, 2:43-53 (teaching that “calculations are avoided of a complete highest resolution integral image for every acquired image in an image stream” in face tracking, thereby “reduces processing overhead for face detection[.]”); Ex-1015, ¶0022 (“By reducing the resolution of the saturation of the image, computational requirements are substantially reduced and a relatively simple method may be used.”). 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 according to Tian's teachings. Ex-1005, 000005. A POSITA would have been motivated to apply Tian's teachings of analyzing reduced-resolution images to achieve the benefit of savings in processing resources and time.

288. It was known that the performance of face recognition may be adversely affected by the reduction of image resolution. Ex-1027, 000004-05. Therefore, a POSITA would have been motivated to perform certain facial analysis techniques (e.g., face detection, head pose estimation) on reduced-resolutions images, while other techniques (e.g., face recognition) on full-quality images.

289. A POSITA would have found it obvious and been motivated to apply the teachings of Tian to the combination of Nielsen-372 and Steinberg because this is an application of known techniques to a known system and method ready for improvement. As explained for Ground 2, it would have been obvious to apply Steinberg's method of identifying individuals using facial analysis to Nielsen-372's people meter. *Supra* §VI.B.1. Tian's techniques (e.g., analyzing a sequence of images, analyzing reduced-resolution images, tracking the location of head in a sequence of images) would have improved Steinberg's method (e.g., improving accuracy, saving processing resources). A POSITA would have found it obvious to further apply the known techniques of Tian to the known system according to the combination of Nielsen-372 and Steinberg to achieve these improvements.

Applying Tian's teachings to the combination of Nielsen-372 and Steinberg would have yielded predictable results because Tian provides experimental assessments of its techniques, such as the accuracy of face detection and head pose estimation for images of different resolutions.

290. A POSITA would have had a reasonable expectation of success when combining the teachings of Nielsen-372, Steinberg, and Tian, given the overlap between Steinberg and Tian and that Tian's image processing techniques were well known. As explained above, the combination simply uses Tian's teachings about image processing and facial analysis, in the manner taught by Tian, to implement functionality in Nielsen-372. A POSITA would have been more than capable of implementing the image processing techniques taught by Tian, such as creating reduced-resolution images by down-sampling an original image, face detection, head pose estimation, and face tracking. *See supra* §I. Furthermore, a POSITA would have been motivated and would have had a reasonable expectation of success in combining the references, as explained above, because the combination would not have changed the principle of operation for any of the teachings of the references relied upon in the combination. Here, Tian's techniques are used in a conventional manner and applied to Nielsen-372 and Steinberg in a known manner.

2. Claim 1

291. It is my opinion that a POSITA would have found claim 1 obvious based on the combined teachings of Nielsen-372, Steinberg, and Tian.

1[p]. An audience measurement system to obtain exposure data for a media exposure environment,

292. As discussed for Ground 2, to the extent the preamble limits claim 1, the combination of Nielsen-372 and Steinberg teaches the preamble. *Supra* §VI.B.2.1[p].

293. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders the preamble obvious.

[1a] the audience measurement system comprising: memory; machine readable instructions; and processor circuitry to execute the machine readable instructions to:

294. As discussed for Ground 2, the combination of Nielsen-372 and Steinberg teaches limitation [1a]. *Supra* §VI.B.2[1a].

295. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[1b] generate an audio signature of media content presented by a television within the media exposure environment;

296. As discussed for Ground 2, Nielsen-372 teaches limitation [1b]. *Supra* §VI.B.2[1b].

297. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[1c] obtain content identifying data corresponding to the presented media content, the content identifying data based on the audio signature of the media content presented by the television within the media exposure environment;
--

298. As discussed for Ground 2, Nielsen-372 teaches limitation [1c].
Supra §VI.B.2[1c].

299. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[1d] analyze a sequence of images of the media exposure environment to detect a head appearing in one or more of the images, the sequence of images obtained by a camera while the media content corresponding to the content identifying data is presented by the television;
--

300. As discussed for Ground 2, the combination of Nielsen-372 and Steinberg teaches limitation [1d]. *Supra* §VI.B.2[1d]. Tian further teaches obtaining a sequence of images of human subjects using a camera. Ex-1005, 000005; *supra* §VI.A.2[1d]. To the extent it is argued that Nielsen-372 and Steinberg does not teach “a *sequence* of images,” a POSITA would have found it obvious and been motivated to apply Tian’s teachings such that a sequence of images is obtained for each person and analyzed. *Supra* §VI.C.1 (explaining why it would have been obvious to obtain and analyze a sequence of images; that analysis is incorporated here).

301. Tian also teaches analyzing a sequence of images to detect a head appearing in one or more of the images using face detection techniques. Ex-1005, 000002 (“Many face detection methods have been developed to detect faces in an arbitrary scene....”), 000005 (“All images in 704 sequences from 97 subjects are used for face detection and head pose estimation. The results show that both face detectors [23, 29] can detect all faces for head region in resolution of 288x384, 144x192, 72x96, and 36x48, but failed for resolution of 18x24.”).

302. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[1e] determine an orientation of the head with respect to the camera; and

303. As discussed for Ground 2, the combination of Nielsen-372 and Steinberg teaches limitation [1e]. *Supra* §VI.B.2[1e]. Tian further teaches determining an orientation of the head with respect to the camera. *Supra* §VI.A.2[1a]; Ex-1005, 000001-03.

304. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[1f] determine audience identification information based on a match of the head to a known person associated with the media exposure environment; and

305. As discussed for Ground 2, the combination of Nielsen-372 and Steinberg teaches limitation [1f]. *Supra* §VI.B.2[1f].

306. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[1g] network interface circuitry to output a signal indicative of the content identifying data and the audience identification information to a data collection facility.

307. As discussed for Ground 2, the combination of Nielsen-372 and Steinberg teaches limitation [1g]. *Supra* §VI.B.2[1g].

308. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

3. Claim 4

309. It is my opinion that a POSITA would have found claim 4 obvious based on the combined teachings of Nielsen-372, Steinberg, and Tian.

4[p]. The audience measurement system of claim 1, wherein the processor circuitry is to:
--

310. The combination of Nielsen-372, Steinberg, and Tian teaches the audience measurement system of claim 1. *Supra* §VI.C.2. As explained below, the combination of Nielsen-372, Steinberg, and Tian further teaches that the audience measurement system performs the further operations recited in claim 4. *Infra* §VI.C.3[4a-4b]. Furthermore, a POSITA would have found it obvious to implement the audience measurement system such that the processor circuitry executes the machine readable instructions to perform operations directed to

identifying a television program and audience members. *Supra* §VI.B.2[1a], §VI.C.2[1a]. It would have been obvious to use the processor circuitry to perform the operations of limitations [4a] and [4b].

[4a] reduce 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

311. The combination of Nielsen-372, Steinberg, and Tian teaches that the processor circuitry of the audience measurement system reduces 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. Tian teaches reducing the resolution of an image to obtain a reduced-resolution image. *Supra* §VI.A.3[4a]; Ex-1005, Abstract, 000002. As explained above, the combination of Nielsen-372, Steinberg, and Tian teaches obtaining a sequence of images of the media exposure environment, detecting a head in one or more of the images, and determining an orientation of the head. *Supra* §VI.C.2[1d-1e]. A POSITA would have been motivated to apply Tian's teaching to Nielsen-372 and Steinberg such that the audience measurement system reduces the resolution of an image and performs face detection or head pose estimation on the reduced-resolution image. *Supra* §VI.C.1 (explaining why it would have been obvious to analyze reduced-resolution images; that analysis is incorporated here).

312. A POSITA would have found it obvious to implement the approach of Tian further in light of Steinberg's teachings of processing images of different resolutions. Ex-1007, ¶0141 ("Firstly an image of the person ... is captured with the camera in flash mode ... Then a second image is captured in non-flash mode, or at a lower resolution or with faster exposure settings."), ¶0142 ("In one embodiment, it is convenient to use a preview image stream of a digital camera. ... [I]t may be captured at a lower resolution than the main image acquisition[.]"), ¶0196 ("Selective compression, or alternatively selective resolution, or both, of images may be suggested or automatically provided where one or more faces receive higher grade compression and/or alternatively higher local resolution than other portions of the image."). Therefore, the combination of Nielsen-372, Steinberg, and Tian teaches the television audience measurement system reducing 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.

[4b] determine the orientation of the head with respect to the camera based on the reduced-resolution image.

313. The combination of Nielsen-372, Steinberg, and Tian teaches that the processor circuitry of the audience measurement system **determines the orientation of the head with respect to the camera** (e.g., pose estimate) **based on the reduced-resolution image**. As explained above, both Steinberg and Tian

teaches determining an orientation of the head with respect to the camera. *Supra* §VI.B.2[1e], §VI.C.2[1e]. Tian teaches determining a pose estimate based on a reduced-resolution image while maintaining a high level of accuracy. Ex-1005, 000005; *supra* §VI.A.3[4b]. A POSITA would have found it obvious to apply Tian's teachings of determining the orientation of the head with respect to the camera based on the reduced-resolution image to the audience measurement system taught by Nielsen-372 and Steinberg. *Supra* §VI.C.1 (explaining why it would have been obvious to analyze reduced-resolution images; that analysis is incorporated here). A POSITA would have been motivated to do so because performing the head pose estimation on the reduced-resolution image would have saved processing resources and time without substantially reducing the accuracy of the determination. Therefore, the combination of Nielsen-372, Steinberg, and Tian teaches the audience measurement system determining the orientation of the head with respect to the camera based on the reduced-resolution image.

4. Claim 5

314. It is my opinion that a POSITA would have found claim 5 obvious based on the combined teachings of Nielsen-372, Steinberg, and Tian.

5. The audience measurement system of claim 4, wherein the processor circuitry is to determine the audience identification information by: (i) generating a facial signature from a region of a second image of the one or more of the images corresponding to a location of the head in the reduced-resolution image, and (ii) comparing the generated facial signature to a database of facial signatures.

315. The combination of Nielsen-372, Steinberg, and Tian teaches the audience measurement system of claim 4. *Supra* §VI.C.3. As explained below, the combination of Nielsen-372, Steinberg, and Tian further teaches that the audience measurement system performs the further operations recited in claim 5. Furthermore, a POSITA would have found it obvious to implement the audience measurement system such that the processor circuitry executes the machine readable instructions to perform operations directed to identifying a television program and audience members. *Supra* §VI.B.2[1a], §VI.C.2[1a]. It would have been obvious to use the processor circuitry to perform the operations of claim 5.

316. Steinberg teaches **generating a facial signature** (e.g., faceprint) **from a region of an image corresponding to a location of the head** (e.g., face candidate region). *Supra* §VI.B.2[1f]. Steinberg teaches generating a faceprint for a face candidate region. Ex-1007, ¶¶0092, ¶¶0096, ¶¶0126-0127, ¶¶0130, FIGs. 1(a)-(c); *supra* §VI.B.2[1f]. The faceprint is a “facial signature,” as the term is used in the ’243 patent. The ’243 patent states that a facial signature may be a vector that is represented by a pattern or map of a region that corresponds to a face in an

image. Ex-1001, 6:3-44. Steinberg explains that a faceprint is a feature vector corresponding to parameters associated with a face region. Ex-1007, ¶0077 (“An ‘archived faceprint’ is a set of face classifier parameters associated with a particular face [sic] region ultimately extracted from a parent image and preferably normalized, and stored in the main recognition database, preferably along with links to the parent image and the face region.”), ¶0124 (“The faceprint may be stored for each face region. The faceprints that are what may be advantageously used in accordance with a preferred embodiment to achieve actual recognition of people or objects or other identities.”), ¶0126 (“The role of the Face Recognition module [1140] is to generate a feature vector, or Faceprint, for a face candidate region identified in an image.”). A POSITA would have understood that the facial signature in the ’243 patent and the faceprint in Steinberg are both vectors that represent patterns or features associated with a region corresponding to a face.

317. Steinberg teaches that the face candidate region is a region of an image corresponding to a location of the face. Ex-1007, ¶0118 (“Once at least one new image is detected it is preferably then analyzed by the Face Detection module [1130]. This module employs a means to determine if facial candidate regions are present in the new image.”), ¶0120 (“These are then passed to the Face Normalization module [1190] which analyzes the lighting, pose, facial features and/or orientation of each face candidate region.”). A POSITA would have

understood that this is the same location as the head. Therefore, Steinberg teaches generating a facial signature from a region of an image corresponding to a location of the head.

318. The combination of the teachings of Steinberg and Tian teaches generating the facial signature from a region of **a second image of the one or more of the images** corresponding to a location of the head **in the reduced-resolution image**. Tian teaches generating a reduced-resolution image by reducing the resolution of a first image of the one or more of the images. *Supra* §VI.C.3[4a]. A POSITA would have been motivated to perform face detection (e.g., determining the location of the face or head) on a reduced resolution image. *Supra* §VI.C.1 (explaining why it would have been obvious to analyze reduced-resolution images; that analysis is incorporated here). As Tian explains, face detection can be performed on a reduced resolution image with high accuracy. Ex-1005, 000005. Therefore, a POSITA would have been motivated to detect a face/head location on a reduced-resolution image to save processing power before performing face recognition to identify a person.

319. Furthermore, Tian teaches determining a face/head location in a first image and only tracking changes to face features in subsequent images. Ex-1005, 000001, 000003; *supra* §VI.A.4. Under this approach, if no change is detected, the face is determined to be at the same location in subsequent images.

320. A POSITA would have found it obvious to apply Tian's teaching to Nielsen-372 and Steinberg such that when the audience measurement system processes a sequence of images, it determines a face or head location in a first image and use the determined location for subsequent images. *Supra* §VI.C.1 (explaining why it is obvious to determine a face or head location in a first image and use the location for subsequent images; that analysis is incorporated here). Further in light of Tian's teachings of analyzing reduced-resolution images, a POSITA would have found it obvious to determine the location of a person's face in a reduced-resolution version of a first image and use this location to perform facial recognition based on the sequence of higher resolution images. In this manner, after a location of the head is determined in a reduced-resolution image generated from a first image, facial recognition is performed on a region corresponding to the location in one or more images of the sequence, including a second image.

321. Steinberg teaches **comparing the generated facial signature** (e.g., faceprint) **to a database of facial signatures**. *Supra* §VI.C.2[1f]; Ex-1007, ¶0092, ¶0096, ¶¶0126-0127, ¶0130, FIGs. 1(a)-(c). Steinberg further teaches **determining the audience identification information** by generating the facial signature and comparing the generated facial signature to a database of facial signatures. *Supra* §VI.C.2[1f]; Ex-1007, ¶0099, 0135, 0156-0157, 0178, FIGs. 4(a)-(b), FIG. 9(a).

Therefore, the combination of Nielsen-372, Steinberg, and Tian teaches the limitations of claim 5.

5. Claim 6

322. It is my opinion that a POSITA would have found claim 6 obvious based on the combined teachings of Nielsen-372, Steinberg, and Tian.

6[p]. The audience measurement system of claim 4, wherein the processor circuitry is to:

323. The combination of Nielsen-372, Steinberg, and Tian teaches the audience measurement system of claim 4. *Supra* §VI.C.3. As explained below, the combination of Nielsen-372, Steinberg, and Tian further teaches that the audience measurement system performs the further operations recited in claim 6. Furthermore, a POSITA would have found it obvious to implement the audience measurement system such that the processor circuitry executes the machine readable instructions to perform operations directed to identifying a television program and audience members. *Supra* §VI.C.2[1a]. It would have been obvious to use the processor circuitry to perform the operations of limitations [6a]-[6c].

[6a] identify a region corresponding to the head within the first image from which the reduced-resolution image was obtained;
--

324. The combined disclosure of Steinberg and Tian teaches **identifying a region corresponding to the head within an image**. *Supra* §VI.B.2[1d], §VI.C.2[1d]. Steinberg teaches analyzing images to detect a “face region,” which

is a region corresponding to the face and head. *Supra* §VI.B.2[1d]; Ex-1007, ¶0005 (“Face detection has to do with the problem of locating regions within a digital image or video sequence which have a high probability of representing a human face.”), ¶0118 (“Once at least one new image is detected it is preferably then analyzed by the Face Detection module [1130]. This module employs a means to determine if facial candidate regions are present in the new image.”); *see also* Abstract, ¶0057, , ¶0128, ¶0130, ¶0151, FIG.3. Furthermore, Tian teaches determining a face/head location in a first image and use this location for subsequent images unless a change is detected. *Supra* §VI.A.4, §VI.C.4; Ex-1005, 000001, 000003.

325. Tian further teaches **obtaining a reduced-resolution image from the first image**. *Supra* ¶VI.A.3[4a], ¶VI.C.3[4a]. Therefore, the combined disclosure of Nielsen-372, Steinberg, and Tian teaches identifying a region corresponding to the head within the first image from which the reduced-resolution image was obtained.

[6b] analyze a portion corresponding to the identified region of a second image; and based on the analysis of the region of the second image, determine that the head matches the known person.

326. The combined disclosure of Steinberg and Tian teaches **analyze a portion corresponding to the identified region of a second image; and based**

on the analysis of the region of the second image, determine that the head matches the known person.

327. As explained with respect to claim 5, a POSITA would have found it obvious to apply Tian’s teaching to Nielsen-372 and Steinberg such that when the television audience measurement system processes a sequence of images, it determines a face or head location in a first image and use the determined location for subsequent images. *Supra* §VI.C.4.

328. Steinberg teaches analyzing a portion corresponding to the identified region of the head or face to determine that the head matches the known person based on the analysis (e.g., by generating a faceprint based on the portion and determining that the generated faceprint matches a stored faceprint of the known person). *Supra* §VI.B.2[1f], §VI.C.2[1f], §VI.C.4; Ex-1007, ¶0126 (“The role of the Face Recognition module [1140] is to generate a feature vector, or Faceprint, for a face candidate region identified in an image.”), ¶0092 (“Then, the images and face prints, ... are made available for the face recognition module to extract, enhance and scan the facial regions and compares their faceprints with a database of ‘known faceprints.’”); *see also* ¶0096, 0099, ¶0127, ¶0130, ¶0135, ¶¶0156-0157, ¶0178, FIGs. 1(a)-(c), FIGs. 4(a)-(b).

329. Therefore, the combined disclosure of Nielsen-372, Steinberg, and Tian teaches analyzing a portion corresponding to the identified region of a second

image and based on the analysis of the region of the second image, determine that the head matches the known person.

6. Claim 8

330. It is my opinion that a POSITA would have found claim 8 obvious based on the combined teachings of Nielsen-372, Steinberg, and Tian.

8. The audience measurement system of claim 6, wherein the audience identification information includes an identifier associated with the known person and the processor circuitry is to, responsive to a determination that the head matches the known person, cause the identifier associated with the person to be stored in the memory.

331. The combination of Nielsen-372, Steinberg, and Tian teaches the audience measurement system of claim 6. *Supra* §VI.C.5. As explained below, the combination of Nielsen-372, Steinberg, and Tian further teaches that the audience measurement system performs the further operations recited in claim 8. Furthermore, a POSITA would have found it obvious to implement the audience measurement system such that the processor circuitry executes the machine readable instructions to perform operations directed to identifying a television program and audience members. *Supra* §VI.B.2[1a], §VI.C.2[1a]. It would have been obvious to use the processor circuitry to perform the operations of claim 8.

332. The combination of Nielsen-372, Steinberg, and Tian teaches determining audience identification information based on a match of the head to a known person. *Supra* §VI.B.2[1f], §VI.C.2[1f]. Steinberg teaches the local

metering unit storing viewing records before sending the records to a central facility. Ex-1006, ¶0036 (“[T]he single home unit may perform the functions of storing data and forwarding the stored data to a central facility for subsequent processing.”), ¶0041 (“The metering unit 124 typically collects a set of viewing records and transmits the collected viewing records over a connection 140 to a central office or data processing facility (not shown) for further processing or analysis.”). The viewing records include information regarding the audience. Ex-1006, ¶0048 (teaching “generat[ing] one or more viewing records based on a change in the viewing audience”). Steinberg further teaches storing identification information in association with faceprints. *See, e.g.*, Ex-1007, ¶0099 (“[E]ach faceprint in the database module is preferably associated with a single identity, while each identity is associated with one or more face classes containing one or more faceprints.”), ¶0150 (“The Faceprints are a set of co-ordinates in the known Face Space. The data stored in this set of database tables is directly associated with known Identity Regions, Face Class Regions and Faceprint Vectors in the Face Space[.]”) The combination of Nielsen-372, Steinberg, and Tian teaches determining audience identification information and outputting a signal indicative of the audience identification information. *Supra* §VI.B.2[1f-1g], §VI.C.2[1f-1g]. In light these teachings, the combination of Nielsen-372, Steinberg, and Tian teaches storing audience identification information.

333. The combination of Nielsen-372, Steinberg, and Tian teaches audience identification information including, for example, identity region, identity grouping, face class, and identity of person. *Supra* §VI.B.2[1f], §VI.C.2[1f]; Ex-1007, ¶¶0099, ¶¶0135, ¶¶0156-0157, ¶¶0178. Because such information corresponds to and identifies known persons, it includes identifiers associated with the persons. Furthermore, the identification information are linked to database profiles of known persons. Ex-1007, ¶¶0159 (“Once a match has been determined and the current Faceprint has been added to the Face Recognition Data for the “known person” who was matched, the system next creates a link from the Image Data to the database profile of that “known person” [4270]. A cross-link is also added to the profile of the matched “known person” which associates the image with that person [4280].”). A POSITA would have found it obvious for the audience identification information to include an identifier associated with the known person given the broad meaning of the word “identifier.” A POSITA would have further found it obvious that the profile of a known person identifies the person and a link to the profile is an identifier of person. In fact, the ’243 patent admits that the concept of an identifier was known in the prior art. Ex-1001, 1:43-55 (“Audience measurement of media content ... often involves collection of ... people data (e.g., **identifiers**, demographic data associated with audience members, etc.).” (emphasis added)).

334. The combination of Nielsen-372, Steinberg, and Tian teaches that the audience measurement system includes a memory. *Supra* §VI.B.2[1a], §VI.C.2[1a]. A POSITA would have understood that memory is a commonly used data storage device and found it obvious to cause the identifier associated with the person to be stored in memory. Therefore, the combination of Nielsen-372, Steinberg, and Tian teaches that the audience identification information includes an identifier associated with the known person and, responsive to a determination that the head matches the known person, causing the identifier associated with the person to be stored in the memory.

7. Claim 9

335. It is my opinion that a POSITA would have found claim 9 obvious based on the combined teachings of Nielsen-372, Steinberg, and Tian.

9[p]. An audience measurement system, comprising:
--

336. As discussed for Ground 2, to the extent the preamble limits claim 9, the combination of Nielsen-372 and Steinberg teaches the preamble. *Supra* §VI.B.3.9[p].

337. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders the preamble obvious.

[9a] an audience measurement device at a media exposure environment, the audience measurement device to:

338. As discussed for Ground 2, the combination of Nielsen-372 and Steinberg teaches limitation [9a]. *Supra* §VI.B.3[9a].

339. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[9b] generate an audio signature of media content presented by a television within the media exposure environment;

340. As discussed for Ground 2, Nielsen-372 teaches limitation [9b]. *Supra* §VI.B.3[9b].

341. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[9c] obtain content identifying data corresponding to the presented media content, the content identifying data based on the audio signature;

342. As discussed for Ground 2, Nielsen-372 teaches limitation [9c]. *Supra* §VI.B.3[9c].

343. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[9d] while the media content corresponding to the content identifying data is presented by the television, collect a first image of the media exposure environment with a camera;

344. As discussed for Ground 2, the combination of Nielsen-372 and Steinberg teaches limitation [9d]. *Supra* §VI.B.3[9d].

345. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[9e] attempt to detect a head in the first image;

346. As discussed for Ground 2, the combination of Nielsen-372 and Steinberg teaches limitation [9e]. *Supra* §VI.B.3[9e]. Tian also teaches attempting to detect a head appearing in an image using face detection techniques. Ex-1005, 000002 (“Many face detection methods have been developed to detect faces in an arbitrary scene....”), 000005 (“All images in 704 sequences from 97 subjects are used for face detection and head pose estimation. The results show that both face detectors [23, 29] can detect all faces for head region in resolution of 288x384, 144x192, 72x96, and 36x48, but failed for resolution of 18x24.”).

347. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[9f] determine an orientation of the head with respect to the camera;

348. As discussed for Ground 2, the combination of Nielsen-372 and Steinberg teaches limitation [9f]. *Supra* §VI.B.3[9f]. Tian further teaches determining an orientation of the head with respect to the camera. *Supra* §VI.A.2[1e]; Ex-1005, 000001-03.

349. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[9g] determine audience identification information based on an indication that the head matches a known person associated with the audience measurement system; and

350. As discussed for Ground 2, the combination of Nielsen-372 and Steinberg teaches limitation [9g]. *Supra* §VI.B.3[9g].

351. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[9h] a data collection facility to obtain the content identifying data and the audience identification information from the audience measurement device.

352. As discussed for Ground 2, the combination of Nielsen-372 and Steinberg teaches limitation [9h]. *Supra* §VI.B.3[9h].

353. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

8. Claim 11

354. It is my opinion that a POSITA would have found claim 11 obvious based on the combined teachings of Nielsen-372, Steinberg, and Tian.

11[p]. The audience measurement system of claim 9, wherein the audience measurement device is to:

355. The combination of Nielsen-372, Steinberg, and Tian teaches the audience measurement system of claim 9. *Supra* §VI.C.7. As explained below, the combination of Nielsen-372, Steinberg, and Tian further teaches that the audience measurement device performs the further operations recited in claim 11. *Infra* §VI.C.8[11a-11b].

[11a] reduce a resolution of the first image of the media exposure environment to obtain a reduced-resolution image, and

356. As explained for claim 4, the combination of Nielsen-372, Steinberg, and Tian teaches that the processor circuitry of the audience measurement system reduces a resolution of one of the one or more of the images of the media exposure environment to obtain a reduced-resolution image. *Supra* §VI.C.3[4a]. The “first image” in claim 9 is satisfied by one of the one or more images. *Supra* §VI.B.3[9d], ¶VI.C.7[9d]. The audience measurement device includes the processor circuitry. *Supra* §VI.B.3[9a], §VI.C.7[9a]. Therefore, the combination of Nielsen-372, Steinberg, and Tian teaches that the audience measurement device

reduces a resolution of the first image of the media exposure environment to obtain a reduced-resolution image.

[11b] determine the orientation of the head with respect to the camera based on the reduced-resolution image.
--

357. As explained for claim 4, the combination of Nielsen-372, Steinberg, and Tian teaches that the processor circuitry of the audience measurement system determines the orientation of the head with respect to the camera based on the reduced-resolution image. *Supra* §VI.C.3[4b]. The audience measurement device includes the processor circuitry. *Supra* §VI.B.3[9a], §VI.C.7[9a]. Therefore, the combination of Nielsen-372, Steinberg, and Tian teaches that the audience measurement device determines the orientation of the head with respect to the camera based on the reduced-resolution image.

9. Claim 12

358. It is my opinion that a POSITA would have found claim 12 obvious based on the combined teachings of Nielsen-372, Steinberg, and Tian.

12[p]. The audience measurement system of claim 11, wherein the audience measurement device is to:

359. The combination of Nielsen-372, Steinberg, and Tian teaches the audience measurement system of claim 11. *Supra* §VI.C.8. As explained below, the combination of Nielsen-372, Steinberg, and Tian further teaches that the

audience measurement device performs the further operations recited in claim 12.

Infra §VI.C.9[12a-12b].

[12a] identify a region corresponding to the head within the first image from which the reduced-resolution image was obtained; and

360. As explained for claim 6, the combination of Nielsen-372, Steinberg, and Tian teaches that the processor circuitry of the audience measurement system identifies a region corresponding to the head within the first image from which the reduced-resolution image was obtained. *Supra* §VI.C.5[6a]. The audience measurement device includes the processor circuitry. *Supra* §VI.B.3[9a], §VI.C.7[9a]. Therefore, the combination of Nielsen-372, Steinberg, and Tian teaches that the audience measurement device identifies a region corresponding to the head within the first image from which the reduced-resolution image was obtained.

[12b] analyze a portion of a second image corresponding to the identified region to determine whether the head matches the known person.

361. As explained for claim 6, the combination of Nielsen-372, Steinberg, and Tian teaches that the processor circuitry of the audience measurement system analyzes a portion of a second image corresponding to the identified region, and based on the analysis of the region of the second image, determine that the head matches the known person. *Supra* §VI.C.5[6b]. Steinberg teaches determining **whether** the head matches the known person. *See, e.g.*, Ex-1007, ¶0126 (“If the

Faceprint can be clearly associated with a single identity region then the face is considered to be positively identified and the Image Classification Database [1160] is updated accordingly. If there is no clear association with a known identity region the user is prompted to manually provide recognition data on the face candidate region and may be presented with a list of “closest” choices.”). The audience measurement device includes the processor circuitry. *Supra* §VI.B.3[9a], §VI.C.7[9a]. Therefore, the combination of Nielsen-372, Steinberg, and Tian teaches that the audience measurement device analyzes a portion of a second image corresponding to the identified region to determine whether the head matches the known person.

10. Claim 13

362. It is my opinion that a POSITA would have found claim 13 obvious based on the combined teachings of Nielsen-372, Steinberg, and Tian.

13. The audience measurement system of claim 12, wherein the audience identification information includes an identifier associated with the known person and the audience measurement device is to, responsive to the indication that the head matches the known person, cause recordation of the identifier associated with the known person in a memory.

363. The combination of Nielsen-372, Steinberg, and Tian teaches the audience measurement system of claim 12. *Supra* §VI.C.9. As explained for claim 8, the combination of Nielsen-372 and Steinberg teaches that the audience identification information includes an identifier associated with the known person.

Supra §VI.C.6. The combination of Nielsen-372, Steinberg, and Tian also teaches that the processor circuitry of the audience measurement system causes the identifier associated with the person to be stored in the memory responsive to a determination that the head matches the known person. *Id.*. Because the audience measurement device includes the processor circuitry, the combination teaches that the audience measurement device performs this operation. *Supra* §VI.B.3[9a], §VI.C.7[9a]. The combination of Nielsen-372, Steinberg, and Tian further teaches representing the match with an indication that the head matches the known person. *Supra* §VI.B.3[9g], §VI.C.7[9g]. Furthermore, a POSITA would have understood that “storing” and “recording” are interchangeable terms to describe saving of data in memory. Therefore, the combination of Nielsen-372, Steinberg, and Tian teaches that the audience measurement device causes **recording** of the identifier associated with the known person in a memory responsive to the **indication** that the head matches the known person.

11. Claim 14

364. It is my opinion that a POSITA would have found claim 14 obvious based on the combined teachings of Nielsen-372, Steinberg, and Tian.

14. The audience measurement system of claim 11, wherein the audience measurement device is to determine the audience identification information by: (i) generating a facial signature from a region of a second image corresponding to a location of the head in the reduced-resolution image, and (ii) comparing the generated facial signature to a database of facial signatures.

365. The combination of Nielsen-372, Steinberg, and Tian teaches the audience measurement system of claim 11. *Supra* §VI.C.8. As explained for claim 5, the combination of Nielsen-372, Steinberg, and Tian teaches that the processor circuitry of the audience measurement system determines the audience identification information by: (i) generating a facial signature from a region of a second image of the one or more of the images corresponding to a location of the head in the reduced-resolution image, and (ii) comparing the generated facial signature to a database of facial signatures. *Supra* §VI.C.4. The audience measurement device includes the processor circuitry. *Supra* §VI.B.3[9a], §VI.C.7[9a]. Therefore, the combination of Nielsen-372, Steinberg, and Tian teaches that the audience measurement device determines the audience identification information by: (i) generating a facial signature from a region of a second image corresponding to a location of the head in the reduced-resolution image, and (ii) comparing the generated facial signature to a database of facial signatures.

12. Claim 16

366. It is my opinion that a POSITA would have found claim 16 obvious based on the combined teachings of Nielsen-372, Steinberg, and Tian.

16[p]. A method for obtaining exposure data for a media exposure environment, the method comprising:

367. As discussed for Ground 2, to the extent the preamble limits claim 16, the combination of Nielsen-372 and Steinberg teaches the preamble. *Supra* §VI.B.4.16[p].

368. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders the preamble obvious.

[16a] generating, using a program detector, an audio signature of media content presented by a television within the media exposure environment;

369. As discussed for Ground 2, Nielsen-372 teaches limitation [16a]. *Supra* §VI.B.4[16a].

370. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[16b] obtaining, based on the audio signature, content identifying data corresponding to the presented media content;

371. As discussed for Ground 2, Nielsen-372 teaches limitation [16b]. *Supra* §VI.B.4[16b].

372. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[16c] while the media content corresponding to the content identifying data is presented by the television, capturing first and second images of the media exposure environment with a camera;
--

373. As discussed for Ground 2, the combination of Nielsen-372 and Steinberg teaches limitation [16c]. *Supra* §VI.B.4[16c]. Tian further teaches obtaining a sequence of images of human subjects using a camera. Ex-1005, 000005; *supra* §VI.A.2[1d]. A POSITA would have found it obvious that the sequence of images include first and second images, in which a head is detected.

374. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[16d] analyzing, using an audience detector, the first image to attempt to detect a head in the first image;
--

375. As discussed for Ground 2, the combination of Nielsen-372 and Steinberg teaches limitation [16d]. *Supra* §VI.B.4[16d]. Tian also teaches analyzing an image to attempt to detect a head appearing in an image using face detection techniques. Ex-1005, 000002, 000005.

376. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[16e] determining, using the audience detector, an orientation of the head with respect to the camera;

377. As discussed for Ground 2, the combination of Nielsen-372 and Steinberg teaches limitation [16e]. *Supra* §VI.B.4[16e]. Tian further teaches determining an orientation of the head with respect to the camera. *Supra* §VI.A.2[1e]; Ex-1005, 000001-03.

378. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[16f] determining, using the audience detector, audience identification information based on an indication that the head matches a known person associated with the audience detector; and

379. As discussed for Ground 2, the combination of Nielsen-372 and Steinberg teaches limitation [16f]. *Supra* §VI.B.4[16f].

380. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

[16g] providing the content identifying data and the audience identification information to a data collection facility.

381. As discussed for Ground 2, the combination of Nielsen-372 and Steinberg teaches limitation [16g]. *Supra* §VI.B.4[16g].

382. For the reasons explained above, the combination of Nielsen-372, Steinberg, and Tian renders this limitation obvious.

13. Claim 18

383. It is my opinion that a POSITA would have found claim 18 obvious based on the combined teachings of Nielsen-372, Steinberg, and Tian.

18[p]. The method of claim 16, further comprising:

384. The combination of Nielsen-372, Steinberg, and Tian teaches the method of claim 16. *Supra* §VI.C.12. As explained below, the combination of Nielsen-372, Steinberg, and Tian further teaches that the method comprises the further operations recited in claim 18. *Infra* §VI.C.13[18a-18b].

[18a] reducing a resolution of the first image of the media exposure environment to obtain a reduced-resolution image, and

385. As explained for claim 11, the combination of Nielsen-372, Steinberg, and Tian teaches that the audience measurement system reduces a resolution of the first image of the media exposure environment to obtain a reduced-resolution image. *Supra* §VI.C.8[11a]. Therefore, the combination of Nielsen-372, Steinberg, and Tian teaches reducing a resolution of the first image of the media exposure environment to obtain a reduced-resolution image.

[18b] wherein the determining of the orientation of the head includes determining the orientation of the head with respect to the camera using the reduced-resolution image.

386. The combination of Nielsen-372, Steinberg, and Tian teaches determining an orientation of the head with respect to the camera. *Supra*

§VI.C.12[16e]. As explained for claim 11, the combination of Nielsen-372, Steinberg, and Tian teaches carrying out the determination based on the reduced-resolution image. *Supra* §VI.C.8[11b]. Therefore, the combination of Nielsen-372, Steinberg, and Tian teaches that the determining of the orientation of the head includes determining the orientation of the head with respect to the camera using the reduced-resolution image.

14. Claim 19

387. It is my opinion that a POSITA would have found claim 19 obvious based on the combined teachings of Nielsen-372, Steinberg, and Tian.

19[p]. The method of claim 18, further comprising:

388. The combination of Nielsen-372, Steinberg, and Tian teaches the method of claim 18. *Supra* §VI.C.13. As explained below, the combination of Nielsen-372, Steinberg, and Tian further teaches that the method comprises the further operations recited in claim 19. *Infra* §VI.C.14[19a-19b].

[19a] identifying a region corresponding to the head within the first image from which the reduced-resolution image was obtained;
--

389. As explained for claim 12, the combination of Nielsen-372, Steinberg, and Tian teaches that the audience measurement system identifies a region corresponding to the head within the first image from which the reduced-resolution image was obtained. *Supra* §VI.C.9[12a]. Therefore, the combination of Nielsen-

372, Steinberg, and Tian teaches identifying a region corresponding to the head within the first image from which the reduced-resolution image was obtained.

[19b] analyzing a portion of the second image corresponding to the identified region; and based on the analyzing of the region of the second image, determining whether the head matches the known person.

390. As explained for claim 12, the combination of Nielsen-372, Steinberg, and Tian teaches that the audience measurement system analyzes a portion of a second image corresponding to the identified region to determine whether the head matches the known person. *Supra* §VI.C.9[12b]. Therefore, the combination of Nielsen-372, Steinberg, and Tian teaches analyzing a portion of the second image corresponding to the identified region; and based on the analyzing of the region of the second image, determining whether the head matches the known person.

15. Claim 20

391. It is my opinion that a POSITA would have found claim 20 obvious based on the combined teachings of Nielsen-372, Steinberg, and Tian.

20. The method of claim 19, wherein the audience identification information includes an identifier associated with the known person, and further including recording the identifier associated with the known person in a memory.

392. The combination of Nielsen-372, Steinberg, and Tian teaches the method of claim 19. *Supra* §VI.C.14. As explained for claim 13, the combination of Nielsen-372, Steinberg, and Tian teaches that the audience identification

information includes an identifier associated with the known person. *Supra* §VI.C.10. Furthermore, the combination of Nielsen-372, Steinberg, and Tian teaches that the audience measurement system causes recordation of the identifier associated with the known person in a memory. *Id.* Therefore, the combination of Nielsen-372, Steinberg, and Tian teaches that the audience identification information includes an identifier associated with the known person and that the method further includes recording the identifier associated with the known person in a memory.

VII. BACKGROUND AND QUALIFICATIONS

393. This section contains a summary of my educational background, career history, publications, and other relevant qualifications. My full curriculum vitae is attached as Appendix 1 to this declaration.

394. I am currently a Professor of Empire Innovation in the Department of Computer Science and Engineering at the University at Buffalo. I am also the inaugural director of the University at Buffalo Institute for Artificial Intelligence and Data Science. I received my B.Sc. degree in Computer Science and Mathematics from the Bloomsburg University of Pennsylvania in 1987, and I received my M.Sc. degree in computer science from the University of Maryland College Park, Maryland, in 1989. I continued my studies in the Computer Vision Laboratory at the University of Maryland, where I earned a Ph.D. degree in

Computer Science in 1993. My Ph.D. thesis focused on applying image processing and computer vision techniques to the detailed understanding of document images.

395. In April 1993, I began working as a research scientist at the University of Maryland's Institute for Advanced Computer Studies (UMIACS). In the same year, I became the Director of the Laboratory for Language and Media Processing (LAMP) at UMIACS. I was an assistant research scientist from 1993 to 2000 and an associate research scientist from 2000 to 2008 before becoming a senior research scientist in 2009.

396. While on the University of Maryland faculty, I was the principal or co-principal investigator for numerous research projects. I received millions of dollars of funding from government agencies, private companies, and academic institutions. My research's focus was widely related to computer vision, pattern recognition, and human language technologies. In document image analysis, my projects have included page decomposition, structural analysis and classification page segmentation, logo recognition, document image compression, duplicate detection, image-based retrieval, optical character recognition ("OCR"), and signature verification. In video processing, my projects centered on object recognition, pose estimation, the segmentation of compressed videos, structural representation classification of videos, detection of reformatted video segments, and the performance evaluation of automated video analysis algorithms. I have

focused extensively on vision algorithms and techniques for operating on embedded mobile devices, such as camera phones, cameras, and edge computing devices in both areas.

397. In 2014, I was hired as a Program Manager in the Information Innovation Office at the Defense Advanced Research Projects Agency (DARPA). In my role, I developed, selected, funded, and oversaw research and transition funding in computer vision, human language technologies, and voice analytics. I developed and ran the successful Media Forensics program, which sought to detect, characterize, and uncover the provenance of modified images and videos, including documents and media collected with mobile devices.

398. Over the past decade, my research has expanded to include the general field of artificial intelligence, applied primarily to computer vision. I have published extensively on image quality estimation, text detection in the wild, machine learning applications for front-end and edge computing, and cutting-edge applications in evolutionary computing.

399. In 2005, I co-founded and served as the President of Applied Media Analysis Inc., a company specializing in image processing and computer vision technologies on mobile devices and in multilingual/multimodal data annotation necessary for training systems to operate on these devices. I have also served as a technical consultant in document image understanding, document image retrieval,

image processing, and system integration for government agencies and private companies. Our company was one of the early collaborators with Nokia in using their low-resolution camera phones for capturing and analyzing document images.

400. As a member of the graduate faculty at the University of Maryland and the University of Buffalo, I have co-supervised approximately 20 Ph.D. theses and mentored over 40 Master's degree students and countless undergraduate students on research topics related to my fields of study. I have taught computer vision, document analysis, artificial intelligence, and edge computing courses. I have also developed and taught a short course for the Society of Photo-Optical Instrumentation Engineers (SPIE) on document image analysis and computer vision on Android devices.

401. I am a founding editor of the International Journal of Document Analysis and Recognition (IJ DAR). I have served as the general chair or co-chair at several international conferences. I was the general chair of the field's premier conference, the International Conference on Document Analysis and Recognition, held in Washington, DC, in 2013. I will be the conference's general chair when it is held in San Jose in 2023.

402. I have over 62 refereed journal publications and over 225 refereed conference papers. In 2002, I received an honorary Doctor of Technology Sciences from the University of Oulu in Finland for my contributions to digital

media processing and document image analysis research. I am a Fellow of the Institute of Electrical and Electronics Engineers (IEEE) and the International Association of Pattern Recognition (IAPR).

A. Compensation

403. For my efforts in connection with the preparation of this declaration I have been compensated at my standard rate for this type of consulting activity. My compensation is in no way contingent on the results of these or any other proceedings relating to the above-captioned patent.

B. Materials and Other Information Considered

404. I have considered information from various sources in forming my opinions. I have reviewed and considered each of the exhibits listed in the attached Appendix 2 (Materials Considered in the Preparation of This Declaration) in forming my opinions.

VIII. UNDERSTANDING OF THE LAW

405. I am not an attorney. In forming my opinions in this Declaration, I applied the relevant legal principles provided to me by counsel, which are summarized in Appendix 4.

IX. RESERVATION OF RIGHTS

406. My opinions are based upon the information that I have considered to date. I am unaware of any evidence of secondary considerations with respect to the '243 Patent that would render any of the challenged claims non-obvious. I

reserve the right, however, to supplement my opinions in the future to respond to any arguments raised by the owner of the '243 Patent and to take into account new information that becomes available to me.

407. 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 of Title 18 of the United States Code.

Executed on June 16th, 2023.

By:


DAVID SCOTT DOERMANN

APPENDIX 1: CURRICULUM VITAE OF DAVID SCOTT DOERMANN

David Scott Doermann

Department of Computer Science and Engineering
School of Engineering and Applied Sciences
Institute for Artificial Intelligence and Data Science
University at Buffalo, Buffalo, NY
doermann@buffalo.edu cse.buffalo.edu/~doermann
716-645-1557 (o) 410-493-9043 (c)

Research Synopsis

My research interests include the analysis and processing of electronic documents and document images and the analysis, indexing, and retrieval of information from a range of sources, including images, video, and audio. Recent interests include the applications of deep learning, including generative adversarial networks to document and video analytics. I have contributed extensively to the document analysis research community on problems including page and zone segmentation, labeling, and classification, forms processing, signature detection and verification, handwriting modeling, quality estimation, enhancement, image-based indexing and retrieval, and mobile device applications. My research related to images and video includes image quality estimation, detection and recognition of scene text in images and video, the segmentation and retrieval of compressed sequences of video based on structure and motion, video sequence visualization, genre classification, dynamic event detection, performance evaluation, video-over-IP, wireless video, and pervasive networking and computing and media forensics. I have worked extensively with related technologies for audio analysis and speech recognition. My general interests involve applying language-understanding techniques and constraints for context-aware analysis of various forms of signal-based media. Recent applications of deep learning include generative adversarial networks to document and video analytics, Neural Architecture Search (NAS), Adversarial AI, Edge Computing, and Neural Network Optimization

Education

University of Maryland, College Park

Doctor of Philosophy in Computer Science

May 1993

Thesis: Document Image Understanding: Integrating Recovery and Interpretation

Advisor: Dr. Azriel Rosenfeld

Master of Science in Computer Science

December 1989

Bloomsburg University of Pennsylvania

Bachelor of Science in Computer Science

December 1986

Bachelor of Science in Mathematics

Professional Experience

University at Buffalo

Buffalo, New York

Professor, Empire Innovation in Artificial Intelligence and Machine Learning

August 2018-present

Department of Computer Science and Engineering

Teaching and Research focused on advancing university initiatives in Artificial Intelligence, Computer Vision, Pattern Recognition and Machine Learning.

Director, University at Buffalo Artificial Intelligence and Data Science Institute **September 2018-present**

Director, Engineering Science MS – AI Program

September 2019-present

Amber Video
San Francisco, California
Technical Advisor

August 2018 - Present

Defense Advanced Research Project Agency (DARPA)
Arlington, VA

Program Manager

April 2014 – April 2018

Developed, selected, and oversaw research and transition funding in the areas of computer vision, human language technologies, and voice analytics by coordinating performers on all of the projects, orchestrating consensus, evaluating cross-team management, and continuing to meet fluid program objectives.

University of Maryland College Park
College Park, Maryland

Research Scientist (Emeritus as of September 2018)

December 2008 – August 2018

Associate Research Scientist

April 2000-December 2008

Director, Language, and Media Processing Laboratory

April 1996 – August 2018

Adjunct Member of the Graduate Faculty

December 1994 – August 2018

Assistant Research Scientist

April 1993 - April 2000

Established the Laboratory for Language and Media Processing to serve the growing demand of the intelligence community by working with interdisciplinary teams of faculty and students from computer science, engineering, linguistics, and the iSchool. Worked extensively with undergraduate and graduate students in Computer Science and Engineering to complete research requirements for BS, MS and PHD Degrees. Worked with various government organizations in both unclassified and classified environments to develop transition technologies to operational environments

University of Salford, Manchester UK

Royal Academy of Engineering – Distinguished Visiting Fellow

2011-2013

Topic: Practical systems for mass-digitization using cloud computing.

Applied Media Analysis, Inc.

College Park, Maryland

President/Co-Founder

2001–2014

Focused the need for a cross-platform implementation of computer vision algorithms on mobile devices and developed the architecture to port basic image processing and document analysis capabilities to various devices from a wide range of manufacturers. Provided various levels of program support by providing annotated language data for training and evaluation.

Consulting (long term):

Consulted in the areas of Document Image Understanding, Document Retrieval, Image Processing, and System Integration.

National Security Agency Advisory Board (NSAAB), Ft. Meade, Maryland

Advanced Resource Technologies, Inc, Alexandria, Virginia

Presearch Inc, Fairfax, Virginia

Department of Atmospheric Chemistry, NASA Goddard Space Flight Center - Greenbelt, Maryland.

Legal Consulting:

Served as an expert on contracting disputes and as a document expert to in various matters related to document image analysis, optical character recognition technologies, pattern recognition, image processing and computer vision.

Expert Consultant – Engaged by Wiley Rein and Fielding, 2005 – Analysis of technology in Federal Contracting Award Dispute

Expert Consultant – Engaged by Perkins Coie, 2006 – Examination of Duplicate Document Images per court Order to remove near duplicate document images from discovery material.

Expert Witness – Wrote expert reports and provided deposition testimony on behalf of the Plaintiff in GTX Corp vs. Kofax Image Products et. al. Worked with Perkins Coie LLP, 2006-2008 on issues related to document cleaning and enhancement in patent 7,016,536

Expert Witness – Wrote expert Reports, provided deposition testimony, and provided trial testimony on behalf of the defendant in Nuance Communications vs. Abbyy Software. Worked with Finnegan Henderson Farabow Garrett & Dunner LLP, 2011-2013 on issues related to optical character recognition (OCR) technologies in the patents 5,381,489, 6,038,342, 5,261,009

Expert Witness – Retained by Latham & Watkins LLP to provide analysis, reports, and possible testimony on behalf of the defendants Capital One vs. Intellectual Ventures regarding US Patent 7,260,587

Expert Witness – Testifying invalidity expert for patents in Key Duplication trial of The Hillman Group, Inc. v. KeyMe, LLC, 2020-2021

Expert Witness – Currently retained for patent infringement litigation in the area of image retrieval

Expert Witness – Currently retained for patent infringement litigation in the area of mobile and remote document systems

External Leadership, Honors and Service

DARPA ISAT Study Group

2021-Present

Elected Member selected to identify new areas of development and to recommend future research directions

State of New York Governors Commission on Artificial Intelligence, Robotics and Automation representing the SUNY System

2019-2021

Bethania Kids

Board of Directors

2018-Present

Rochester Institute of Technology, Rochester, NY

Industry Advisory Board of the Department of Computer Science

2018-Present

Research Grants and Contracts

Principal Investigator/Co-Principal Investigator

Center for Identification Technology Research – Generation of High Resolution Morphed Faces for Evaluation

National Science Foundation, 7/21-6/23

Center for Identification Technology Research – Deep Fakes

National Science Foundation, 9/20-9/21

Multi-media Analytics Leading to Intent and Semantic Evidence (MALISE)

SRI/DARPA, 8/20-8/24

Center for Identification Technology Research – Fmonet

National Science Foundation, 3/20-3/21

SPIR: Applications of Machine Learning to HVAC Service Diagnostics and Performance Optimization
NorthPark Innovations Group, 02/19-02/20

Scalable Video Retrieval
National Science Foundation, 10/13-10/14, \$234,225

Document Image Quality Estimation, Enhancement, Classification and Retrieval
National Science Foundation, 10/13-10/14, \$234,225

EAGER: Large Scale Document Image Triage, Indexing and Retrieval,
National Science Foundation, 10/12-10/13, \$300,000

EAGER: Video Analytics for Large Heterogeneous Repositories
National Science Foundation, 10/12-10/13, \$214,195

ADALT – Automatic Detection of Anomalies in Lexical Text
Center for the Advanced Study of Language,, 9/10-3/13, 225,000

Tools for Document Analysis Metadata Extraction
Department of Defense, 1/08-12/10, \$700,000

MADCAT Arabic Document Image Analysis
DARPA/BBN, 12/07-12/13, \$1,650,000

Best of Breed OCR Evaluation
Army Research Laboratory, 11/07 – 10/08, \$100,000

Advanced CIR Algorithm for Forensic Analysis II,
MIPS, w/ Advanced Digital Forensic Solutions, Inc., 06/07-05/08, \$77,777

CLEAT: A Classification, Enhancement and Analysis Tool,
Singapore Centre for Strategic Infocom Tech., 12/06-12/07, \$200,000

Bridge Map & Gazetteer Optical Character Recognition,
Sarnoff Corp., 09/06-09/07, \$75,000

Integration of DocLib Components as Web Services,
Universities-Illinois Institute of Technology, 06/06-02/07, \$50,000

Handwriting Recognition Technology Assessment,
Central Intelligence Agency, 05/06-4/07, \$94,075

Advanced CIR Algorithm for Forensic Analysis,
MIPS, w/ Advanced Digital Forensic Solutions, Inc., 02/06-01/07, \$77,777

Integration of DocLib Components as Web Servicer,
Illinois Institute of Technology, 03/05-04/07, \$50,000

BRIDGE
Center for the Advanced Study of Language, 07/04-06/06, \$667,000

Evaluation and Improvement of MT Using Parallel Corpra,
Mitre Corporation, 05/01-01/03, \$100,000

Defect Detection and Classification
KLA-Tencor, 9/00-02/03, \$ 125,000

Automated Modeling and Analysis of Structured Documents,
Panasonic Information Technologies, 01/01-03/02. \$35,000

Intelligent Set-Top Processing,
Philips Research Laboratory, 06/00-09/01, \$55,000

Task Specific Evaluation of Digital Libraries
University of Maryland Baltimore County, 01/00-11/01, \$99,641

Document Image Decomposition and Support,
Department of Defense, 06/93-12/96, \$907,987

Investigator

Development and Evaluation of Search Technology for Discovery of Evidence in Civil Litigation
NSF, 6/11-5/13, \$1,116,000

Language and Media Processing Laboratory,
Department of Defense, 09/96-04/07, \$9,051,812

Event Modeling and Recognition for (Mutli-Perspective) Recognition,
Department of the Interior, 09/06-09/07, \$200,000

ARDA Workshop-VERAAE,
Northrop-Grumman, 06/05-06/06, \$70,000

Multilingual Access to Large Spoken Archives (MALACH),
National Science Foundation, 9/01-8/07, \$3,300,000 (Investigator)

Activity Detection by Video Content Evaluation (ADVICE),
Department of Defense, 08/00-12/02, \$898,380

Video Frame Ranking Toolkit
Hitachi Corporation, 9/01-8/02, \$50,000

Video Detection of Events,
Philips Laboratories, 10/98-09/99, \$50,000

Automated Video Analysis and Retrieval,
Philips Laboratory, 05/95-02/98, \$75,000

Ph.D. Thesis Supervised

Kamran Etemad	<i>Multi-Scale Discriminant Analysis and Recognition of Signals and Images</i> , 1996 (w/ Rama Chellappa)
Omid Kia	<i>Document Image Compression and Analysis</i> , 1997 (w/ Rama Chellappa)
Vikrant Kobla	<i>Automated Analysis of MPEG-Compressed Video</i> , 1999 (w/ Azriel Rosenfeld)
Huiping Li	<i>Automated Processing and Analysis of Text in Digital Video</i> , 2000 (w/ Rama Chellappa)
Katherine Guo	<i>Forgery Detection by Local Correspondence</i> , 2000 (w/ Azriel Rosenfeld)
Christian Shin	<i>The Roles of Document Structure in Document Image Retrieval and Classification</i> , 2000 (w/ Azriel Rosenfeld)
Huanfeng Ma	<i>Adaptive Analysis and Processing of Structured Multilingual Documents</i> , 2005 (w/ Rama Chellappa)
Yefeng Zheng	<i>Handwriting Identification, Matching and Indexing in Noisy Document Images</i> , 2005 (w/ Rama Chellappa)
Jian Liang	<i>Analysis of Camera Based Document Images</i> , 2006 (w/ Rama Chellappa)
Burcu Karagol-Ayan	<i>Lexical Acquisition from Bi-Lingual Dictionaries</i> , 2007 (w/ Amy Weinberg)
Guangyu Zhu	<i>Document Indexing and Retrieval using Logos and Signatures</i> , 2009 (w/ Rama Chellappa)
Xu Liu	<i>Mobile Document Image Analysis</i> , 2008 (w/ Larry Davis)
Mudit Agrawal	<i>Adaptive Algorithms for the Automated Processing of Document Images</i> (w/ Larry Davis)
Jayant Kumar	<i>Efficient Machine Learning Methods for Document Image Analysis</i> , 2013 (w/ Larry Davis)
Peng Ye	<i>Feature Learning and Active Learning for Image Quality Assessment</i> , 2014 (w/ Rama Chellappa)
Le Kang	<i>Document and Natural Image Applications of Deep Learning</i> , 2015 (w/ Rama Chellappa)
Rajiv Jain	<i>Searching Heterogeneous Document Image Collections</i> , 2015 (w/ Larry Davis)
Sungmin Eum	<i>Image and Video Analytics for Document Processing and Event Recognition</i> , 2016 (w/ Joseph Jaja)

Courses Taught

CSE 473/573 - Introduction to Computer Vision and Image Processing (Spring 2019, Spring 2020, Spring 2021, Spring 2022)
 CSE 610 - Automated Analysis of Sporting Event Videos (Spring 2022, Summer 2022)
 CSE 199 – Internet, Computing, and Society (Fall 2021)
 CSE 510 - Edge Intelligence & Computing (Spring 2021)
 EAS 595LEC - Special Topics Fundamentals of Artificial Intelligence (Spring 2020)
 CSE 728 Seminar - Applications of Generative Adversarial Networks (Fall 2019)
 CSE 728 Seminar - Legal, Ethical and Social Issues in Artificial Intelligence (Spring 2019)
 CMSC 828D - Automated Document and Video Analysis, (Various)
 LBSC 690 - Information Technology, (Various)
 SPIE Short Course: Document Image Processing and Analysis on Android Devices, 2013

Professional Activities

Editorial

Founding Editor-in-Chief: Int. Journal on Document Analysis and Recognition 1995-2005
 Managing Editor: Int. Journal on Document Analysis and Recognition 2005-present

Conference Leadership

General Chair: Int. Conference on Document Analysis and Recognition, 2023
 General Chair: Int. Conference on Document Analysis and Recognition, 2013
 General Chair: Int. Conference on Mobile and Ubiquitous Multimedia, 2004
 General Chair: Summit on Arabic and Chinese Handwriting, 2006

 Co-General Chair: Int. Workshop on Document Analysis Systems, 2010
 Co-General Chair: Int. Workshop on Camera Based Document Analysis and Recognition, 2005-2007
 Co-General Chair: Center for Discrete Mathematics and Theoretical Computer Science
 Workshop on Video Mining, 2002

Organizer and General Chair: Symposium on Document Image Understanding Tech, 1995-2005

Program Chair: Int. Conf. on Document Analysis and Recognition, 2007

Track Chair: Multimedia and Document Analysis, Processing and Retrieval, Int. Conference on Pattern Recognition, 2010

Track Chair: Int. Conf. on Multimedia & Expo 2006

Area Chair: Document Analysis, Biometrics and Pattern Recognition Applications, Int. Conference on Pattern Recognition, 2014

Area Chair: Document Analysis Track, Int. Conference on Pattern Recognition, 2012

Member: Int. Conference on Document Analysis and Recognition Awards Committee, 2007-2007

Publicity Chair: Int. Conf. on Document Analysis and Recognition, 2017

Publicity Chair: Int. Conf. on Document Analysis and Recognition, 2005, 2011

Conference Committees

Program Committee: Int. Workshop on Frontiers in Arabic Handwriting Recognition, 2010
 Program Committee: Int. Conference on Advances in Multimedia, 2009, 2010
 Program Committee: Int. Conference on Document Analysis and Recognition, 1999-2017
 Program Committee: Int. Conference on Frontiers in Handwriting Recognition, 2012
 Program Committee: SPIE-Document Recognition and Retrieval, 1997-2014
 Program Committee: Int. Workshop on Video Mining, 2008
 Program Committee: Int. Workshop on Mobile Multimedia Processing, 2008, 2010
 Program Committee: Int. Conference on Information Sciences, Signal Processing and It's Applications 2007
 Program Committee: Int. Conference on Image and Video Retrieval, 2009
 Program Committee: Int. Workshop on Document Analysis Systems, 2002- 2008, 2012-2014
 Program Committee: Digital Access to Textual Cultural Heritage, 2014
 Program Committee: Multilingual OCR, 2011-2013
 Program Committee: SPIE-Multimedia on Mobile Devices, 2005-2012
 Program Committee: Int. Workshop on Service-Oriented Community Coordinated Multimedia, 2009
 Program Committee: Int. Symposium on Signal Processing 2007
 Program Committee: Annual ACM Symposium on Applied Computing –Document Engineering

2006-2007, 2010 -2012

Program Committee: Int. Workshop on Graphics Recognition 2001-2013

Program Committee: Int. Conference on Mobile and Ubiquitous Multimedia, 2001-2003, 2007-2008, 2011-2012

Program Committee: Workshop on Analytics for Noisy Unstructured Text Data, 2006-2011

Program Committee: LCPR 2006

Program Committee: Document Image Analysis for Libraries, 2006

Program Committee: Int. Conference on Pattern Recognition, 2002, 2004, 2006

Program Committee: Annual ACM Symposium on Applied Computing 2005

Program Committee: CBDAR 2005, 2011, 2013

Program Committee: IEEE Int. Conference on Multimedia & Expo, 2005, 2006

Program Committee: Int. Conference on Multimodal Interfaces 2005

Program Committee: ACM Symposium on Applied Computing 2005

Program Committee: Int. Conference on Image Analysis and Processing 1999-2005

Program Committee: IEEE Int. Symposium on Multimedia Software Engineering 2004

Program Committee: Int. Workshop on Distributed Shared Memory on Clusters 2003

Program Committee: Int. Workshop on Multimedia Data and Document Engineering 2003

Program Committee: Second Int. Workshop on Web Document Analysis 2001, 2003

Program Committee: Document Layout Interpretation and its Applications, 1999-2003

Program Committee: Int. Workshop on Information Retrieval 2001

Program Committee: Event, 2001

Program Committee: DIAUDD, 1999

Program Committee: Int. Workshop on Performance Evaluation Issues in Multilingual OCR, 1999

Program Committee: Brazilian Symposium on Document Image Analysis, 1997

Program Committee: Workshop on Syntactic & Structural Pattern Recognition, 1995

Other Activities

IAPR Publications and Publicity Committee, 2001-2014

Chair, IAPR Publications and Publicity Committee, 2010-2014

External Examiner: Alan McCabe, Handwritten Signature Verification Using Complementary Statistical Models”, School of Information Technology, James Cook University of North Queensland, November 2003.

External Examiner: Jarvi Sadri Automatic Segmentation and Recognition of Unconstrained Handwritten Numeral Strings, Doctoral Thesis, Computer Science Department, Concordia University, Montreal, April 2007.

Reviewer: National Science Foundation

Reviewer: Computer Vision and Pattern Recognition

Reviewer: Graphical Models and Image Processing

Reviewer: IEEE Transactions on Pattern Analysis and Machine Intelligence

Reviewer: Pattern Recognition

Reviewer: Pattern Recognition Letters

Awards and Recognition

Award for Excellence: The Under Secretary of Defense (Acquisition, Technology and Logistics) for support to the CENTCOM Technology Tiger Team. Transition of key technology to deployed troops.

Results Matter, 2016: Presented to a program manager by the Director of DARPA for successful program development, implementation and transition that exemplifies the DARPA mission of preventing strategic surprise.

Best Student Paper: Sungmin Eum and David Doermann, “Content Selection Using Frontalness Evaluation of Multiple Frames”, International Conference on Pattern Recognition, 2016 (Piero Zamperoni Award)

Fellow of the IEEE: for contributions to Research and development of automatic analysis and processing of document page imagery 2014

Fellow of the IAPR: for contributions to the field of document image analysis and in recognition of service to the IAPR, 2014

Distinguished Visiting Fellowship, Royal Academy of Engineering: Practical systems for mass-digitization using cloud computing. University of Salford, Manchester UK, 2011-2013

Best Student Paper: Jayant Kumar, Peng Ye and David Doermann. "Learning Document Structure for Retrieval and Classification." International Conference on Pattern Recognition, 2012

Best Student Paper Award: Mudit Agrawal and David Doermann. “Context-Aware and Content-Based Dynamic Voronoi Page Segmentation”, International Workshop on Document Analysis systems, 2011

Best Paper Award: Xu Liu, David Doermann and Huiping Li, “Imaging as an Alternative Data Channel for Camera Phones” International Conference on Mobile and Ubiquitous Multimedia, 2006

Best Student Paper: Yi Li, Yefeng Zheng, David Doermann and Stefan Jaeger, “A New Algorithm for Detecting Text Line in Handwritten Documents”, International Workshop on Frontiers in Handwriting Recognition, 2006

Honorary Doctorate of Technology Sciences: University of Oulu for contributions to digital media processing and document analysis research, 2002.

University of Maryland Invention of the Year Finalist: 2007 and 2009.

Pattern Recognition 1994 Outstanding Paper Honorable Mention.

Journal Publications

- J65. Y. Yang et al., “Long term 5G network traffic forecasting via modeling non-stationarity with deep learning,” Commun Eng, vol. 2, no. 1, Art. no. 1, Jun. 2023, doi: 10.1038/s44172-023-00081-4.
- J64. P. Yan, S. Ahmed, and D. Doermann, “Context-Aware Chart Element Detection.” arXiv, May 06, 2023. doi: 10.48550/arXiv.2305.04151.
- J63. X. Gong et al., “Federated Learning with Privacy-Preserving Ensemble Attention Distillation,” IEEE Transactions on Medical Imaging, pp. 1–1, 2022, doi: 10.1109/TMI.2022.3213244.

- J62. Z. Zhang, Y. Zhu, R. Rai, and D. Doermann, "PIMNet: Physics-Infused Neural Network for Human Motion Prediction," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 8949–8955, Oct. 2022, doi: 10.1109/LRA.2022.3188892.
- J61. J. Hyysalo et al., "Smart mask—Wearable IoT solution for improved protection and personal health," *Internet of Things*, vol. 18, p. 100511, 2022.
- J60. J. Zhao et al., "Data-adaptive binary neural networks for efficient object detection and recognition," *Pattern Recognition Letters*, vol. 153, pp. 239–245, 2022.
- J59. S. Xu, C. Liu, B. Zhang, J. Lü, G. Guo, and D. Doermann, "BiRe-ID: Binary Neural Network for Efficient Person Re-ID," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 18, no. 1s, pp. 1–22, 2022.
- J58. Y. Zhai et al., "Adaptive Two-Stream Consensus Network for Weakly-Supervised Temporal Action Localization," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 01, pp. 1–16, 2022.
- J57. H. Chen et al., "Binarized neural architecture search for efficient object recognition," *International Journal of Computer Vision*, vol. 129, no. 2, pp. 501–516, 2021.
- J56. R. Wang, B. Zhang, L. Zhuo, Q. Ye, and D. Doermann, "Cogradient Descent for Dependable Learning," *arXiv preprint arXiv:2106.10617*, 2021.
- J55. S. Xue, H. Chen, C. Xie, B. Zhang, X. Gong, and D. Doermann, "Fast and Unsupervised Neural Architecture Evolution for Visual Representation Learning," *IEEE Computational Intelligence Magazine*, vol. 16, no. 3, pp. 22–32, 2021.
- J54. J. Zhao et al., "Data-Adaptive Binary Neural Networks for Efficient Object Detection and Recognition," *Pattern Recognition Letters*, 2021.
- J53. M. Mao, Y. Tian, B. Zhang, Q. Ye, W. Liu, and D. Doermann, "IffDetector: inference-aware feature filtering for object detection," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
- J52. C. Liu et al., "Rectified Binary Convolutional Networks with Generative Adversarial Learning," *Int J Comput Vis*, Jan. 2021.
- J51. Z. Zhang, R. Rai, S. Chowdhury, and D. Doermann, "MIDPhyNet: Memorized infusion of decomposed physics in neural networks to model dynamic systems," *Neurocomputing*, vol. 428, pp. 116–129, Mar. 2021, doi: 10.1016/j.neucom.2020.11.042.
- J50. H. Chang, Y. Chen, B. Zhang, and D. Doermann, "Multi-UAV Mobile Edge Computing and Path Planning Platform Based on Reinforcement Learning," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2021.
- J49. Zhang, Duona, Wenrui Ding, Baochang Zhang, Chunhui Liu, Jungong Han, and David Doermann. 2021. "Learning Modulation Filter Networks for Weak Signal Detection in Noise." *Pattern Recognition* vol. 109, p. 107590, Jan. 2021.
- J48. H. Chen, B. Zhang, X. Zheng, J. Liu, R. Ji, D. Doermann, G. Guo, and others, "Binarized neural architecture search for efficient object recognition," *International Journal of Computer Vision*, vol. 129, no. 2, pp. 501–516, 2021.
- J47. W. Zhao, T. Ma, X. Gong, B. Zhang, and D. Doermann, "A Review of Recent Advances of Binary Neural Networks for Edge Computing," *IEEE Journal on Miniaturization for Air and Space Systems*, pp. 1–1, 2020.
- J46. A. Behjat, C. Zeng, R. Rai, I. Matei, D. Doermann, and S. Chowdhury, "A physics-aware learning architecture with input transfer networks for predictive modeling," *Applied Soft Computing*, vol. 96, p. 106665, Nov. 2020.
- J45. Davila, K, S Setlur, D Doermann, UK Bhargava, and V Govindaraju. 2020. "Chart Mining: A Survey of Methods for Automated Chart Analysis." *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- J44. Liu, C., Ding, W., Hu, Y., Xia, X., Zhang, B., Liu, J. and Doermann, D., 2020. Circulant Binary Convolutional Networks for Object Recognition. *IEEE Journal of Selected Topics in Signal Processing*.

- J43. Bhowmik, S., Sarkar, R., Das, B. and Doermann, D., 2019. GiB: A Game Theory Inspired Binarization Technique for Degraded Document Images. *IEEE Transactions on Image Processing*, 28(3), pp.1443-1455
- J42. Shibaprasad Sen, Ankan Bhattacharyya, Pawan Kumar Singh, Ram Sarkar, Kaushik Roy, and David Doermann. "Application of Structural and Topological Features to Recognize Online Handwritten Bangla Characters." *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)* 17, no. 3 (2018): 20.
- J41. Showmik Bhowmik, Ram Sarkar, Mita Nasipuri, and David Doermann. "Text and non-text separation in offline document images: a survey." *International Journal on Document Analysis and Recognition (IJ DAR)* 21, no. 1-2 (2018): 1-20.
- J40. Sungmin Eum, and David Doermann, Planar content selection in images and videos using frontality, In *Pattern Recognition Letters*, 2017, ISSN 0167-8655, <https://doi.org/10.1016/j.patrec.2017.08.020>.
- J39. H. Chang, Y. Chen, B. Zhang, and D. Doermann, "Multi-UAV Mobile Edge Computing and Path Planning Platform Based on Reinforcement Learning," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2021
- J38. Qixiang Ye and David Doermann, Text detection and recognition in imagery: A survey, *IEEE transactions on pattern analysis and machine intelligence* 37 (2015), no. 7, 1480-1500.
- J37. Roy, Ankush, Biswajit Halder, Utpal Garain, and David S. Doermann. "Machine-assisted authentication of paper currency: an experiment on Indian banknotes." *International Journal on Document Analysis and Recognition (IJ DAR)* 18, no. 3 (2015): 271-285.
- J36. Jayant Kumar, Peng Ye, and David Doermann, Structural similarity for document image classification and retrieval, *Pattern Recognition Letters* 43 (2014), 119-126.
- J35. Mudit Agrawal and David Doermann. "Clutter noise removal in binary images." *International Journal on Document Analysis and Recognition*, 16, pp. 351-369, December 2013.
- J34. Ankush Roy, Biswajit Halder, Utpal Garain and David Doermann. "Automatic Authentication of Banknotes." *IEEE Transactions on Information Forensics & Security*, 2012.
- J33. Peng Ye and D. Doermann. "No-Reference Image Quality Assessment using Visual Codebooks." *IEEE Transactions on Image Processing*, 21:7, pp. 3129-3138, July 2012.
- J32. T. Steinherz, D. Doermann, E. Rivlin and N. Intrator. Off-Line Loop Investigation for Handwriting Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2), pages 193-209, February 2009.
- J31. L. Yi, Y. Zheng, D. S. Doermann and S. Jaeger. Script-Independent Text Line Segmentation in Freestyle Handwritten Documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1313-1329, August 2008.
- J30. Xu Liu, D. Doermann and Huiping Li. VCode - Pervasive Data Transfer Using Video Barcode. *IEEE Transactions on Multimedia*, 10(3), pages 361-371, April 2008.
- J29. Xu Liu and David Doermann. Mobile Retriever: Access to Digital Documents from their Physical Source. *International Journal on Document Analysis and Recognition*, 11(1), pages 19-27, September 2008.
- J28. Guangyu Zhu, Yefeng Zheng, David Doermann and Stefan Jaeger. Signature Detection and Matching for Document Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11), pages 2015-2031, November 2009.
- J27. Guangyu Zhu, Xiaodong Yu, Yi Li and David Doermann. Language Identification for Handwritten Document Images Using A Shape Codebook. *Pattern Recognition*, 42, pages 3184-3191, December 2009.
- J26. Daniel DeMenthon and David Doermann. Video Retrieval of Near-Duplicates using k-Nearest Neighbor Retrieval of Spatio-Temporal Descriptors. *Multimedia Tools and Applications (MTAP)*, 30, 2006.
- J25. J. Liang, D. DeMenthon and D. Doermann. Mosaicing of Camera-captured Documents without Pose Restriction. *Computer Vision and Image Understanding*, 2006. (Submitted).

- J24. J. Liang, D. DeMenthon and D. Doermann. Geometric Rectification of Camera-captured Document Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, July 2006. (Submitted).
- J23. Y. Zheng and D. Doermann. Robust Point Matching for Nonrigid Shapes By Preserving Local Neighborhood Structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4), pages 643-649, April 2006.
- J22. Y. Zheng, H. Li and D. Doermann. A Parallel Line Detection Algorithm Based on HMM Decoding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5), pages 777-792, 2005.
- J21. J. Liang, D. Doermann and H. Li. Camera-Based Analysis of Text and Documents: A Survey. *International Journal on Document Analysis and Recognition*, 7(2+3), pages 83-104, July 2005.
- J20. M. Balcells-Capellades, D. DeMenthon and D. Doermann. An Appearance-based Approach for Consistent Labeling of Humans and Objects in Video. *Pattern Analysis and Applications*, pages 1433-7451, November 2004. (<http> version available online from Springer, printed version 2005).
- J19. W. Byrne, D. Doermann, M. Franz, S. Gustman, J. Hajic, D. Oard, M. Picheny, J. Psutka, B. Ramabhadran, D. Soergel, T. Ward, and W.-J. Zhu. Automatic Recognition of Spontaneous Speech for Access to Multilingual Oral History Archives. *IEEE Transactions on Speech and Audio Processing*, Special Issue on Spontaneous Speech Processing, 12(4), pages 420-435, July 2004.
- J18. Y. Zheng, H. Li, and D. Doermann. Machine Printed Text and Handwriting Identification in Noisy Document Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3), pages 337-353, 2004.
- J17. H. Ma, and D. Doermann. Adaptive Hindi OCR Using Generalized Hausdorff Image Comparison. *ACM Transactions on Asian Language Information Processing*, 26(2), pages 198-213, 2004.
- J16. H. Ma, B. Karagol-Ayan, D. Doermann, D. Oard, and J. Wang. Parsing and Tagging of Bilingual Dictionaries. *Traitement Automatique Des Langues*, 44(2), pages 125-150, 2003.
- J15. K. Guo, D. Doermann and A. Rosenfeld. Forgery Detection by Local Correspondence. *International Journal on Pattern Recognition and Artificial Intelligence*, 15(4), pages 579-641, 2001.
- J14. C. Shin, D. Doermann, and A. Rosenfeld. Classification of Document Pages Using Structure-Based Features. *International Journal on Document Analysis and Recognition*, 3(4), pages 232-247, 2001.
- J13. O. Kia and D. Doermann. Residual coding in document image compression. *IEEE Transactions on Image Processing*, 9(6):961-969, 2000.
- J12. H. Li, D. Doermann, and O. Kia. Automatic text detection and tracking in digital video. *IEEE Transactions on Image Processing - Special Issue on Image and Video Processing for Digital Libraries*, 9(1):147-156, 2000.
- J11. O. Kia and D. Doermann. Document image coding for processing and retrieval. *Journal of VLSI Signal Processing*, 20:121-135, 1998.
- J10. D. Doermann, E. Rivlin and A. Rosenfeld. The Function of Documents. *International Journal of Computer Vision*, 16, pages 799-814, 1998.
- J9. D. Doermann. The Indexing and Retrieval of Document Images: A Survey. *Computer Vision and Image Understanding*, 70(3), pages 287-298, 1998.
- J8. D. Doermann, H. Li and O. Kia. The Detection of Duplicates in Document Image Databases. *Image and Vision Computing*, 16, pages 907-920, 1998.
- J7. V. Kobla and D. Doermann. Indexing and Retrieval of MPEG-compressed video. *The Journal of Electronic Imaging*, pages 294-307, 1998.
- J6. O. Kia, D. Doermann, A. Rosenfeld, and R. Chellappa. Symbolic compression and processing of document images. *Computer Vision and Image Understanding*, 70(3):335-349, 1998.
- J5. K. Etemad, D. Doermann, and R. Chellappa. Multiscale document page segmentation using soft decision integration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 92-96, 1997.
- J4. D. S. Doermann, E. Rivlin, and I. Weiss. Applying algebraic and differential invariants for logo recognition. *Machine Vision and Applications*, 9(2):73-86, 1996.
- J3. O. Hori and D. Doermann. Table-form structure analysis based on box-driven reasoning. *IEICE Transactions on Information and Systems*, pages 542-547, 1996.

- J2. D. S. Doermann and A. Rosenfeld. Recovery of temporal information from static images of handwriting. *International Journal of Computer Vision*, 52(1-2), pages 143-164, 1994.
- J1. D. S. Doermann, V. Varma and A. Rosenfeld. Instrument grasp: A model and its effects on handwritten strokes. *Pattern Recognition*, 27(2), pages 233-245, 1994.

Books

- B4. David Scott Doermann, Karl Tombre, et al., *Handbook of document image processing and recognition*, 2014.
- B3. Yefeng Zheng, David Doermann and Huiping Li. *Handwritten Document Image Processing*. VDM Verlag Dr. Muller, 2008.
- B2. A. Rosenfeld, D. Doermann and D. DeMenthon. *Video Mining*. Kluwer Academic Publishers, 2003
- B1. B. Perry, S-K. Chang, J. Dinsmore, D. Doermann, A. Rosenfeld, and S. Stevens. *Content Based Access to Multimedia Information: From Technology Trends to State of the Art*. Kluwer, 1999.

Book Chapters

- B4. S. Jaeger, H. Ma and D. Doermann. *Machine Learning in Document Analysis and Recognition: Combining Classifiers with Informational Confidence*. Chapter: tba. Springer, LNCS, 2007. (to appear).
- B3. S. Tuljakov, S. Jaeger, V. Govindaraju and D. Doermann. *Machine Learning in Document Analysis and Recognition: Review of Classifier Combination Methods*. Chapter: tba. Springer, LNCS, 2007. (to appear).
- B2. D. Dori, D. Doermann, C. Shin, et. al. *Handbook on Optical Character Recognition 2nd Document Image Analysis*, chapter *The Representation of Document Structure: A Generic Object-Process Approach*, pages 421–456. World Scientific, 1997.
- B1. O. Hori and D.S. Doermann. Quantitative measurement of the performance of raster-to-vector conversion algorithms. In *Graphics Recognition: Methods and Applications*, pages 57–68. Springer, 1996.

Technical Reports (not appearing in journals)

- T11. S. Kibey. *Tools for Advanced Video Metadata Modeling*. Technical Report: LAMP-TR-141, University of Maryland, College Park, February 2007.
- T10. S. Jaeger. *Uncertainty in Neural and Physical Processes*. Technical Report: LAMP-TR-137/CAR-TR-1016/CS-TR-4841/UMIACS-TR-2006-56, University of Maryland, College Park, December 2006.
- T9. Y. Li, Y. Zheng, D. Doermann and S. Jaeger. *Script-Independent Text Line Segmentation in Freestyle Handwritten Documents*. Technical Report: LAMP-TR-136/ CS-TR-4836/ UMIACS-TR-2006-51/ CFAR-TR-1017, University of Maryland, College Park, November 2006.
- T8. M. Luo, D. DeMenthon, X. Yu and D. Doermann. *SOFTCBIR: Object Searching in Videos Combining Keypoint Matching and Graduated Assignment*. Technical Report: LAMP-TR-132/CAR-TR-1013/CS-TR-4804/UMIACS-TR-2006-25, University of Maryland, College Park, May 2006.
- T7. G. Zi. *GroundTruth Generation and Document Image Degradation*. Technical Report: LAMP-TR-121/CAR-TR-1008/CS-TR-4699/UMIACS-TR-2005-08, University of Maryland, College Park, May 2005.
- T6. Y. Zheng, H. Li, and D. Doermann. *A Parallel Line Detection Algorithm Based on HMM Decoding*. Technical Report: LAMP-TR-109/CAR-TR-994/CS-TR-4545/UMIACS-TR-2003-1113, University of Maryland, College Park, December 2003.

- T5. H. Ma, and D. Doermann. Adaptive Hindi OCR Using Generalized Hausdorff Image Comparison. Technical Report: LAMP-TR-105/CFAR-TR-987/CS-TR-4519/UMIACS-TR-2003-87, University of Maryland, College Park, August 2003.
- T4. O. Okun, D. Doermann, and M. Pietikainen. Page Segmentation and Zone Classification: The State of the Art. Technical Report: LAMP-TR-036/CAR-TR-927/CS-TR-4079, University of Maryland, College Park, November 1999.
- T3. R. Jones, D. DeMenthon, and D. Doermann. Building mosaics from video using MPEG motion vectors. Technical Report: LAMP-TR-035/CAR-TR-918/CS-TR-4034, University of Maryland, College Park, July 1999.
- T2. D. Doermann. Document image understanding - 1996. Technical Report LAMP-TR-008/CAR-TR-853/CS-TR-3775, University of Maryland, College Park, 1997.
- T1. V. Kobla, D.S. Doermann, K.I. Lin, and C. Faloutsos. Feature normalization for video indexing and retrieval. Technical Report LAMP-TR-003/CAR-TR-847/CS-TR-3732, University of Maryland, College Park, 1996.

Conference and ArXiv Publications

- C229. J. Lal, A. Mitkari, M. Bhosale, and D. Doermann, "LineFormer: Rethinking Line Chart Data Extraction as Instance Segmentation." arXiv, May 02, 2023. doi: 10.48550/arXiv.2305.01837.
- C228. R. Wang, Z. Liu, B. Zhang, G. Guo, and D. Doermann, "Few-Shot Learning with Complex-Valued Neural Networks and Dependable Learning," *Int J Comput Vis*, vol. 131, no. 1, pp. 385–404, Jan. 2023, doi: 10.1007/s11263-022-01700-x.
- C227. L. Song et al., "PREF: Predictability Regularized Neural Motion Fields." arXiv, Sep. 21, 2022. doi: 10.48550/arXiv.2209.10691.
- C226. X. Gong et al., "Self-supervised Human Mesh Recovery with Cross-Representation Alignment." arXiv, Sep. 10, 2022. doi: 10.48550/arXiv.2209.04596.
- C225. X. Gong, L. Khaidem, W. Zhu, B. Zhang, and D. Doermann, "Uncertainty Learning towards Unsupervised Deformable Medical Image Registration," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 2484–2493.
- C224. R. Wang, L. Yang, B. Zhang, W. Zhu, D. Doermann, and G. Guo, "Confidence Dimension for Deep Learning based on Hoeffding Inequality and Relative Evaluation," arXiv preprint arXiv:2203.09082, 2022.
- C223. X. Gong et al., "Preserving Privacy in Federated Learning with Ensemble Cross-Domain Knowledge Distillation," 2022.
- C222. Fang, C., Hu, Y., Zhang, B. and Doermann, D., 2021. The Fusion of Neural Architecture Search and Destruction and Construction Learning: First Classified. In *Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10-15, 2021, Proceedings, Part VIII* (pp. 480-489). Springer International Publishing.
- C220. X. Gong, X. Xia, W. Zhu, B. Zhang, D. Doermann, and L. Zhuo, "Deformable Gabor Feature Networks for Biomedical Image Classification," 2021, pp. 4004–4012, Accessed: Jan. 26, 2021.
- C219. X. Gong, S. Chen, B. Zhang, and D. Doermann, "Style Consistent Image Generation for Nuclei Instance Segmentation," 2021, pp. 3994–4003, Accessed: Jan. 26, 2021.
- C218. T. Ma, Y. Wang, J. Shao, B. Zhang, and D. Doermann, "Orthogonal Features Fusion Network for Anomaly Detection," in *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, Dec. 2020, pp. 33–37, doi: 10.1109/VCIP49819.2020.9301755.
- C217. Hu, Y., Jiang, X., Liu, X., Zhang, B., Han, J., Cao, X., Doermann, D., 2020. NAS-Count: Counting-by-Density with Neural Architecture Search. ArXiv Prepr. ArXiv200300217.
- C216. Chen, Hanlin, Baochang Zhang, Song Xue, Xuan Gong, Hong Liu, Rongrong Ji, and David Doermann. 2020. "Anti-Bandit Neural Architecture Search for Model Defense." ArXiv:2008.00698 [Cs], August. <http://arxiv.org/abs/2008.00698>.

- C215. Zhuo, L., Zhang, B., Chen, H., Yang, L., Chen, C., Zhu, Y., Doermann, D., 2020. CP-NAS: Child-Parent Neural Architecture Search for 1-bit CNNs. ArXiv Prepr. ArXiv200500057.
- C214. Chanda, S., GV, A.C., Brun, A., Hast, A., Pal, U., Doermann, D., 2019. Face Recognition-A One-Shot Learning Perspective, in: 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). IEEE, pp. 113–119.
- C213. Gv, A.C., Chanda, S., Pal, U., Doermann, D., 2019. One-Shot Learning-Based Handwritten Word Recognition, in: Asian Conference on Pattern Recognition. Springer, Cham, pp. 210–223.
- C212. Chen, H., Zhuo, L.A., Zhang, B., Zheng, X., Liu, J., Doermann, D. and Ji, R., 2019. Binarized Neural Architecture Search. arXiv preprint arXiv:1911.10862.
- C211. Zhuo, L.A., Zhang, B., Chen, C., Ye, Q., Liu, J. and Doermann, D., 2019. Calibrated Stochastic Gradient Descent for Convolutional Neural Networks.
- C210. Zhang, B., Chen, C., Ye, Q., Liu, J. and Doermann, D., 2019, July. Calibrated Stochastic Gradient Descent for Convolutional Neural Networks. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 9348-9355).
- C209. Liu, C., Ding, W., Xia, X., Zhang, B., Gu, J., Liu, J., Ji, R. and Doermann, D., 2019. Circulant Binary Convolutional Networks: Enhancing the Performance of 1-bit DCNNs with Circulant Back Propagation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2691-2699).
- C208. Li, Y., Lin, S., Zhang, B., Liu, J., Doermann, D., Wu, Y., Huang, F. and Ji, R., 2019. Exploiting Kernel Sparsity and Entropy for Interpretable CNN Compression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2800-2809).
- C207. Gu, J., Li, C., Zhang, B., Han, J., Cao, X., Liu, J. and Doermann, D., 2019, July. Projection convolutional neural networks for 1-bit cnns via discrete back propagation. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 8344-8351).
- C206. Lin, S., Ji, R., Yan, C., Zhang, B., Cao, L., Ye, Q., Huang, F. and Doermann, D., 2019. Towards optimal structured CNN pruning via generative adversarial learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2790-2799).
- C205. Zhu, Y., Zhou, Y., Xu, H., Ye, Q., Doermann, D. and Jiao, J., 2019. Learning Instance Activation Maps for Weakly Supervised Instance Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3116-3125).
- C204. Jiang, X., Xiao, Z., Zhang, B., Zhen, X., Cao, X., Doermann, D. and Shao, L., 2019. Crowd Counting and Density Estimation by Trellis Encoder-Decoder Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 6133-6142).
- C203. Eum, Sungmin, Hyungtae Lee, Heesung Kwon, and David Doermann. "IOD-CNN: Integrating Object Detection Networks for Event Recognition." arXiv preprint arXiv:1703.07431 (2017).
- C202. Sungmin Eum, and David Doermann. "Content selection using frontalness evaluation of multiple frames." In Pattern Recognition (ICPR), 2016 23rd International Conference on, pp. 3404-3409. IEEE, 2016.
- C201. Jingtao Xu, Peng Ye, Qiaohong Li, Yong Liu, and David Doermann, No-reference document image quality assessment based on high order image statistics, Image Processing (ICIP), 2016 IEEE International Conference on, IEEE, 2016, pp. 3289-3293.
- C200. Eum, Sungmin, Hyungtae Lee, and David S. Doermann. "JH2R: Joint Homography Estimation for Highlight Removal." In BMVC, pp. 49-1. 2015.
- C199. Sravanthi Bondugula, Varun Manjunatha, Larry S Davis, and David Doermann, Shoe: Sibling hashing with output embeddings, Proceedings of the 23rd ACM International Conference on Multimedia, ACM, 2015, pp. 823-826.
- C198. Xianzhi Du, David Doermann, and Wael Abd-Almageed, A graphical model approach for matching partial signatures, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1465-1472.
- C197. Hongxing Gao, Marcal Rusiñol, Dimosthenis Karatzas, Josep Lladós, Rajiv Jain, and David Doermann, Novel line verification for multiple instance focused retrieval in document collections,

- Document Analysis and Recognition (ICDAR), 2015 13th International Conference on, IEEE, 2015, pp. 481-485.
- C196. Rajiv Jain and David Doermann, Localized document image change detection, Document Analysis and Recognition (ICDAR), 2015 13th International Conference on, IEEE, 2015, pp. 786-790.
- C195. Le Kang, Peng Ye, Yi Li, and David Doermann, Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks, Image Processing (ICIP), 2015 IEEE International Conference on, IEEE, 2015, pp. 2791-2795.
- C194. Pawan Kumar Singh, Ram Sarkar, Mita Nasipuri, and David Doermann, Word-level script identification for handwritten Indic scripts, Document Analysis and Recognition (ICDAR), 2015 13th International Conference on, IEEE, 2015, pp. 1106-1110.
- C193. Xianzhi Du, David Doermann, and Wael Abd-Almageed, Signature matching using supervised topic models, Pattern Recognition (ICPR), 2014 22nd International Conference on, IEEE, 2014, pp. 327-332.
- C192. Le Kang, Peng Ye, Yi Li, and David Doermann, A deep learning approach to document image quality assessment, Image Processing (ICIP), 2014 IEEE International Conference on, IEEE, 2014, pp. 2570-2574.
- C191. Sungmin Eum and David Doermann, Sharpness-aware document image mosaicing using graphcuts, Image Processing (ICIP), 2014 IEEE International Conference on, IEEE, 2014, pp. 2575-2579.
- C190. Shan Gao, Zhenjun Han, David Doermann, and Jianbin Jiao, Depth structure association for RGB-D multi-target tracking, Pattern Recognition (ICPR), 2014 22nd International Conference on, IEEE, 2014, pp. 4152-4157.
- C189. Rajiv Jain and David Doermann, Combining local features for offline writer identification, Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on, IEEE, 2014, pp. 583-588.
- C188. Le Kang, Yi Li, and David Doermann, Orientation robust text line detection in natural images, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 4034-4041.
- C187. Le Kang, Peng Ye, Yi Li, and David Doermann, Convolutional neural networks for no-reference image quality assessment, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1733-1740.
- C186. Le Kang, Peng Ye, Yi Li, and David Doermann, A deep learning approach to document image quality assessment, Image Processing (ICIP), 2014 IEEE International Conference on, IEEE, 2014, pp. 2570-2574.
- C185. Paul Rodrigues, David Zajic, David Doermann, Michael Bloodgood, and Peng Ye, Detect-ing structural irregularity in electronic dictionaries using language modeling, arXiv preprint arXiv:1410.8149 (2014).
- C184. Qixiang Ye and David S Doermann, Robust scene text detection using integrated feature discrimination, Image Processing (ICIP), 2014 IEEE International Conference on, IEEE, 2014, pp. 1678-1682.
- C183. David Zajic, Michael Maxwell, David Doermann, Paul Rodrigues, and Michael Bloodgood, Correcting errors in digital lexicographic resources using a dictionary manipulation language, arXiv preprint arXiv:1410.7787 (2014).
- C182. Jialing Zou, Qixiang Ye, Yanting Cui, David Doermann, and Jianbin Jiao, A belief based correlated topic model for trajectory clustering in crowded video scenes, Pattern Recognition (ICPR), 2014 22nd International Conference on, IEEE, 2014, pp. 2543-2548.
- C181. Jingtao Xu, Peng Ye, Yong Liu, and David Doermann, No-reference video quality assessment via feature learning, Image Processing (ICIP), 2014 IEEE International Conference on, IEEE, 2014, pp. 491-495.
- C180. Peng Ye and David Doermann, Active sampling for subjective image quality assessment, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 4249-4256.

- C179. Peng Ye, Jayant Kumar, and David Doermann, Beyond human opinion scores: blind im-age quality assessment based on synthetic scores, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 4241-4248.
- C178. Rajiv Jain and David Doermann, VisualDiff: Document image verification and change detection, Document Analysis and Recognition (ICDAR), 2013 12th International Conference on, IEEE, 2013, pp. 40-44.
- C177. Rajiv Jain and David Doermann, Writer identification using an alphabet of contour gradient descriptors, Document Analysis and Recognition (ICDAR), 2013 12th International Conference on, IEEE, 2013, pp. 550-554.
- C176. Rajiv Jain, Douglas W Oard, and David Doermann, Scalable ranked retrieval using document images, IS&T /SPIE Electronic Imaging, International Society for Optics and Photonics, 2013, pp. 90210K-90210K.
- C175. Jayant Kumar and David Doermann, Unsupervised classification of structurally similar document images, Document Analysis and Recognition (ICDAR), 2013 12th International Conference on, IEEE, 2013, pp. 1225-1229.
- C174. Peng Ye and David Doermann, Combining preference and absolute judgements in a crowd-sourced setting, Proc. of Intl. Conf. on Machine Learning, 2013, pp. 1-7.
- C173. Jayant Kumar, Peng Ye and D. Doermann. "A Dataset for Quality Assessment of Camera Captured Document Images." International Workshop on Camera-Based Document Analysis and Recognition (CBDAR), August 2013.
- C172. Peng Ye and David Doermann. "Document Image Quality Assessment: A Brief Survey." Intl. Conf. on Document Analysis and Recognition (ICDAR), 2013.
- C171. Jayant Kumar and David Doermann. "Unsupervised Classification of Structurally Similar Document Images." Intl. Conf. on Document Analysis and Recognition (ICDAR 13), 2013.
- C170. Peng Ye, Jayant Kumar, Le Kang and David Doermann. "Real-time No-Reference Image Quality Assessment based on Filter Learning." Intl. Conf. on Computer Vision and Pattern Recognition (CVPR), 2013.
- C169. Qixiang Ye and D. Doermann. "Scene Text Detection via Integrated Discrimination of Component Appearance and Consensus." International Workshop on Camera-Based Document Analysis and Recognition (CBDAR), August 2013.
- C168. Xianzhi Du, Wael Abd-Almageed and David S. Doermann. "Large-scale Signature Matching using Multi-Stage Hashing." 12th International Conference on Document Analysis and Recognition, pp. 976--980, 2013.
- C167. Peng Ye and D. Doermann. "Learning features for predicting OCR accuracy." International Conference on Pattern Recognition (ICPR), pp. 3204--3207, 2012.
- C166. Peng Ye, Jayant Kumar, Le Kang and David Doermann. "Unsupervised Feature Learning Framework for No-reference Image Quality Assessment." Intl. Conf. on Computer Vision and Pattern Recognition (CVPR 2012), pp. 1098-1105, 2012.
- C165. Le Kang, David Doermann, Huiagu Cao, Rohit Prasad and Prem Natarajan. "Local Segmentation of Touching Characters using Contour based Shape Decomposition." Document Analysis Systems, pp. 460-464, 2012.
- C164. Rajiv Jain and David Doermann. "Logo Retrieval in Document Images." Document Analysis Systems, pp. 135 - 139, 2012.
- C163. Michael Bloodgood, Peng Ye, Paul Rodrigues, David Zajic and David Doermann. " A Random Forest System Combination Approach for Error Detection in Digital Dictionaries ." Innovative hybrid approaches to the processing of textual data, EACL 2012 Workshop, pp. 78-86, April 2012.
- C162. Jayant Kumar, Francine Chen and David Doermann. "Sharpness Estimation for Document and Scene Images." International Conference on Pattern Recognition (ICPR 2012), pp. 3292-3295, 2012.
- C161. Le Kang, Jayant Kumar, Peng Ye and David Doermann. "Learning Text-line Segmentation using Codebooks and Graph Partitioning." International Conference on Frontiers in Handwriting Recognition (ICFHR), 2012. (63-68).

- C160. Jayant Kumar, Peng Ye and David Doermann. "Learning Document Structure for Retrieval and Classification." International Conference on Pattern Recognition (ICPR 2012), pp. 1558-1561, 2012.
- C159. Zhiyi Song, Safa Ismael, Stephen Grimes, David Doermann and Stephanie Strassel. "Linguistic Resources for Handwriting Recognition and Translation Evaluation." Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), 2012.
- C158. Jayant Kumar, Le Kang, David Doermann and Wael Abd-Almageed. "Segmentation of Handwritten Textlines in Presence of Touching Components." Intl. Conf. on Document Analysis and Recognition (ICDAR 11), pp. 109-113, 2011.
- C157. Jayant Kumar, Rohit Prasad, Huiagu Cao, W. Abd-Almageed, David Doermann and Prem Natarajan. "Shape Codebook based Handwritten and Machine Printed Text Zone Extraction." Document Recognition and Retrieval, pp. 7874:1-8, San Francisco, January 2011.
- C156. Mudit Agrawal and David Doermann. "Stroke-like Pattern Noise Removal in Binary Document Images." International Conference on Document Analysis and Recognition, pp. 17-21, 2011.
- C155. Rajiv Jain and David Doermann. "Offline Writer Identification using K-Adjacent Segments." International Conference on Document Analysis and Recognition, pp. 769-773, 2011.
- C154. Peng Ye and David Doermann. "No-Reference Image Quality Assessment based on Visual Codebook." Int. Conf. on Image Processing (ICIP 2011), pp. 3150–3153, 2011.
- C153. Le Kang and David Doermann. "Template based Segmentation of Touching Components in Handwritten Text Lines." 11th Intl. Conf. on Document Analysis and Recognition (ICDAR'11), pp. 569-573, 2011.
- C152. Jayant Kumar, Jaishanker Pillai and David Doermann. "Document Image Classification and Labeling using Multiple Instance Learning." Intl. Conf. on Document Analysis and Recognition (ICDAR 11), pp. 1059-1063, 2011.
- C151. Jayant Kumar and David Doermann. "Fast Rule-line Removal using Integral Images and Support Vector Machines." Intl. Conf. on Document Analysis and Recognition (ICDAR 11), pp. 584-588, 2011.
- C150. Paul McNamee, James Mayfield, Dawn Lawrie, Doug Oard and David Doermann. "Cross Language Entity Linking." IJCNLP: International Joint Conference on Natural Language Processing, 2011. (TO APPEAR).
- C149. Paul McNamee, James Mayfield, Doug Oard, David Doermann, Tan Xu and Ke Wu. "Cross-Language Entity Linking in Maryland during a Hurricane." TAC, 2011.
- C148. Utpal Garain, Jiaul Paik, Tamaltaru Pal, Prasenjit Majumder, David Doermann and Douglas Oard. "Overview of the FIRE 2011 RISOT Task." FIRE, 2011.
- C147. Utpal Garain, David Doermann and Douglas D. Oard. "Maryland at FIRE 2011: Retrieval of OCREd Bengali." FIRE, 2011.
- C146. David Zajic, Michael Maxwell, David Doermann, Paul Rodrigues and Michael Bloodgood. "Correcting Errors in Digital Lexicographic Resources Using a Dictionary Manipulation Language." Electronic lexicography in the 21st century: new applications for new users (eLEX2011), pp. 297-301, 2011.
- C145. Paul Rodrigues, David Zajic, Michael Bloodgood, Peng Ye and David Doermann. "Detecting Structural Irregularity in Electronic Dictionaries Using Language Modeling." Electronic lexicography in the 21st century: new applications for new users (eLEX2011), pp. 227-232, 2011.
- C144. David Doermann, Elena Zotkina and Huiping Li. "GEDI - A Groundtruthing Environment for Document Images." Ninth IAPR International Workshop on Document Analysis Systems (DAS 2010), 2010.
- C143. Mudit Agrawal and David Doermann. "Context-Aware and Content-Based Dynamic Voronoi Page Segmentation." The Ninth IAPR International Workshop on Document Analysis Systems, pp. 73-80, 2010.
- C142. W. Seo, Mudit Agrawal and David Doermann. "Performance Evaluation Tools for Zone Segmentation and Classification (PETS)." International Conference on Pattern Recognition, pp. 503-506, 2010.

- C141. Jayant Kumar, W. Abd-Almageed, Le Kang and David Doermann. "Handwritten Arabic Text Line Segmentation using Affinity Propagation." Document Analysis Systems, pp. 135-142, 2010.
- C140. Radu Dondera, D. Doermann and L. Davis. Action recognition based on human movement characteristics. WMVC, pages 1-8, 2009.
- C139. W. Abd-Almageed, Jayant Kumar and David Doermann. Page Rule-Line Removal using Linear Subspaces in Monochromatic Handwritten Arabic Documents. Intl. Conf. on Document Analysis and Recognition (ICDAR 09), pages 768-772, 2009.
- C138. Mudit Agrawal and David Doermann. Clutter Noise Removal in Binary Document Images. International Conference on Document Analysis and Recognition (ICDAR '09), pages 556-560, 2009.
- C137. Mudit Agrawal and David Doermann. Voronoi++: A Dynamic Page Segmentation approach based on Voronoi and Docstrum features. International Conference on Document Analysis and Recognition (ICDAR '09), pages 1011-1015, 2009.
- C136. Guangyu Zhu and D. Doermann. Logo Matching for Document Image Retrieval. International Conference on Document Analysis and Recognition (ICDAR 2009), pages 606-610, 2009.
- C135. Guangyu Zhu, Xiaodong Yu, Yi Li and David Doermann. Unconstrained Language Identification Using A Shape Codebook. The 11th International Conference on Frontiers in Handwriting Recognition (ICFHR 2008), pages 13-18, 2008.
- C134. Mudit Agrawal and David Doermann. Re-Targetable OCR with Intelligent Character Segmentation. DAS, September 2008.
- C133. Guangyu Zhu, Yefeng Zheng and David Doermann. Signature-based Document Image Retrieval. The 10th European Conference on Computer Vision (ECCV 2008), pages 1-14, 2008
- C132. Guangyu Zhu, Xiaodong Yu, Yi Li and David Doermann. Learning Visual Shape Lexicon for Document Image Content Recognition. The 10th European Conference on Computer Vision (ECCV 2008), pages 1-14, 2008.
- C131. Mudit Agrawal and David Doermann. Re-Targetable OCR with Intelligent Character Segmentation. DAS, September 2008.
- C130. Xu Liu, D. Doermann and H. Li. A Camera-based Mobile Data Channel: Capacity and Analysis. ACM International Conference on Multimedia(to appear), October 2008.
- C129. Xu Liu, David Doermann and H. Li. Camera Phone Based Tools for the Visually Impaired. The First International Workshop on Mobile Multimedia Processing(to appear), December 2008.
- C128. Xu Liu and David Doermann. A Camera Phone Based Currency Reader for the Visually Impaired. The Tenth International ACM SIGACCESS Conference on Computers and Accessibility(to appear), October 2008.
- C127. X. Yu, Y. Li, C. Fermuller and D. Doermann. Object Detection Using Shape Codebook. British Machine Vision Conference (BMVC'07), December 2007 (accepted).
- C126. G. Zhu and D. Doermann. Automatic Document Logo Detection. 9th International Conference on Document Analysis and Recognition (ICDAR'07), 2007 (accepted).
- C125. Ryan Farrell, David Doermann and Larry S. Davis. Learning Higher-order Transition Models in Medium-scale Camera Networks. Workshop on Omnidirectional Vision, Camera Networks and Nonclassical Cameras (ICCV'07), pages 1-8, 2007.
- C124. Xu Liu and D. Doermann. Mobile Retriever - Finding Document with a Snapshot. CBDAR 07, pages 29-34, September 2007.
- C123. Z. Lin, L. S. Davis and D. Doermann. Hierarchical Part-Template Matching for Human Detection and Segmentation. IEEE International Conference on Computer Vision (ICCV'07), 2007. (accepted).
- C122. Z. Lin, L. S. Davis, D. Doermann and D. DeMenthon. An Interactive Approach to Pose-Assisted and Appearance-based Segmentation of Humans. Workshop on Interactive Computer Vision (ICV'07), 2007. (accepted).
- C121. Z. Lin, L. S. Davis, D. Doermann and D. DeMenthon. Simultaneous Appearance Modeling and Segmentation for Matching People under Occlusion. Asian Conference on Computer Vision (ACCV'07), 2007. (accepted).

- C120. J. Hannuksela, P. Sangi, J. Heikkila, X. Liu and D. Doermann. Document Image Mosaicing with Mobile Phones. International Conference on Image Analysis and Processing (ICIAP'07), pages 1-8, September 2007.
- C119. X. Liu, H. Li and D. S. Doermann. Imaging as an Alternative Data Channel for Camera Phones. MUM06, 193-198, December 2006. (Best Paper).
- C118. B. Karagol-Ayan, D. Doermann and A. Weinberg. Adaptive Transformation-based Learning for Improving Dictionary Tagging. Conference on European Chapter of the Association for Computational Linguistics, pages 257-264, April 2006.
- C117. G. Zhu, S. Jaeger and D. Doermann. A Robust Stamp Detection Framework on Degraded Documents. International Conference on Document Recognition and Retrieval XIII (IS&T, SPIE), pages 1-9, January 2006.
- C116. J. Liang, D. DeMenthon and D. Doermann. Camera-Based Document Image Mosaicing. Proceedings of International Conference on Document Recognition and Retrieval XIII (IS&T, SPIE), pages 476-479, 2006.
- C115. B. Karagol-Ayan, D. Doermann and A. Weinberg. Morphology Induction from Limited Noisy Data Using Approximate String Matching. ACL Special Interest Group in Computational Phonology (SIGPHON 2006), pages 60-68, June 2006.
- C114. Y. Li, Y. Zheng and D. Doermann. Detecting Text Line in Handwritten Documents. ICPR'06, pages 1030-1033, 2006.
- C113. Y. Li, Y. Zheng, D. Doermann and S. Jaeger. A New Algorithm for Detecting Text Line in Handwritten Documents. 10th International Workshop on Frontiers in Handwriting Recognition, pages 35-40, 2006.
- C112. Stefan Jaeger, Guangyu Zhu, David Doermann, Kevin Chen and Summit Sampat. DOCLIB: a Software Library for Document Processing. Proceedings of International Conference on Document Recognition and Retrieval XIII (IS&T, SPIE), pages 1-9, January 2006. (San Jose).
- C111. C. Shin and D. Doermann. Document Page Image Classification Based on Similarity of Visual Appearance. 8th IASTED International Conference on Signal and Image Processing, August 2006. (to appear).
- C110. H. Ma and D. S. Doermann. Font identification using the grating cell texture operator. SPIE Conference on Document Recognition and Retrieval XXII, pages 148-156, 2005.
- C109. Y. Zheng and D. Doermann. Robust Point Matching for Two-Dimensional Nonrigid Shapes. Proceedings in the ICASSP'04 IEEE International Conference on Computer Vision, pages 1561-1566, October 2005.
- C108. H. Ma and D. Doermann. Adaptive OCR with Limited User Feedback. Int. Conf. on Document Analysis and Recognition (ICDAR'05), pages 814-818, August 2005.
- C107. Y. Zheng and D. Doermann. Handwriting Matching and Its Application to Handwriting Synthesis. ICDAR, pages 861-865, 2005.
- C106. M. Huang, D. DeMenthon, D. Doermann and L. Golebiowski. Document Ranking by Layout Relevance. ICDAR, pages 362-366, August 2005.
- C105. X. Liu and D. Doermann. Using Computer Vision to Detect Web Browser Display Errors. 3rd Web Document Analysis Workshop (on ICDAR 2005, pages 5-9, July 2005).
- C104. J. Liang, D. DeMenthon and D. Doermann. Unwarping Images of Curved Documents Using Global Shape Optimization. International Workshop on Camera-based Document Analysis and Recognition, pages 25-29, 2005.
- C103. J. Liang, D. Doermann and H. Li. Camera-Based Analysis of Text and Documents: A Survey. International Journal on Document Analysis and Recognition, 2005.
- C102. K. Chen, S. Jaeger, G. Zhu and D. Doermann. DOCLIB: A document processing research tool. Symposium on Document Image Understanding Technology, pages 159-163, 2005.
- C101. X. Liu, D. Doermann and H. Li. Fast Camera Motion Estimation for Hand-Devices and Applications. International Conference on Mobile and Ubiquitous Multimedia, pages 103-108, 2005.

- C100. J. Liang, D. DeMenthon and D. Doermann. Flattening Curved Document Images. CVPR, pages 338-345, 2005.
- C99. Y. Yu and D. Doermann. Model of Object-Based Coding for Surveillance Video. Proceedings in the ICASSP'04 IEEE International Conference on Computer Vision, pages 693-696, March 2005.
- C98. S. Jaeger, H. Ma and D. Doermann. Identifying Script on Word-Level with Informational Confidence. Int. Conf. on Document Analysis and Recognition, pages 416-420, 2005.
- C97. NM. Ghanem, D.S. Doermann, L.S. Davis, and D.F. DeMenthon. Mining tools for surveillance video. SPIE International Symposium on Electronic Imaging, Storage and Retrieval Methods and Applications for Multimedia, 5307, pages 259-270, 2004.
- C96. N. Ghanem, D. DeMenthon, D. Doermann and L. Davis. Representation and Recognition of Events in Surveillance Video Using Petri Nets. IEEE Workshop on Event Mining 2004, CVPR2004, pages 112-112, 2004.
- C95. H. Ma, and D. Doermann. Word Level Script Identification on Scanned Document Images. SPIE Conference on Document Recognition and Retrieval, pages 124-135, 2004.
- C94. H. Kang and D. Doermann. Product Approximation by Minimizing the Upper Bound of Bayes Error Rate for Bayesian Combination of Classifiers. ICPR, pages 252-255, 2004.
- C93. J. Liang and D. Doermann. Content Features for Logical Document Labeling. Proc. SPIE Conference on Document Recognition and Retrieval X, pages 189-196, 2003.
- C92. Y. Zheng, H. Li, and D. Doermann. Background Line Detection with a Stochastic Model. Symposium on Document Image Analysis and Retrieval (in conjunction with IEEE CVPR'03, pages VII, 424-431, 2003.
- C91. D. Doermann, and A. Karunanidhi. Video Analysis for Pervasive Environments. International Conference on Multimedia and Expo, (ICME), pages 161-164, 2003.
- C90. X. Gilbert, H. Li, and D. Doermann. Sports Video Classification Using HMM. International Conference on Multimedia and Expo, (ICME), pages 345-348, 2003.
- C89. M. Balcells-Capellades, D. Doermann, D. DeMenthon, and R. Chellappa. An Appearance Based Approach for Human and Object Tracking. IEEE International Conference on Image Processing, pages 85-88, September 2003.
- C88. B. Karagol-Ayan, D. Doermann, and B. Dorr. Acquisition of Bilingual MT Lexicons from OCREd Dictionaries. Proceedings of the Ninth Machine Translation Summit, pages 208-215, 2003.
- C87. H. Ma, and D. Doermann. Gabor Filter Based Multi-class Classifier for Scanned Document Images. International Conference on Document Analysis and Recognition (ICDAR), pages 968-972, 2003.
- C86. D. Doermann, J. Liang, and H. Li. Progress in Camera Based Document Image Analysis. International Conference on Document Analysis and Recognition (ICDAR), pages 606-617, 2003.
- C85. H. Kang, and D. Doermann. Evaluation of the Information-Theoretic Construction of Multiple Classifier Systems. International Conference on Document Analysis and Recognition (ICDAR), pages 789-793, 2003.
- C84. H. Kang, and D. Doermann. Combining Multiple Classifiers based on Third-Order Dependency. International Conference on Document Analysis and Recognition (ICDAR), pages 21-25, 2003.
- C83. Y. Zheng, H. Li, and D. Doermann. Text Identification in Noisy Document Images Using Markov Random Field. International Conference on Document Analysis and Recognition (ICDAR), pages 599-603, 2003.
- C82. Y. Zheng, H. Li, and D. Doermann. A Model-based Line Detection Algorithm in Documents. International Conference on Document Analysis and Recognition (ICDAR), pages 44-48, 2003.
- C81. H. Ma, and D. Doermann. Bootstrapping Structured Page Segmentation. SPIE Conference Document Recognition and Retrieval, pages 179-188, JAN 2003.
- C80. J. Liang, D. Doermann, M. Ma and J. Guo. Page Classification through Logical Labeling. International Conference on Pattern Recognition, pages 477-480, 2002.
- C79. A. Karunanidhi, D. Doermann, N. Parekh, and V. Rautio. Video analysis applications for pervasive environments. International Conference on Mobile and Ubiquitous Multimedia, pages 48-55, 2002.

- C78. Y. Zheng, H. Li, and D. Doermann. Segmentation and Identification of Handwriting in Noisy Documents. IAPR Conference on Document Analysis System, pages 95-105, 2002.
- C77. D. Doermann and J. Liang. Binary Document Image Similarity Using Multiple Texture Features. Symposium on Document Image Understanding Technology, pages 181-193, April 2001.
- C76. J. Liang, and D. Doermann. Logical Labeling of Document Images Using Layout Graph Matching with Adaptive Learning. IAPR Conference on Document Analysis System, pages 212-223, 2002.
- C75. A.F. Smeaton, P. Over, C.J. Costello, A.P. de Vries, D. Doermann, A. Hauptmann, M.E. Rorvig, J.R. Smith, and L. Wu. The TREC2002 Video Track: Information Retrieval on Digital Video Information. Research and Advanced Technology for Digital Libraries. European Conference ECDL2002, pages 266-275, 2002.
- C74. D. Doermann, N. Intrator, E. Rivlin, and T. Steinherz. Hidden Loop Recovery for Handwriting Recognition. International Workshop on Frontiers in Handwriting Recognition, pages 375-380, 2002.
- C73. H. Li, and D. Doermann. Text Quality Estimation in Digital Video. SPIE Conf. on Document Recognition and Information Retrieval, pages 232-243, 2002.
- C72. D. Doermann, H. Ma, B. Karagol-Ayan, and D. Oard. Lexicon Acquisition from Bilingual Dictionaries. SPIE Photonic West Electronic Imaging Conference, pages 37-48, 2002.
- C71. M. Rautiainen, and D. Doermann. Temporal Color Correlograms in Video Retrieval. International Conference on Pattern Recognition, pages 267-270, 2002.
- C70. P. Gupta, D. Doermann, and D. DeMenthon. Beam Search for Feature Selection in Automatic SVM Defect Classification. International Conference on Pattern Recognition, pages 212-215, 2002.
- C69. C. Wolf, and D. Doermann. Binarization of Low Quality Text using a Markov Random Field Model. International Conference on Pattern Recognition, pages 160-163, 2002.
- C68. V.Y. Mariano, J. Min, J.-H. Park, R. Kasturi, D. Mihalcik, D. Doermann, and T. Drayer. Performance Evaluation of Object Detection Algorithms. International Conference on Pattern Recognition, pages 965-969, 2002.
- C67. K. Darwish, D. Doermann, R. Jones, D. Oard, and M. Rautiainen. TREC-10 Experiments at Maryland: CLIR and Video. TREC 2001, page 552, 2001.
- C66. Y. Qi, D. Doermann and D. DeMenthon. Hybrid Independent Component Analysis and Support Vector Machine Learning Scheme For Face Detection. ICASSP, 2001.
- C65. D. Doermann and D. DeMenthon. Data and content based adaptation for video over low bandwidth networks. In SPIE- Multimedia and Systems Applications III, pages 69–78, 2000.
- C64. A. Koivisto, P. Pietikainen, J. Sauvola, and D. Doermann. Live multimedia adaptation through wireless hybrid networks. In ICME, 2000.
- C63. C. Shin, D. Doermann, and A. Rosenfeld. Classification of document page images based on visual similarity of layout structures. In SPIE Conference on Document Recognition and Retrieval (VII), pages 182–190, JAN 2000.
- C62. M. Vuilleumier-Stuckelberg D. DeMenthon and D. Doermann. Hidden Markov models for images. In ICPR, volume 3, pages 147–150, 2000.
- C61. D. Doermann and D. Mihalcik. Tools and techniques for video performance evaluation. In ICPR, volume 4, pages 167–170, 2000.
- C60. K. Guo, D. Doermann, and A. Rosenfeld. Off-line skilled forgery detection using stroke and substroke features. In ICPR, volume 2, pages 355–359, 2000.
- C59. K. Yoon, D. DeMenthon, and D. Doermann. Event detection from mpeg video in the compressed domain. In ICPR, volume 1, pages 819–822, 2000.
- C58. H. Li and D. Doermann. Superresolution-based enhancement of text in digital video. In ICPR, volume 1, pages 847–850, 2000.
- C57. H. Li and D. Doermann. A video text detection system based on automated training. In ICPR, volume 2, pages 223–226, 2000.
- C56. V. Kobla, D. DeMenthon, and D. Doermann. Detection of slow-motion replay sequences for identifying sports videos. In Proceedings of IEEE 1999 Workshop on Multimedia Signal Processing, pages 135–140, 1999.

- C55. H. Li and D.S. Doermann. Text enhancement in digital video using multiple frame integration. In ACM –Multimedia 99, Orlando, Florida, pages 19–22, 1999.
- C54. R. Jones, D. DeMenthon, and D. Doermann. Building mosaics from video using mpeg motion vectors. In ACM Conference on Multimedia, volume 2, pages 29–32, NOV 1999.
- C53. M.V. Stuckelberg and D. Doermann. Model based graphics recognition. In GREC, pages 136–143, 1999.
- C52. M.V. Stuckelberg and D. Doermann. On musical score recognition using probabilistic reasoning. In Proceedings of the International Conference on Document Analysis and Recognition, pages 115–118, 1999.
- C51. L. Xingyuan, D. Doermann, W. Oh, and W. Gao. A robust method for unknown forms analysis. In Proceedings of the International Conference on Document Analysis and Recognition, pages 531–534, 1999.
- C50. H. Li, O. Kia, and D. Doermann. Text enhancement in digital video. In Proceedings of SPIE - Conference on Document Recognition and Retrieval VI, pages 2–9, Jan 1999.
- C49. V. Kobla, D. DeMenthon, and D. Doermann. Special effect edit detection using videotrails: a comparison with existing techniques. In Proceedings of SPIE - Conference on Storage and Retrieval for Image and Video Databases VII, volume 3656, pages 302–313, 1999.
- C48. V. Kobla, D.S. Doermann, and C. Faloutsos. Developing high-level representations of video clips using videotrails. In Proceedings of SPIE - Conference on Storage and Retrieval for Image and Video Databases VI, volume 3312, pages 81–92, 1998.
- C47. D. DeMenthon, D.S. Doermann, and V. Kobla. Video summarization by curve simplification. In Proceedings of ACM - Multimedia 98, Bristol, England, volume 433981, pages 211–218, 1998.
- C46. H. Li and D. Doermann. Automatic identification of text in digital video key frames. In Proceedings of ICPR, pages 129–132, 1998.
- C45. H. Li, D. Doermann, and O. Kia. Text extraction and recognition in digital video. In IAPR Workshop on Document Analysis Systems, pages 119–128, 1998.
- C44. H. Li and D. Doermann. Automatic text tracking in digital videos. In Proceedings of IEEE 1998 Workshop on Multimedia Signal Processing, pages 21–26, 1998.
- C43. J. Sauvola, D. Doermann, H. Kauniskangas, C. Shin, M. Koivusaari, and M. Pietikainen. Graphical tools and techniques for querying document databases. In Brazilian Symposium on Document Image Analysis, pages 213–224, 1997.
- C42. J. Sauvola, H. Kauniskangas, D. Doermann, and M. Pietikainen. Techniques for automated testing of document analysis algorithms. In Brazilian Symposium on Document Image Analysis, pages 201–212, 1997.
- C41. H. Kauniskangas, J. Sauvola, M. Pietikainen, and D. Doermann. Content-based image retrieval using composite features. In Proceedings of the 1997 Scandinavian Conference on Image Analysis, pages 35–42, 1997.
- C40. O. Kia and D. Doermann. Integrated segmentation and clustering for enhanced compression of document images. In International Conference on Document Analysis and Recognition, page 406, 1997.
- C39. D. Doermann, H. Li, and O. Kia. The detection of duplicates in document image databases. In Proceedings of the International Conference on Document Analysis and Recognition, pages 314–318, 1997.
- C38. D. Doermann, A. Rosenfeld, and E. Rivlin. The function of documents. In Proceedings of the International Conference on Document Analysis and Recognition, pages 1077–1081, 1997.
- C37. K. Guo, D. Doermann, and A. Rosenfeld. Local correspondence for detecting random forgeries. In Proceedings of the International Conference on Document Analysis and Recognition, pages 319–323, 1997.
- C36. D. Doermann. The retrieval of document images: A brief survey. In Proceedings of the International Conference on Document Analysis and Recognition, pages 945–949, 1997.

- C35. D. Doermann, J. Sauvola, S. Haapakoski, H. Kauniskangas, T. Seppanen, and M. Pietikainen. A distributed management system for testing document image database analysis algorithms. In *Proceedings of the International Conference on Document Analysis and Recognition*, pages 989–995, 1997.
- C34. S. Zhong, D. Doermann, and A. Rosenfeld. Image indexing with minimum adaptive spatial segmentation. In *VISUAL 1997*, 1997.
- C33. V. Kobla and D. Doermann. Extracting features for indexing MPEG-compressed video. In *Proceedings of the IEEE First Workshop on Multimedia Signal Processing*, pages 337–342, 1997.
- C32. O. Kia and D. Doermann. OCR-based rate-distortion analysis of residual coding. In *International Conference on Image Processing*, pages 690–693, 1997.
- C31. O. Kia and D. Doermann. The role of compressed document images in transmission and retrieval. In *IEEE First Workshop on Multimedia Signal Processing*, pages 331–336, 1997.
- C30. V. Kobla, D. Doermann, and C. Faloutsos. VideoTrails: Representing and visualizing structure in video sequences. In *Proceeding of the ACM International Multimedia Conference*, pages 335–346, 1997.
- C29. J. Sauvola, D. Doermann, and M. Pietikainen. Locally adaptive document skew detection. In *Proceedings of the SPIE - Document Recognition IV*, pages 96–108, 1997.
- C28. V. Kobla, D. Doermann, K. Lin, and C. Faloutsos. Compressed domain video indexing techniques using DCT and motion vector information in MPEG video. In *Proceedings of the SPIE Conference on Storage and Retrieval for Still Image and Video Databases V*, volume 3022, pages 200–211, 1997.
- C27. V. Kobla, D.S. Doermann, and K-I. Lin. Archiving, indexing, and retrieval of video in compressed domain. In *SPIE Conference of Multimedia Storage and Archiving Systems*, volume 2916, pages 78–89, 1996.
- C26. O. Kia, D. Doermann, and R. Chellappa. Compressed domain document retrieval and analysis. In *SPIE Conference of Multimedia Storage and Archiving Systems*, volume 2916, pages 176–187, 1996.
- C25. D. Doermann, J. Sauvola, H. Kauniskangas, C. Shin, M. Pietikainen, and A. Rosenfeld. The development of a general framework for intelligent document image retrieval. In *International Workshop on Document Analysis Systems*, pages 605–632, 1996.
- C24. O. Kia and D. Doermann. Structure preserving document image compression and transmission. In *Proceedings of the International Conference on Image Processing*, pages 193–196, 1996.
- C23. O. Kia and D. Doermann. Structural compression for document analysis. In *Proceedings of the International Conference on Pattern Recognition*, pages 664–668, 1996.
- C22. E. Rivlin D. Doermann and A. Rosenfeld. A theory of document functionality. In *Proceedings of FLAIRS*, pages 429–433, 1996.
- C21. D. S. Doermann and O. Kia. Hybrid Thinning through Reconstruction. In *Proceedings of the International Conference on Document Analysis and Recognition*, pages 632–635, 1995.
- C20. O. Hori and D. S. Doermann. Robust Table-form Structure Analysis based on Box-Driven-Reasoning. In *Proceedings of the International Conference on Document Analysis and Recognition*, pages 218–221, 1995.
- C19. O. Hori and D. S. Doermann. Quantitative Measurement of the Performance of Raster-to-vector Conversion Algorithms. In *Proceedings of the First International Workshop on Graphics Recognition*, pages 272–281, 1995.
- C18. D. Doermann and S. Yao. Generating Synthetic Data for Text Analysis Systems. In *Proceedings of the Fourth Symposium on Document Analysis and Information Retrieval*, pages 449–467, 1995.
- C17. D. Doermann. Document Understanding at Maryland. In *Proceedings of the ARPA Image understanding Workshop*, pages 817–826, 1994.
- C16. K. Etemad, D. Doermann and R. Chellappa. Document page decomposition by integration of distributed soft decisions. In *Proceedings of the International Conference on Neural Networks*, pages 4022–4027, 1994.

- C15. K. Etemad, D. Doermann and R. Chellappa. Page Segmentation Using Decision Integration and Wavelet Packet Basis. In Proceedings of the International Conference on Pattern Recognition, pages 345–349, 1994.
- C14. D. S. Doermann and A. Rosenfeld. The processing of form documents. In Proceedings of the International Conference on Document Analysis and Recognition, pages 497–501, 1993.
- C13. D. S. Doermann and R. Furuta. Image based typographic analysis of documents. In Proceedings of the International Conference on Document Analysis and Recognition, pages 769–773, 1993.
- C12. D. S. Doermann, E. Rivlin and I. Weiss. Logo recognition using geometric invariants. In Proceedings of the International Conference on Document Analysis and Recognition, pages 894–897, 1993.
- C11. D. S. Doermann and R. Furuta. Image based typographic analysis of documents. Technical Report 93-008, Hypermedia Research Lab, Department of Computer Science, Texas A&M University, 1993.
- C10. D. S. Doermann. Document Image Understanding: Integrating Recovery and Interpretation. PhD thesis, University of Maryland, College Park, MD, 1993.
- C9. D. Doermann and A. Rosenfeld. The interpretation and recognition of interfering strokes. In Proceedings of the International Workshop on Frontiers in Handwriting Recognition, pages 41–50, 1993.
- C8. D. S. Doermann and A. Rosenfeld. Recovery of temporal information from static images of handwriting. In Proceedings of Computer Vision and Pattern Recognition, pages 162–168, 1992.
- C7. D. S. Doermann and A. Rosenfeld. Temporal clues in handwriting. In Proceedings of the International Conference on Pattern Recognition, pages 317–320, 1992.
- C6. V. Varma and D. S. Doermann. Modeling of a grasp for handwriting. In Modeling and Simulation Conference, pages 2133–2140, Pittsburgh, PA, 1992.
- C5. D. S. Doermann and V. Varma. Simulating pressure variations in handwriting. In Proceedings Modeling and Simulation Conference, pages 2141–2148, Pittsburgh, PA, 1992.
- C4. D. S. Doermann (with Igor J. Eberstein). Use of the Goddard Massively Parallel Processor for acid rain modeling. AGU, Spring Meeting, 1991.
- C3. D. S. Doermann and A. Rosenfeld. Mailpiece preprocessing: The population of character parts. In Proceedings of the USPS Advanced Technology Conference, pages 961–972, 1990.
- C2. D. S. Doermann (with I. J. Eberstein and R. W. Stewart). Calculated chemical effects of the El Chichon eruption on the middle stratosphere. AGU, Spring Meeting, Baltimore, MD, 1990.
- C1. D. S. Doermann (with I. J. Eberstein and R. W. Stewart). Random walk plume model with chemistry using the Goddard Massively Parallel Processor, SIAM Conference on Applied Probability in Science and Engineering, New Orleans, LA, March 5-9, 1990.

Invited Contributions

Keynote Talks

- K7. Document Integrity in an AI World, Virtual Milano, ICPR, 2021
- K6. Image and Video Forensics”, Workshop on Information Forensics, Abu-Dhabi, UAE, 2016
- K5. The Role of Document Analysis in the Paperless Office: One perspective”, Nancy, France, 2015
- K4. The Emerging Impact of Document Image Classification and Retrieval”, International Conference on Multimedia Technology, Guangzhou, China 2013
- K3. The evolution of Document Authentication”, International Conference on Frontiers in Handwriting Recognition, Kolkata, India 2010.
- K2. Progress in Camera Based Document Image Analysis. International Conference on Document Analysis and Recognition, Edinburgh, UK, 2003.
- K1. Document Image Understanding, International Symposium on Document Image Management, Curitiba, Brazil, 1998

Invited Talks

- I13. D. Doermann: Computer Forensics for Document Images, Summer School on Document Analysis, 2018.
- I12. D. Doermann: Document Image Classification and Retrieval, iDIPS Summer School, 2014
- I11. Y. Zheng, H. Li, and D. Doermann. Background Line Detection with a Stochastic Model. Symposium on Document Image Understanding Technology (SDIUT'03), pages 97-109, APR 2003.
- I10. H. Ma, B. Karagol-Ayan, and D. Doermann. Segmenting and Tagging Structured Content. Symposium on Document Image Understanding Technology (SDIUT'03), pages 53-64, APR 2003.
- I9. D. Doermann, and G. Zi. Groundtruth Image Generation from Electronic Text (Demonstration). Symposium on Document Image Understanding Technology (SDIUT'03), pages 309-312, APR 2003.
- I8. C. Shin and D. Doermann. Classification of document page images. In SDIUT99, pages 166–175, 1999.
- I7. D. S. Doermann. "Indexing Video Databases", University of Oulu, Oulu, Finland, 1998.
- I6. D. S. Doermann. "Document Image Understanding", 3rd Annual Intelligence Community Information and Classification Management Conference, College Park, Maryland 1997.
- I5. D. S. Doermann. "Media Processing Research at Maryland", Symposium on Document Image Understanding Technology, Annapolis, MD, 1997.
- I4. D. S. Doermann. "Document and Video Understanding at Maryland", Symposium on Document Image Understanding Technology, Bowie, MD, 1995.
- I3. D. S. Doermann. "Analysis of Structured Documents", DIMUND Workshop on Page Decomposition, Character Recognition and Data Standards, Harpers Ferry, WV, 1993.
- I2. D. S. Doermann. "The DIMUND Document Information Server", DIMUND Workshop on Page Decomposition, Character Recognition and Data Standards, Harpers Ferry, WV, 1993.
- I1. D. S. Doermann. "Recovery in Document Image Analysis", Advanced Information Processing and Analysis Symposium, Tysons Corner, VA, 1993.

I have reviewed and certify that this curriculum vitae is a current and accurate statement of my professional record.



David S. Doermann

January 10, 2023

Date

APPENDIX 2: MATERIALS CONSIDERED IN THE PREPARATION OF THIS DECLARATION

Exhibit No.	Description
1001	U.S. Patent No. 11,470,243 (“the ’243 Patent”)
1002	Prosecution History for the ’243 Patent
1004	U.S. Patent Application Publication No. 2002/0059577 (“Lu-577”)
1005	Ying-li Tian, <i>Evaluation of Face Resolution for Expression Analysis</i> , 2004 Conference on Computer Vision and Pattern Recognition Workshop, Washington, DC, USA, 2004, pp. 82-82, doi: 10.1109/CVPR.2004.334 (“Tian”)
1006	U.S. Patent Application Publication No. 2010/0274372 (“Nielsen-372”)
1007	U.S. Patent Application Publication No. 2010/0066822 (“Steinberg”)
1008	U.S. Patent No. 4,858,000 (“Lu-000”)
1009	U.S. Patent No. 4,677,466 (“Lert”)
1010	U.S. Patent Application Publication No. 2011/0081052
1011	Brent Merritt, <i>A Brief History of Media Measurement</i> , Media, available at https://medium.com/@brentmerritt/a-brief-history-of-media-measurement-f1f28aa807ce
1012	U.S. Patent Application Publication No. 2011/0265110
1013	Bernhard Fröba and Christian Küblbeck, <i>Robust Face Detection At Video Frame Rate Base on Edge Orientation Features</i> , International Conference on Audio- and Video-Based Biometric Person Authentication (2001)
1014	U.S. Patent No. 7,315,631
1015	European Patent Application No. 0932114
1016	Erik Larson, <i>Watching American Watch TV</i> , The Atlantic Monthly (1992)
1017	U.S. Patent No. 6,173,074
1018	U.S. Patent No. 7,359,889
1019	Cha Zhang and Zhengyou Zhang, <i>A Survey of Recent Advances in Face Detection</i> , MSR-TR-2010-66, available at https://www.microsoft.com/en-us/research/publication/a-survey-of-recent-advances-in-face-detection/

Exhibit No.	Description
1020	E. Murphy-Chutorian and M. M. Trivedi, <i>Head Pose Estimation in Computer Vision: A Survey</i> , in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 4, pp. 607-626, April 2009, doi: 10.1109/TPAMI.2008.106
1021	W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, <i>Face recognition: A literature survey</i> , ACM Comput. Surv. 35, 4 (December 2003), 399–458
1022	G. Shakhnarovich, L. Lee and T. Darrell, <i>Integrated face and gait recognition from multiple views</i> , Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, pp. I-I, doi: 10.1109/CVPR.2001.990508
1023	Qaim Rizvi, <i>A Review on Face Detection Methods</i> , Journal of Management Development and Information Technology (2011)
1024	Duane M. Blackburn, Mike Bone, P. Jonathon Phillips, <i>Facial Recognition Vendor Test 2000</i> , Evaluation Report
1025	Vidit Jain and Erik Learned-Miller, <i>FDDB: A Benchmark for Face Detection in Unconstrained Settings</i> (2010)
1026	Avery Wang, <i>The Shazam Music Recognition Service</i> , Commun. ACM 49 (2006)
1027	J. Wang, C. Zhang, and H. Shum, <i>Face Image Resolution versus Face Recognition Performance Based on Two Global Methods</i> (2004)
1028	Declaration of Gordon MacPherson for Tian

APPENDIX 3: CHALLENGED CLAIMS

1[p]. An audience measurement system to obtain exposure data for a media exposure environment,

[1a] the audience measurement system comprising:

memory;

machine readable instructions; and

processor circuitry to execute the machine readable instructions to:

[1b] generate an audio signature of media content presented by a television within the media exposure environment;

[1c] obtain content identifying data corresponding to the presented media content, the content identifying data based on the audio signature of the media content presented by the television within the media exposure environment;

[1d] analyze a sequence of images of the media exposure environment to detect a head appearing in one or more of the images, the sequence of images obtained by a camera while the media content corresponding to the content identifying data is presented by the television;

[1e] determine an orientation of the head with respect to the camera; and

[1f] determine audience identification information based on a match of the head to a known person associated with the media exposure environment; and

[1g] network interface circuitry to output a signal indicative of the content identifying data and the audience identification information to a data collection facility.

4[p]. The audience measurement system of claim 1, wherein the processor circuitry is to:

[4a] reduce 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

[4b] determine the orientation of the head with respect to the camera based on the reduced-resolution image.

5. The audience measurement system of claim 4, wherein the processor circuitry is to determine the audience identification information by: (i) generating a facial signature from a region of a second image of the one or more of the images corresponding to a location of the head in the reduced-resolution image, and (ii) comparing the generated facial signature to a database of facial signatures.

6[p]. The audience measurement system of claim 4, wherein the processor circuitry is to:

[6a] identify a region corresponding to the head within the first image from which the reduced-resolution image was obtained;

[6b] analyze a portion corresponding to the identified region of a second image; and based on the analysis of the region of the second image, determine that the head matches the known person.

8. The audience measurement system of claim 6, wherein the audience identification information includes an identifier associated with the known person and the processor circuitry is to, responsive to a determination that the head matches the known person, cause the identifier associated with the person to be stored in the memory.

9[p]. An audience measurement system, comprising:

[9a] an audience measurement device at a media exposure environment, the audience measurement device to:

[9b] generate an audio signature of media content presented by a television within the media exposure environment;

[9c] obtain content identifying data corresponding to the presented media content, the content identifying data based on the audio signature;

[9d] while the media content corresponding to the content identifying data is presented by the television, collect a first image of the media exposure environment with a camera;

[9e] attempt to detect a head in the first image;

[9f] determine an orientation of the head with respect to the camera;

[9g] determine audience identification information based on an indication that the head matches a known person associated with the audience measurement system; and

[9h] a data collection facility to obtain the content identifying data and the audience identification information from the audience measurement device.

11[p]. The audience measurement system of claim 9, wherein the audience measurement device is to:

[11a] reduce a resolution of the first image of the media exposure environment to obtain a reduced-resolution image, and

[11b] determine the orientation of the head with respect to the camera based on the reduced-resolution image.

12[p]. The audience measurement system of claim 11, wherein the audience measurement device is to:

[12a] identify a region corresponding to the head within the first image from which the reduced-resolution image was obtained; and

[12b] analyze a portion of a second image corresponding to the identified region to determine whether the head matches the known person.

13. The audience measurement system of claim 12, wherein the audience identification information includes an identifier associated with the known person and the audience measurement device is to, responsive to the indication that the head matches the known person, cause recordation of the identifier associated with the known person in a memory.

14. The audience measurement system of claim 11, wherein the audience measurement device is to determine the audience identification information by: (i) generating a facial signature from a region of a second image corresponding to a location of the head in the reduced-resolution image, and (ii) comparing the generated facial signature to a database of facial signatures

16[p]. A method for obtaining exposure data for a media exposure environment, the method comprising:

[16a] generating, using a program detector, an audio signature of media content presented by a television within the media exposure environment;

[16b] obtaining, based on the audio signature, content identifying data corresponding to the presented media content;

[16c] while the media content corresponding to the content identifying data is presented by the television, capturing first and second images of the media exposure environment with a camera;

[16d] analyzing, using an audience detector, the first image to attempt to detect a head in the first image;

[16e] determining, using the audience detector, an orientation of the head with respect to the camera;

[16f] determining, using the audience detector, audience identification information based on an indication that the head matches a known person associated with the to a data collection facility. audience detector; and

[16g] providing the content identifying data and the audience identification information to a data collection facility.

18[p]. The method of claim 16, further comprising:

[18a] reducing a resolution of the first image of the media exposure environment to obtain a reduced-resolution image, and

[18b] wherein the determining of the orientation of the head includes determining the orientation of the head with respect to the camera using the reduced-resolution image.

19[p]. The method of claim 18, further comprising:

[19a] identifying a region corresponding to the head within the first image from which the reduced-resolution image was obtained;

[19b] analyzing a portion of the second image corresponding to the identified region; and based on the analyzing of the region of the second image, determining whether the head matches the known person.

20. The method of claim 19, wherein the audience identification information includes an identifier associated with the known person, and further including recording the identifier associated with the known person in a memory.

APPENDIX 4: UNDERSTANDING OF THE LAW

I have applied the following legal principles provided to me by counsel in arriving at the opinions set forth in this report.

Legal Standard for Prior Art

I am not an attorney. I have been informed by attorneys of the relevant legal principles and have applied them to arrive at the opinions set forth in this declaration.

I understand that the petitioner for inter partes review may request the cancelation of one or more claims of a patent based on grounds available under 35 U.S.C. § 102 and 35 U.S.C. § 103 using prior art that consists of patents and printed publications.

Anticipation and Prior Art

I understand that § 102 specifies when a challenged claim is invalid for lacking novelty over the prior art, and that this concept is also known as “anticipation.” I understand that a prior art reference anticipates a challenged claim, and thus renders it invalid by anticipation, if all elements of the challenged claim are disclosed in the prior art reference. I understand the disclosure in the prior art reference can be either explicit or inherent, meaning it is necessarily present or implied. I understand that the prior art reference does not have to use the same words as the challenged claim, but all of the requirements of the claim

must be disclosed so that a person of ordinary skill in the art could make and use the claimed subject-matter.

In addition, I understand that § 102 also defines what is available for use as a prior art reference to a challenged claim.

Under § 102(a), a challenged claim is anticipated if it was patented or described in a printed publication in the United States or a foreign country before the challenged claim's date of invention.

Under § 102(b), a challenged claim is anticipated if it was patented or described in a printed publication in the United States or a foreign country more than one year prior to the challenged patent's filing date.

Under § 102(e), a challenged claim is anticipated if it was described in a published patent application that was filed by another in the United States before the challenged claim's date of invention, or was described in a patent granted to another that was filed in the United States before the challenged claim's date of invention.

I understand that a challenged claim's date of invention is presumed to be the challenged patent's filing date. I also understand that the patent owner may establish an earlier invention date and "swear behind" prior art defined by § 102(a) or § 102(e) by proving (with corroborated evidence) the actual date on which the

named inventors conceived of the subject matter of the challenged claim and proving that the inventors were diligent in reducing the subject matter to practice.

I understand that the filing date of a patent is generally the filing date of the application filed in the United States that issued as the patent. However, I understand that a patent may be granted an earlier effective filing date if the patent owner properly claimed priority to an earlier patent application.

I understand that when a challenged claim covers several structures, either generically or as alternatives, the claim is deemed anticipated if any of the structures within the scope of the claim is found in the prior art reference.

I understand that when a challenged claim requires selection of an element from a list of alternatives, the prior art teaches the element if one of the alternatives is taught by the prior art.

Legal Standard for Obviousness

I understand that even if a challenged claim is not anticipated, it is still invalid if the differences between the claimed subject matter and the prior art are such that the claimed subject matter would have been obvious to a person of ordinary skill in the pertinent art at the time of the alleged invention.

I understand that obviousness must be determined with respect to the challenged claim as a whole.

I understand that one cannot rely on hindsight in deciding whether a claim is obvious.

I also understand that an obviousness analysis includes the consideration of factors such as (1) the scope and content of the prior art, (2) the differences between the prior art and the challenged claim, (3) the level of ordinary skill in the pertinent art, and (4) “secondary” or “objective” evidence of non-obviousness.

Secondary or objective evidence of non-obviousness includes evidence of: (1) a long felt but unmet need in the prior art that was satisfied by the claimed invention; (2) commercial success or the lack of commercial success of the claimed invention; (3) unexpected results achieved by the claimed invention; (4) praise of the claimed invention by others skilled in the art; (5) taking of licenses under the patent by others; (6) deliberate copying of the claimed invention; and (7) contemporaneous and independent invention by others. However, I understand that there must be a relationship between any secondary evidence of non-obviousness and the claimed invention.

I understand that a challenged claim can be invalid for obviousness over a combination of prior art references if a reason existed (at the time of the alleged invention) that would have prompted a person of ordinary skill in the art to combine elements of the prior art in the manner required by the challenged claim. I understand that this requirement is also referred to as a “motivation to combine,”

“suggestion to combine,” or “reason to combine,” and that there are several rationales that meet this requirement.

I understand that the prior art references themselves may provide a motivation to combine, but other times simple common sense can link two or more prior art references. I further understand that obviousness analysis recognizes that market demand, rather than scientific literature, often drives innovation, and that a motivation to combine references may come from market forces.

I understand obviousness to include, for instance, scenarios where known techniques are simply applied to other devices, systems, or processes to improve them in an expected or known way. I also understand that practical and common-sense considerations should be applied in a proper obviousness analysis. For instance, familiar items may have obvious uses beyond their primary purposes.

I understand that the combination of familiar elements according to known methods is obvious when it yields predictable results. For instance, obviousness bars patentability of a predictable variation of a technique even if the technique originated in another field of endeavor. This is because design incentives and other market forces can prompt variations of it, and predictable variations are not the product of innovation, but rather ordinary skill and common sense.

I understand that a particular combination may be obvious if it was obvious to try the combination. For example, when there is a design need or market

pressure to solve a problem and there are a finite number of identified, predictable solutions, a person of ordinary skill has good reason to pursue the known options within his or her technical grasp. This would result in something obvious because the result is the product not of innovation but of ordinary skill and common sense. However, I understand that it may not be obvious to try a combination when it involves unpredictable technologies.

It is further my understanding that a proper obviousness analysis focuses on what was known or obvious to a person of ordinary skill in the art, not just the patentee. Accordingly, I understand that any need or problem known in the field of endeavor at the time of invention and addressed by the patent can provide a reason for combining the elements in the manner claimed.

It is my understanding that the Manual of Patent Examining Procedure §2143 sets forth the following as exemplary rationales that support a conclusion of obviousness:

- Combining prior art elements according to known methods to yield predictable results;
- Simple substitution of one known element for another to obtain predictable results;
- Use of known technique to improve similar devices (methods, or products) in the same way;

- Applying a known technique to a known device (method, or product) ready for improvement to yield predictable results;
- Choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success;
- Known work in one field of endeavor may prompt variations of it for use in either the same field or a different one based on design incentives or other market forces if the variations are predictable to one of ordinary skill in the art;
- Some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention.

A person of ordinary skill in the art looking to overcome a problem will often use the teachings of multiple publications together like pieces of a puzzle, even though the prior art does not necessarily fit perfectly together. Therefore, I understand that references for obviousness need not fit perfectly together like puzzle pieces. Instead, I understand that obviousness analysis takes into account inferences, creative steps, common sense, and practical logic and applications that a person of ordinary skill in the art would employ under the circumstances.

I understand that a claim can be obvious in light of a single reference, if the elements of the challenged claim that are not explicitly or inherently disclosed in the reference can be supplied by the common sense of one of skill in the art.

I understand that obviousness also bars the patentability of applying known or obvious design choices to the prior art. One cannot patent merely substituting one prior art element for another if the substitution can be made with predictable results. Likewise, combining prior art techniques that are interoperable with respect to one another is generally obvious and not patentable.

In order for a claim to be found invalid based upon a modification or combination of the prior art, there must be reasonable expectation that a person of ordinary skill would have successfully modified or combined the prior art to arrive at the claimed arrangement. This does not mean that it must be certain that a person of ordinary skill would have been successful – the law only requires that the person of ordinary skill in the art would have perceived a reasonable expectation of success in modifying or combining the prior art to arrive at the claimed invention.

In sum, my understanding is that obviousness invalidates claims that merely recite combinations of, or obvious variations of, prior art teachings using understanding and knowledge of one of skill in the art at the time and motivated by the general problem facing the inventor at the time. Under this analysis, the prior art references themselves, or any need or problem known in the field of endeavor

at the time of the invention, can provide a reason for combining the elements of or attempting obvious variations on prior art references in the claimed manner.

Legal Standard for Claim Construction

I understand that before any invalidity analysis can be properly performed, the scope and meaning of the challenged claims must be determined by claim construction.

I understand that a patent may include two types of claims, independent claims and dependent claims. I understand that an independent claim stands alone and includes only the limitations it recites. I understand that a dependent claim depends from an independent claim or another dependent claim. I understand that a dependent claim includes all the limitations that it recites in addition to the limitations recited in the claim (or claims) from which it depends.

In comparing the challenged claims to the prior art, I have carefully considered the patent and its file history in light of the understanding of a person of skill at the time of the alleged invention.

I understand that to determine how a person of ordinary skill would have understood a claim term, one should look to sources available at the time of the alleged invention that show what a person of skill in the art would have understood disputed claim language to mean. It is my understanding that this may include what is called “intrinsic” evidence as well as “extrinsic” evidence.

I understand that, in construing a claim term, one should primarily rely on intrinsic patent evidence, which includes the words of the claims themselves, the remainder of the patent specification, and the prosecution history. I understand that extrinsic evidence, which is evidence external to the patent and the prosecution history, may also be useful in interpreting patent claims when the intrinsic evidence itself is insufficient. I understand that extrinsic evidence may include principles, concepts, terms, and other resources available to those of skill in the art at the time of the invention.

I understand that words or terms should be given their ordinary and accepted meaning unless it appears that the inventors were using them to mean something else or something more specific. I understand that to determine whether a term has special meaning, the claims, the patent specification, and the prosecution history are particularly important, and may show that the inventor gave a term a particular definition or intentionally disclaimed, disavowed, or surrendered claim scope.

I understand that the claims of a patent define the scope of the rights conferred by the patent. I understand that because the claims point out and distinctly claim the subject matter which the inventors regard as their invention, claim construction analysis must begin with and is focused on the claim language itself. I understand that the context of the term within the claim as well as other claims of the patent can inform the meaning of a claim term. For example, because

claim terms are normally used consistently throughout the patent, how a term is used in one claim can often inform the meaning of the same term in other claims. Differences among claims or claim terms can also be a useful guide in understanding the meaning of particular claim terms.

I understand that a claim term should be construed not only in the context of the particular claim in which the disputed term appears, but in the context of the entire patent, including the entire specification. I understand that because the specification is a primary basis for construing the claims, a correct construction must align with the specification.

I understand that the prosecution history of the patent as well as art incorporated by reference or otherwise cited during the prosecution history are also highly relevant in construing claim terms. For instance, art cited by or incorporated by reference may indicate how the inventor and others of skill in the art at the time of the invention understood certain terms and concepts. Additionally, the prosecution history may show that the inventors disclaimed or disavowed claim scope, or further explained the meaning of a claim term.

With regard to extrinsic evidence, I understand that all evidence external to the patent and prosecution history, including expert and inventor testimony, dictionaries, and learned treatises, can also be considered. For example, technical dictionaries may indicate how one of skill in the art used or understood the claim

terms. However, I understand that extrinsic evidence is considered to be less reliable than intrinsic evidence, and for that reason is generally given less weight than intrinsic evidence.

I understand that in general, a term or phrase found in the introductory words or preamble of the claim, should be construed as a limitation if it recites essential structure or steps, or is necessary to give meaning to the claim. For instance, I understand preamble language may limit claim scope: (i) if dependence on a preamble phrase for antecedent basis indicates a reliance on both the preamble and claim body to define the claimed invention; (ii) if reference to the preamble is necessary to understand limitations or terms in the claim body; or (iii) if the preamble recites additional structure or steps that the specification identifies as important.

On the other hand, I understand that a preamble term or phrase is not limiting where a challenged claim defines a structurally complete invention in the claim body and uses the preamble only to state a purpose or intended use for the invention. I understand that to make this determination, one should review the entire patent to gain an understanding of what the inventors claim they invented and intended to encompass in the claims.

I understand that 35 U.S.C. § 112 ¶ 6 created an exception to the general rule of claim construction called a “means plus function” limitation. These types of

terms and limitations should be interpreted to cover only the corresponding structure described in the specification, and equivalents thereof. I also understand that a limitation is presumed to be a means plus function limitation if (a) the claim limitation uses the phrase “means for”; (b) the “means for” is modified by functional language; and (c) the phrase “means for” is not modified by sufficient structure for achieving the specified function.

I understand that a structure is considered structurally equivalent to the corresponding structure identified in the specification only if the differences between them are insubstantial. For instance, if the structure performs the same function in substantially the same way to achieve substantially the same result. I further understand that a structural equivalent must have been available at the time of the issuance of the claim.