

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

Reply Declaration of Dr. David Scott Doermann

I, David Scott Doermann, declare as follows:

1. I previously submitted a declaration in this proceeding (Ex-1003). I have been asked to respond to certain opinions of Dr. Eli Saber and arguments in the Patent Owner's Response. I have considered the additional materials listed in the attached Appendix 1. I make this declaration based upon my own personal knowledge and, if called upon to testify, would testify competently to the matters contained herein.

2. Dr. Saber states that "a POSITA would read the caption of Table 1 [of Tian] as indicating all images, including the lower resolution images, come 'from the Cohn-Kanade database'." Ex-2020, ¶5. This is different from my understanding of Tian. Tian's "Database" subsection discusses the "DFAT subset of Cohn-Kanade expression database" used by Tian's experiment. Ex-1005, 000005. This subsection states that "[t]he image sequences ... were digitized into 640x480 pixel arrays[,] " for which "[t]he size of head region is about 280x380 pixels[,] " without mentioning other resolutions. *Id.* Based on this paragraph, I understand Tian to obtain images of an original resolution of 640x480 from the Cohn-Kanade database.

3. Tian's Table 1 shows five images, respectively marked with "288x384 (Original)," "144x192," "72x96," "36x48," and "18x24." The caption of Table 1 is reproduced below:

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. Based on the images shown in Table 1 and the caption, I understand that the first image marked with “288x384 (Original)” is the head region of an example original image obtained from the Cohn-Kanade database. The other four images are head regions of reduced-resolution versions of this example’s original image that Tian creates by down-sampling.

4. To validate my understanding, I obtained a copy of the Cohn-Kanade database. I was previously familiar with the dataset while at the University of Maryland but no longer had access. Dr. Jeffrey Cohn lists the database on his personal website (<https://www.jeffcohn.net/Resources/>) under Databases, and provides an agreement form for its use. As instructed, I provided the form as a faculty member at the University at Buffalo, and received instructions to download it. The source and website from which I obtained the Cohn-Kanade database indicate that any reference to the database should include the citation “Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, 46-53.” Tian cites this paper as a reference [16] and states that “[m]ore details about the [Cohn-Kanade] database can be found at paper [16].” Ex-1005, 000005, 07. I reviewed the Cohn-

Kanade database and determined that images in the database have resolutions of 640x480 or 640x490. I did not find reduced-resolution images in the database. This is consistent with my reading of Tian and my recollection about the Cohn-Kanade database, which I encountered in earlier research. My review of the Cohn-Kanade database confirms my understanding that Tian obtains original images from this database and creates reduced-resolution images by down-sampling.

5. Regarding performing face detection and head pose estimation on reduced-resolution images, Dr. Saber argues that the additional processing requirements to reduce the resolution would outweigh the reduction in processing resources by analyzing reduced-resolution images. Ex-2020, ¶¶38-44. I disagree with Dr. Saber.

6. As explained in my original declaration, facial analysis is resource-consuming. Ex-1003, ¶69. For example, regardless of the specific method used (including those mentioned by Dr. Saber (Ex-2020, ¶44)), face detection and head pose estimation both require feature extraction over the image analyzed. This normally includes shifting a window (e.g., 5x5 pixels) across the entire image multiple times (e.g., in different directions or for identifying different types of features) and performing complex computations (e.g., multiple multiplications and summations) when the window is placed at each pixel position. Because complex

computations are performed for each pixel, significant processing resources are saved when facial analysis is performed on a reduced-resolution image.

7. I further explained in my original declaration that resolution-reduction techniques included “simply subsampling the image and dropping pixels.” Ex-1003, ¶15. This process merely involves reading values of selected pixels in an original image for use as pixel values of a subsampled image. Only the selected pixels are read and no mathematical calculation is performed to modify the pixel values. As a POSITA would have understood, this process requires minimal computation.

8. The amount of processing needed for reducing the resolution of an image is trivial when compared with the resource savings for performing facial analysis on reduced-resolution images. As explained in my original declaration, “[t]he processing resources and time can be reduced if the resolution of the image is reduced.” Ex-1003, ¶69. I further cited to background references to support this explanation. For example, Exhibit 1014 teaches a face detection approach including deriving an integral image from an acquired image and using the integral image for extracting features to be used in face detection. Ex-1014, 1:23-58, Fig. 2. Exhibit 1014 teaches subsampling an image before performing face detection on the subsampled image so that “calculations are avoided of a complete highest resolution integral image.” Ex-1014, 2:29-53. This “reduces processing overhead

for face detection[.]” *Id.* Exhibit 1014 would not have proposed this method if the processing resources for subsampling outweigh the savings of processing resources for performing face detection on the subsampled image.

9. Dr. Saber argues that antialiasing and interpolation are computationally expensive. Ex-2020, ¶¶40-43. However, antialiasing is not required for resolution reduction or facial analysis. For example, Tian teaches down-sampling and analyzing images without mentioning either antialiasing or the method of interpolation. In some situations these may help, but Tian showed that face detection can be accomplished, without noticeable reduction in performance. Tian teaches reducing the resolution by integer factors. For example, in Table 1, the 144x192 image is obtained by reducing the resolution of the original 288x384 image by a factor of four; the 72x96 image is obtained by reducing the resolution by another factor of four. Ex-1005, 000002. This can be accomplished by subsampling the images to pick one pixel out of every four. Even if considered, antialiasing and interpolation are significantly simpler than facial analysis tasks such as face detection and head pose estimation in terms of the computation involved.

10. Dr. Saber argues that selecting images suitable for face recognition based on head orientation determination “would decrease the chance of successful

facial recognition.” Ex-2020, ¶¶62. To support this argument, Dr. Saber provides an example:

Consider an audience measurement system that captures 10 consecutive frames of video of an audience. Even with each of the 10 frames only having a 25% chance of success (e.g., given various head movements between frames that may obscure certain features), performing facial recognition on all 10 frames would still be more likely to identify an audience member than only a single face recognition operation on a single frame that has a 50% chance of success.

Ex-2020, ¶¶61. Dr. Saber’s analysis of this example ignores that an unsuccessful face recognition results in identification of a wrong person. Performing face recognition on 10 frames each with a 25% chance of success would have resulted in more misidentifications (e.g., 75% chance) than correct identifications (e.g., 25% chance). Even if results from different images are aggregated, the probability of misidentifications would be very high. The performance of a system embodying this approach would likely be worse than a system that can correctly identify individuals half of the time.

11. Dr. Saber argues that “the starting resolution of the head region in an exemplary image captured by an audience measurement system is considerably lower than the resolution of Tian’s starting image” based on his discussion of two example images. Ex-2020, ¶¶55-59. Dr. Saber assumed that the example images have resolutions of 1280x960 and 1200x960, respectively. Ex-2020, ¶¶56-57. The

assumed resolutions are about 1.2 megapixels.¹ This assumption is unwarranted because cameras in 2011 were capable of capturing images of significantly higher resolutions (e.g., 25 megapixels). *E.g.*, Ex-1032, ¶186 (“Cameras are currently available at retail with 25 Megapixel resolution[.]”).

12. Here, I do not provide an opinion as to whether Dr. Saber’s method for estimating the resolution of head regions in the two examples is correct because Dr. Saber does not provide details about his method. Ex-2020, ¶¶55-59. But I note that Dr. Saber’s analysis would have reached different results if his assumption of image resolution was consistent with the state of the art. This is because the resolutions of the head regions scale proportionally with the resolutions of the images.

13. For his first example, Dr. Saber assumed an image resolution of 1280x960 and estimated a head region resolution of 60x60. Ex-2020, ¶56. However, if the image resolution is assumed to be 2560x1920 (~5 megapixels), the head region estimated under Dr. Saber’s method would have had a resolution of 120x120, well above the 36x48 resolution needed in Tian for high-accuracy face detection and head pose estimation.² If the image resolution is assumed to be

¹ An image of the resolution 1280x960 has $1280 \times 960 = 1.23$ million pixels. Because each megapixel is one million pixels, this image has a resolution of 1.23 megapixels.

² Dr. Saber claims that the minimum viable size for head pose estimation is 72x96. Ex-2020, ¶58. This is incorrect because Tian reports a 98% head pose estimation

5120x3840 (~20 megapixels), the head region estimated under Dr. Saber's method would have had a resolution of 240x240, close to Tian's original resolution of 288x384. The resolutions of 120x120 and 240x240, after being reduced by a factor of four for example, becomes 60x60 and 120x120, both of which are suitable for face detection and head orientation determination under Tian. The same analysis applies to Dr. Saber's second example.

14. Patent Owner argues that paragraph 7 of Lu-577 disparages the approach of Lert. POR, 43. This is incorrect. Paragraph 7 states “[c]hanges in the methods of measuring the reception of television or radio programming are required because of a planned change-over from analog to digital broadcasting.” Ex-1004, ¶7. Lert teaches identifying a television program by extracting an audio signature. Ex-1009, 3:57-4:42, 4:61-5:19, Figs. 1-2. Lu-577's paragraph 7 does not mention Lert's approach and does not indicate that Lert's signature-based approach is affected by the change-over from analog to digital broadcasting. Ex-1004, ¶7.

15. I declare that all statements made herein of my knowledge are true, 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

accuracy for 18x24 and a 100% face detection accuracy for 36x48. Ex-1005, 000005.

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 July 23, 2024

By:



Dr. David Scott Doermann

**Appendix 1: Additional Materials Considered
in the Preparation of this Declaration**

Paper / Exhibit No.	Description
Paper 17	Patent Owner's Response
Ex. 2020	Supplemental Declaration of Dr. Eli Saber
Ex. 1031	Takeo Kanade, Jeffrey F. Cohn, & Yingli Tian, <i>Comprehensive Database for Facial Expression Analysis</i> , Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 2000, pp. 46-53.
Ex. 1032	U.S. Patent Application Publication No. 2012/0159996 ("Sutton")