

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

PETITIONER’S ORAL HEARING DEMONSTRATIVES

**TVISION INSIGHTS, INC.,
Petitioner,
v.
THE NIELSEN COMPANY (US), LLC,
Patent Owner.**

Case No. IPR2023-01014
U.S. Patent 11,470,243

Petitioner's Demonstratives
October 16, 2024

Instituted Grounds

Grounds	Claims	Statutory Bias	Prior Art
1	4-6, 8, 11-14, and 18-20	§ 103	Lu-577 and Tian
3	4-6, 8, 11-14, and 18-20	§ 103	Nielsen-372, Steinberg, and Tian

'243 Patent – Claim 1

- [1p].** 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.

'243 Patent – Claims 4-6

[4p]. 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.

[6p]. 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;

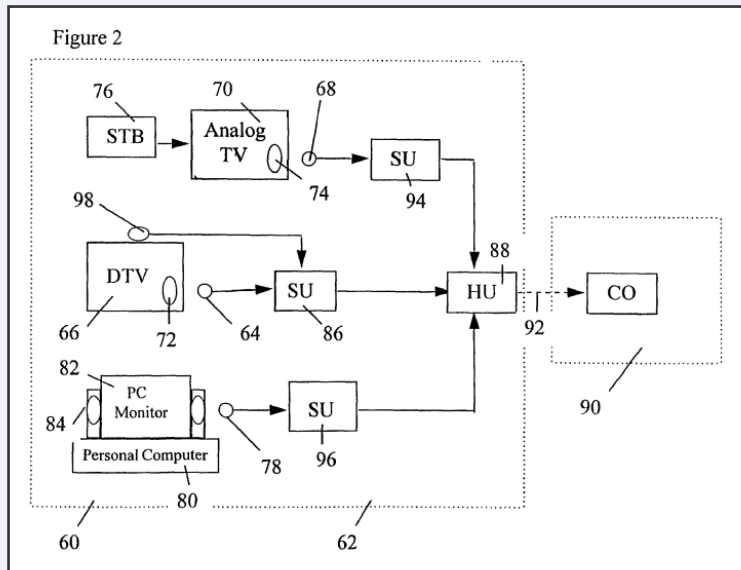
[6b] analyze a portion corresponding to the identified region of a second image; and based on the analysis of the region of the second image, determine that the head matches the known person.

Ground 1

Lu-577

- **Audience measurement system:**
 - Identifying TV program based on audio
 - Identifying audience based on head detection and face recognition

Figure 2



Tian

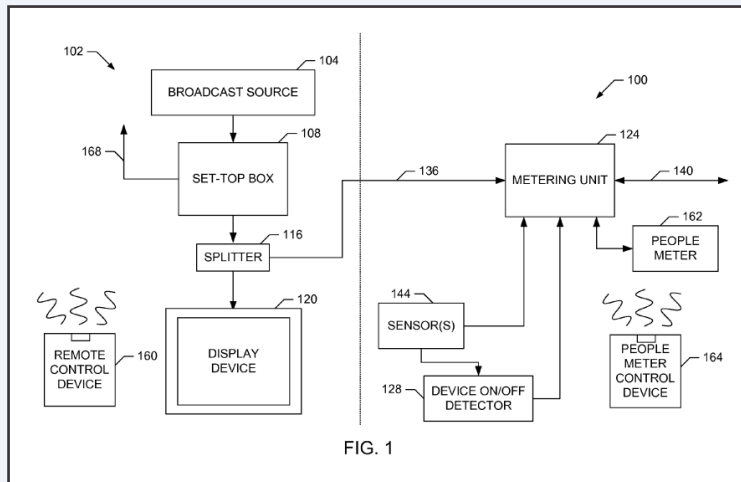
- **Techniques for improving facial analysis:**
 - Reducing image resolution and analyzing reduced-resolution images
 - Head pose estimation
 - Face tracking across a sequence of images

Face Process						
		288x384 (Original)	144x192	72x96	36x48	18x24
Face	FD	100%	100%	100%	100%	0%
Acquisition	HPE	98.5%	98%	98.2%	97.8%	98%

Ground 3

Nielsen-372

- **Audience measurement system:**
 - Identifying TV program based on audio
 - Identifying audience



Steinberg

- **Facial analysis techniques:**
 - Face detection
 - Face orientation determination
 - Face recognition

Tian

- **Techniques for improving facial analysis:**
 - Facial analysis on sequences of images
 - Face tracking
 - Reducing image resolution and analyzing reduced-resolution images

Remaining Issues in Dispute

Remaining issues in dispute



Tian teaches “reduc[ing] a resolution”

Motivation to combine

Motivation to apply Tian’s teachings of resolution reduction and head orientation determination

Motivation to combine Nielsen-372 with Steinberg

Grounds 1 and 3 teach face recognition on corresponding region/portion of second image

Tian teaches resolution reduction by down-sampling

Evaluation of Face Resolution for Expression Analysis

Ying-li Tian

IBM T. J. Watson Research Center,
PO Box 704, Yorktown Heights, NY 10598
Email: yltian@us.ibm.com

Abstract

Most automatic facial expression analysis (AFE) systems attempt to recognize facial expressions from data collected in a highly controlled environment with very high resolution frontal faces (face regions greater than 200 x 200 pixels). However, in real environments, the face image is often in lower resolution and with head motion. It is unclear that the performance of AFE systems for low resolution face images. The general approach to AFE consists of 3 steps: face acquisition, facial feature extraction, and facial expression recognition. 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. A total of five different resolutions of the head region are studied (288x384, 144x192, 72x96, 36x48, and 18x24) based on a widely used public database [16]. The lower resolution images are down-sampled from the originals.

1. Introduction

Facial expression analysis includes both measurement of facial motion and recognition of expression. The general approach to Automatic Facial Expression Analysis (AFE) systems consists of 3 steps (see Figure 1): face acquisition, facial feature extraction and representation, and facial expression recognition.

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. In order to handle large head motion, head finding, head tracking and pose estimation can be applied to a facial expression analysis system.

After the face is located, the next step is to extract and represent the facial changes caused by facial expressions. In facial feature extraction for expression analysis, there are mainly two types of approaches: geometric feature-based methods and appearance-based methods. The geometric facial features present the shape and locations of fa-



Figure 1: Basic Systems.

cial components (i.e., facial features). The facial components to form a feature vector in appearance-based wavelets, are applied regions in a face image on the different effects of in-plane faces can be eliminated the feature extraction step of expression analysis.

Facial expression systems to identify system (FACS) actions expressions.

Many recent advances achieved (2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100). Their system can deal with limited out-of-business and accurate face-component minimized 16 of the 30 combinations. Recognition [6, 20] attempted a variation (in the upper part) by using same database [13]. In that database, subjects

In this paper, we investigate the effects of different image resolutions for each step of facial expression analysis. 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. As shown in Figure

Undisputed teachings of Tian satisfy “reduc[ing] a resolution”

The **Institution Decision** recognizes plain teachings of Tian:

At least at this preliminary stage of these proceedings, we determine that Petitioner’s showing that the cited art teaches the “reducing the resolution” limitation as recited in claim 4 and similarly recited in claims 11 and 18 to be **compelling**.¹⁹ Tian **plainly** teaches obtaining reduced-resolution images and teaches **down-sampling** as shown in the Petition. *See* Pet. 26.

Patent Owner acknowledged that Tian teaches **“resolution reduction”**:

system and method. *Supp. Dec.*, 12-14. First, *Tian* relies on pre-processing for resolution reduction. *Tian*, 000002 (characterizing Table 1 as “[a]n example in five

No dispute that resolution-reduction techniques were known by 2011



Dr. Doermann



Dr. Saber

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.

Q. By 2011, would a POSITA have been capable of using these methods to reduce the resolution of the image?

A. Yes.

Q. By 2011, would a POSITA have been able to use downsampling to reduce the resolution of the image?

A. Yes. Again, I mean, it depends on the application that you are apply -- working with and whether -- you know, the requirements for realtime and not realtime. There are parameters that you have -- that has to be considered. But generally speaking, a POSITA would be familiar with downsampling.

Q. Would a POSITA have been able to implement downsampling using software running on a computer?

A. Again, depending on the application and the parameters of the application. But generally speaking, a POSITA would be able to implement downsampling using software on a computer.

Nielsen's shifting arguments about Tian

POR: Tian does not create down-sampled images



(*emphasis added*). Petitioner's characterization of Tian's language is wrong, as *Tian* does not *create* the down-sampled images; it just accesses existing down-sampled images. *Dec 48* (Dr. Saber opining that a POSITA would not understand that *Tian*

Sur-Reply: Tian is ambiguous



perform resolution reduction. *Cf. Ex. 2020*, ¶ 8. Petitioner does not point to anything in Tian that would **preclude** Tian from obtaining reduced-resolution images from an alternative, third-party source. *GMG*, 25 (finding the art was ambiguous in part because there was “no preclusion” of an alternative).

PO seeks to deny Tian's clear teachings based on its characterization of the Cohn-Kanade Database

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

TVISION INSIGHTS

Petitioner,

v.

THE NIELSEN COMPANY

Patent Owner

Case IPR2023-010

Patent No. 11,470,

'243 Patent, 15:56-58, 16:49-51, 17:41-43. *Tian* does not teach actually reducing the resolution of the images, much less circuitry, instructions, or a device to do so. Instead, *Tian* discloses a pre-existing “public database” of images that “are down-sampled” versions of the original images. Dec. 46, 48. *Tian* fails to disclose the

PATENT OWNER'S RESPONSE TO PETITION
FOR INTER PARTES REVIEW

POR:

Tian obtains down-sampled images from Cohn-Kanade database

Reply:

- **Tian's "Database" subsection**: image sequences from the database digitized into the same resolution
- **Tian's Table 1**: highest resolution image marked "Original"
- **Paper [16] (Ex. 1031)**: images sequences in Cohn-Kanade database digitized into original resolution
- **Dr. Doermann's confirmation**: Cohn-Kanade database did not include reduced resolution images

Sur-Reply:



PO's argument contradicted by Tian's description of the Cohn-Kanade Database

Tian

3.1. Database

The DFAT subset of Cohn-Kanade expression database [16] is used for our experiments. 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. The size of head region is about 280x380 pixels. More details about the database can be found at paper [16].



Dr. Doermann

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.



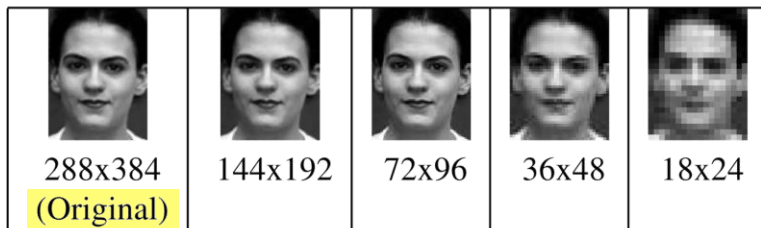
Dr. Saber

54. Per Tian, the original images in the database are at a full resolution of 640x480 pixels. Tian, 000005. The “size of the head region” within these original images, as used by Tian, is 288x384. Tian, 000005, Table 3 (showing the “original” and highest-resolution image with a head region of 288x384). Thus, the head portion

Statements quoted by PO are consistent with Tian's teachings of resolution reduction

Tian

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.



"example"

**"in five different
resolutions"**

**"from Cohn-
Kanade database"**



Dr. Doermann

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.

Tian's teachings confirmed by paper describing the Cohn-Kanade database

Tian

3.1. Database

The DFAT subset of Cohn-Kanade expression database [16] is used for our experiments. 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. The size of head region is about 280x380 pixels. More details about the database can be found at [paper \[16\]](#).

Paper [16]: "Comprehensive Database for Facial Expression Analysis"

Image sequences for frontal views were digitized into either 640x490 or 640x480 pixel arrays with 8-bit gray-scale or 24-bit color values. To date, 1917

frame rate omitting odd fields. The image data may be transformed or down-sampled to examine the effects of reduced resolution, sampling rate, image compression, and luminance.

The Cohn-Kanade database was known to not include reduced-resolution images



Dr. Doermann



Dr. Saber

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.

Q. In your opinion, the Cohn-Kanade database would have been known to a POSITA in 2011, correct.

A. A POSITA that would be working in this area would be familiar with, obviously, the literature in this area so would know about this facial -- a POSITA working in facial expression analysis would know about the database.

Dr. Doermann obtained the Cohn-Kanade database as used by Tian



Dr. Doermann

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-

Dr. Saber did not bother to confirm if the Cohn-Kanade database included reduced-resolution images



Dr. Saber

Q. Did you in any way confirm your understanding that the Cohn-Kanade database include the images in five different resolutions?

A. **I didn't need to.** Tian clearly states in Table 1 -- again, I'm reading Tian as a POSITA and responding to your petition, the prior art, and the four corners of Tian.

Tian states "An example in five different resolutions from Cohn-Kanade database." So a POSITA reading Tian would not understand Tian to teach reducing a resolution.

Remaining issues in dispute

Tian teaches “reduc[ing] a resolution”

Motivation to combine



Motivation to apply Tian's teachings of resolution reduction and head orientation determination

Motivation to combine Nielsen-372 with Steinberg

Grounds 1 and 3 teach face recognition on corresponding region/portion of second image

TVision's obviousness showings for combining Lu-577 and Tian

1. Lu-000 and Tian are similarly directed to image processing and facial analysis.	UNDISPUTED
2. Lu-577 explicitly invites employing "known head location and face recognition software."	UNDISPUTED
3. Applying Tian's teaching of performing facial analysis on a sequence of images can improve the confidence level of person identification and allow the selective performance of face recognition on suitable images.	UNDISPUTED
4. The combination involves an application of known techniques (e.g., Tian's facial analysis techniques) to a known system (e.g., Lu-577's audience measurement system) and method ready for improvement. Doing so would have yielded predictable results at least because Tian provides experimental assessments of its techniques.	UNDISPUTED
5. The combination simply uses Tian's teachings in the manner taught by Tian to implement functionalities in Lu-577.	UNDISPUTED
6. Applying Tian's face tracking saves processing resources and time by reducing the times face detection needs to be applied.	DISPUTE ABANDONED IN SUR-REPLY
7. Applying Tian's teaching of analyzing reduced-resolution images saves processing resources for face detection and post estimation without substantially reducing their accuracy.	AT ISSUE
8. Applying Tian's head pose estimation allows improving the performance of face recognition by determining whether an image is suitable for face recognition and determining whether a person is watching the television.	AT ISSUE

Motivation to Analyze Reduced-Resolution Images

Dr. Doermann: face detection and head pose estimation on reduced-resolution images save processing resources

Petition

1003, ¶68. A POSITA would have found it obvious to apply Tian's teaching to Lu-577's system to perform facial analysis on reduced-resolution images. Ex-1003, ¶69. Doing so can reduce the time and resource consumption of tasks such as face detection and pose estimation, without substantially reducing the accuracy of these tasks. Ex-1013, 000001; Ex-1014, 2:43-53; Ex-1015, ¶0022; Ex-1005,



Dr. Doermann

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,

PO's Alleged Error 1:



Petitioner has not compared cost of resolution reduction plus cost of facial analysis on reduced-resolution image with cost of facial analysis on nonreduced-resolution image

Dr. Doermann did the exact comparison that PO asks for



Dr. Doermann

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,

Face analysis is a complex process that includes various steps

PO's Exhibit 2034

Face detection algorithms usually share common steps. Firstly, some data dimension reduction is done, in order to achieve a admissible response time. Some pre-processing could also be done to adapt the input image to the algorithm prerequisites. Then, some algorithms analyze the image as it is, and some others try to extract certain relevant facial regions. The next phase usually involves extracting facial features or measurements. These will then be weighted, evaluated or compared to decide if there is a face and where is it. Finally, some algorithms have a learning routine and they include new data to their models.



Dr. Saber

Q. Do you agree that the steps mentioned in the paragraph are common steps for face detection algorithms?

A. Well, generally speaking, I mean, from a general perspective, yeah. That's -- they seem reasonable.

Dr. Doermann: facial analysis (e.g., the feature extraction step) includes complex computations



Dr. Doermann

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.

Dr. Saber admits that feature extraction for face detection and orientation determination includes complex computations



Dr. Saber

A simple "feature extraction method" includes sliding a kernel "across the whole image" and computing convolution, which requires "25 multiplications and 24 additions for every pixel."



Pixel Values:

$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	$p_{1,4}$	$p_{1,5}$
$p_{2,1}$	$p_{2,2}$	$p_{2,3}$	$p_{2,4}$	$p_{2,5}$
$p_{3,1}$	$p_{3,2}$	$p_{3,3}$	$p_{3,4}$	$p_{3,5}$
$p_{4,1}$	$p_{4,2}$	$p_{4,3}$	$p_{4,4}$	$p_{4,5}$
$p_{5,1}$	$p_{5,2}$	$p_{5,3}$	$p_{5,4}$	$p_{5,5}$

Kernel:

$k_{1,1}$	$k_{1,2}$	$k_{1,3}$	$k_{1,4}$	$k_{1,5}$
$k_{2,1}$	$k_{2,2}$	$k_{2,3}$	$k_{2,4}$	$k_{2,5}$
$k_{3,1}$	$k_{3,2}$	$k_{3,3}$	$k_{3,4}$	$k_{3,5}$
$k_{4,1}$	$k_{4,2}$	$k_{4,3}$	$k_{4,4}$	$k_{4,5}$
$k_{5,1}$	$k_{5,2}$	$k_{5,3}$	$k_{5,4}$	$k_{5,5}$

Calculations for each pixel:

$$\begin{aligned} & k_{1,1} \times p_{1,1} + k_{1,2} \times p_{1,2} + k_{1,3} \times p_{1,3} + k_{1,4} \times p_{1,4} + k_{1,5} \times p_{1,5} \\ & + k_{2,1} \times p_{2,1} + k_{2,2} \times p_{2,2} + k_{2,3} \times p_{2,3} + k_{2,4} \times p_{2,4} + k_{2,5} \times p_{2,5} \\ & + k_{3,1} \times p_{3,1} + k_{3,2} \times p_{3,2} + k_{3,3} \times p_{3,3} + k_{3,4} \times p_{3,4} + k_{3,5} \times p_{3,5} \\ & + k_{4,1} \times p_{4,1} + k_{4,2} \times p_{4,2} + k_{4,3} \times p_{4,3} + k_{4,4} \times p_{4,4} + k_{4,5} \times p_{4,5} \\ & + k_{5,1} \times p_{5,1} + k_{5,2} \times p_{5,2} + k_{5,3} \times p_{5,3} + k_{5,4} \times p_{5,4} + k_{5,5} \times p_{5,5} \end{aligned}$$

Dr. Saber admits that performing facial analysis on fewer pixels significantly saves processing resources



Dr. Saber

Q. In performing that computation, is that amount of computation proportional to the number of pixels in the image?

A. Well, maybe you didn't understand my answer correctly.

So as I'm taking the three-by-three window across the entire image -- so if the image is 512-by-512, it would be so many operations. If the image is 1024-by-1024, obviously it's going to be a lot more operations because it's a bigger image. So it's related to the size of the image and the size of the kernel.

Dr. Doermann: resolution reduction requires minimal computation



Dr. Doermann

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.

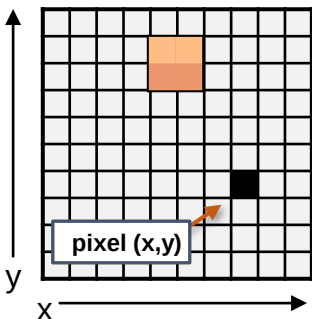
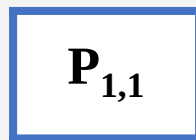
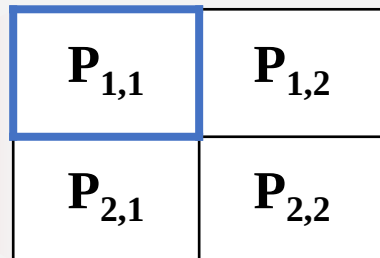
Dr. Saber's example down-sampling method selects 1/4 of the pixels and includes minimal computations



Dr. Saber

[Y]ou could pick a pixel out of every four, ...and obviously downsample the image.

Generally speaking, it's fairly simple for you to skip every other pixel in the horizontal and vertical direction and just select one.



Calculations for each pixel:

$$\begin{aligned} & k_{1,1} \times p_{1,1} + k_{1,2} \times p_{1,2} + k_{1,3} \times p_{1,3} + k_{1,4} \times p_{1,4} + k_{1,5} \times p_{1,5} \\ & + k_{2,1} \times p_{2,1} + k_{2,2} \times p_{2,2} + k_{2,3} \times p_{2,3} + k_{2,4} \times p_{2,4} + k_{2,5} \times p_{2,5} \\ & + k_{3,1} \times p_{3,1} + k_{3,2} \times p_{3,2} + k_{3,3} \times p_{3,3} + k_{3,4} \times p_{3,4} + k_{3,5} \times p_{3,5} \\ & + k_{4,1} \times p_{4,1} + k_{4,2} \times p_{4,2} + k_{4,3} \times p_{4,3} + k_{4,4} \times p_{4,4} + k_{4,5} \times p_{4,5} \\ & + k_{5,1} \times p_{5,1} + k_{5,2} \times p_{5,2} + k_{5,3} \times p_{5,3} + k_{5,4} \times p_{5,4} + k_{5,5} \times p_{5,5} \end{aligned}$$

Dr. Saber admits that multiple resolution reduction methods are all simpler than facial analysis



Transcript of Dr. Eli Saber

Date: July 2, 2024
Case: TVision Insights, Inc. -v- The Nielsen Company (US), LLC (PTAB)

Planet Depos
Phone: (502) 233-2345
Email: transcripts@planetdepos.com
www.planetdepos.com

WORLDWIDE COURT REPORTING & LITIGATION TECHNOLOGY

IPROBES-0514 TVision Insights, Inc. -v- The Nielsen Company (US), LLC



Dr. Saber

A. So if you're saying **downsampling** is just picking one out of every four pixels -- that's downsampling -- that would be relatively **easier than** doing **face detection**.

A. So if **pixel binning** is defined as the addition of four pixels and computing the sum, then that would be relatively **easier than** a **face detection** algorithm, generally speaking.

A. But generally speaking, **bilinear interpolation** would be relatively **easier to do than face detection**, generally speaking.

A. But **face orientation determination** would be, relatively speaking, **more complex than subsampling**.

A. But **pixel binning**, if you're just summing four pixels, it would be relatively **easier computationally-wise than face orientation detection**.

A. But **bilinear interpolation** would probably relatively be easier -- if it's properly implemented, it would probably be easily -- **easier than face orientation determination**.

Even if there is a tradeoff between resource savings and costs, it does not obviate the motivation to combine



“It’s not necessary to show that a combination is the best option, only that it be a suitable option.”

Intel Corp. v. Qualcomm Inc., 21 F.4th 784, 800 (Fed. Cir. 2021)



“A given course of action often has simultaneous advantages and disadvantages, and this does not necessarily obviate motivation to combine.”

Medichem, S.A. v. Rolabo, S.L., 437 F.3d 1157, 1165 (Fed. Cir. 2006)

PO's Alleged Error 2:



Petitioner does not analyze the computational cost of antialiasing

PO's antialiasing argument contradicts its own expert

PO

No reasonable expectation of success from forgoing antialiasing of down-sampled image



Dr. Saber

- '243 patent does not claim antialiasing
- '243 patent discloses resolution reduction without antialiasing
- Tian does not perform antialiasing

The '243 patent does not claim or otherwise mention antialiasing

The '243 patent does not claim antialiasing

- Q. In your view, can Claim 4 be met without performing anti-aliasing?
- A. Well, if you reduce the resolution, you've met the claim element. There are obviously different ways of doing the resolution. The resolution would be - have better quality how you've done it. But if you reduce the resolution, you've met the claim element, on a first image.

The '243 patent discloses a resolution-reduction method (e.g., pixel binning) without a separate antialiasing step

- Q. If you reduce the resolution of the image using binning, you do not need a separate anti-aliasing step, correct?
- A. Well, they're not utilizing a separate anti-aliasing step here.



Dr. Saber

Tian's experimental results show high accuracy achieved without antialiasing

Dr. Saber: antialiasing not performed on Tian's images

20. Further, *Tian's* disclosure of using images that are only down-sampled, without any other disclosed pre-processing steps, would not result in a useable audience measurement system, because the down-sampled image would not be ready for analysis. Down-sampling is vulnerable to artifacts resulting from aliasing.

Yet, Tian demonstrates high accuracy for face detection and head orientation determination on reduced-resolution images

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%

PO's Alleged Error 3:



Petitioner improperly analyzes the cost for facial analysis using convolution, but not Tian's neural-network approach

Dr. Saber: neural network approach includes feature extraction and convolution

A simple "feature extraction method" includes sliding a kernel "across the whole image" and computing convolution, which requires "25 multiplications and 24 additions for every pixel."



Dr. Saber

- Q. Do you know of any head orientation determination method that does not depend on feature extraction?
- A. Well, neural network-type techniques usually have a different type of approach to them. They still depend on some kind of features, but they're generally a little different where you have potentially -- if you're doing a supervised type of thing, you've shown the network some input, the corresponding output, and you have essentially tried to converge the network to give you that kind of thing, it still has to depend on features, in a sense. You need to have some type of features to start with, whether it's the pixel gray levels themselves or edges or corners or whatever. But you need to have some type of features. Different algorithms have different ways of analyzing those features.

PO's Alleged Error 4:



Petitioner does not account for the fact that the resolution-reduction process would be performed more often than the head-orientation determination process could be skipped

Savings of computational resources achieved by analyzing one reduced-resolution image



Dr. Doermann

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,

Claim 4 does not require reducing the resolution of multiple images.

4. The audience measurement system of claim 1, wherein the processor circuitry is to:
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
determine the orientation of the head with respect to the camera based on the reduced-resolution image.

Motivation to Determination Head Orientation

Selectively performing face recognition based on face orientation determination improves performance

Petition

apply this teaching to Lu-577's system. Ex-1003, ¶65 A POSITA would have found it obvious to determine a person's head pose to determine whether an image is suitable for face recognition and understood that this improves the performance of face recognition. Ex-1005, 000002; Ex-1010, ¶0003; Ex-1021, 000042; Ex-



Dr. Doermann

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.

PO ignores that certain head orientations are not suitable for face recognition



PO: selectively performing face recognition undermines performance by reducing face recognition opportunities



Dr. Doermann

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. ... In fact, Tian teaches performing facial expression analysis only for frontal or near frontal view faces.



Dr. Saber

Q. What orientations of the face are not suitable for performing face recognition?

A. Do you mean orientation of the head?

Q. Yes

A. So if I'm looking backwards, obviously it's not going to be suitable for face recognition.

Nielsen ignores that unsuccessful face recognitions undermine performance



Nielsen's Example: For example, performing facial recognition on 10 consecutive frames, each with a 25% chance of success, is significantly more likely to identify an audience member than performing face recognition on a single frame with a 50% chance of success.



Dr. Doermann: unsuccessful face recognition results in identifying a wrong person

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.



Dr. Saber: unsuccessful facial recognition refers to misidentification

- Q. So when facial recognition is unsuccessful, it is possible that no face is recognized, or it is possible that a wrong person is identified. Is that correct?
- A. So assuming you detected the face correctly and now you're trying to recognize it. So it could be that the face doesn't exist in your database. So it's a new face you've never seen, so the algorithm will not be successful in finding the face in the database. That doesn't mean it's not successful. It just says -- it's been successful because it didn't match to anything in the database. So it's actually told you that this face doesn't exist. So that was a successful scenario. If it matches Tom to John, then it's an unsuccessful scenario unless John is a twin of Tom and the two are identical twins.

Pose estimation facilitates determining if a person is watching television

Dr. Doermann: pose estimation allows determining whether a person is watching television

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

Dr. Saber: Tian's frontal view indicates that the person "is looking directly toward the camera[.]"

Frontal or
near frontal



Tian, Table 2 (annotated). Though these three images are placed in the **same pose category**, one image is looking directly toward the camera (blue) and the other images are looking to the left of the camera (red). This distinction indicates *Tian*

Remaining issues in dispute

Tian teaches “reduc[ing] a resolution”

Motivation to combine

Motivation to apply Tian’s teachings of resolution reduction and head orientation determination



Motivation to combine Nielsen-372 with Steinberg

Grounds 1 and 3 teach face recognition on corresponding region/portion of second image

The combination applies two aspects of Steinberg's teachings to Nielsen-372

1) using a camera to capture images of individuals

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).

2) "approach" of identifying individuals by detecting and recognizing faces

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

The Petition relies on Steinberg's teaching about obtaining images by a camera

Institution Decision:

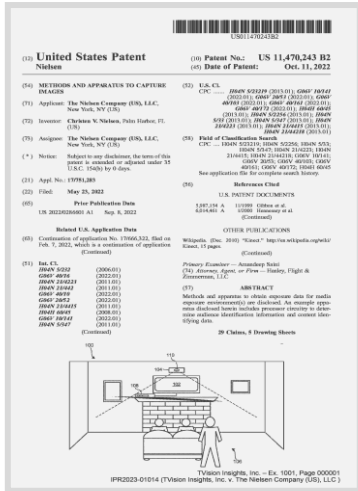
camera” (*id.* at 17:22–25) (emphasis added). For each of these limitations in the independent claims that contains the first recitation of “a camera” in the claims and that serves as the antecedent basis for the later references to “the camera” in the claims, Petitioner cites teachings in Steinberg, not Nielsen-372. *See* Pet. 53, 65, 69. In this regard, the Petition states:

Petition:

The combination teaches limitation [1d]. Ex-1003, ¶¶227-235. Steinberg teaches **obtaining a sequence of images by a camera** (e.g., digital camera, image capture device). Ex-1007, ¶0010. Steinberg’s image capture device (e.g., a digital camera) captures images and handles certain image processing operations. Ex-1007, ¶¶0128-0129, FIGs. 1(b)-1(c); Ex-1003, ¶227.

Taking photos of individuals using a camera is undisputedly a prior-art feature

Background of the '243 patent:



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. The calculated amount of people in the media exposure



Dr. Saber

Q. By 2011, was it common to use camera to take photos of the audience?

A. Yes.

Basis of PO's mischaracterization: Steinberg's "approach" implemented on Nielsen-372's local metering system

Sentences relied on by PO:

- "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. 1003, ¶ 199

- "Nielsen-372's local metering system includes a camera....A POSITA would have found it obvious to implement [Steinberg's approach] on either the computer or the camera of Nielsen-372[.]"

Id., ¶ 200

- "A POSITA would have had a reasonable expectation of success...given that Nielsen-372 provide[s] the hardware architecture for performing Steinberg's face recognition method.... [T]he combination simply implements Steinberg's teachings about facial analysis [] in Nielsen-372's system."

Id., ¶ 203

Steinberg's "approach" refers to face detection, normalization, and recognition, not image capturing



Dr. Doermann

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

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

PO: physically incorporating Steinberg's teachings to Nielsen-372 creates problems

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

TVISION INSIGHTS,

Petitioner,

v.

THE NIELSEN COMPANY

Patent Owner

Case IPR2023-010
Patent No. 11,470,2

Petitioner's proposed modification would require the ON/OFF device camera ... to face the audience to recognize the faces of the audience members using "Steinberg's methods." ... With the camera facing the audience, the Nielsen-372 camera would no longer be positioned to take images of the screen to detect when the display is ON or OFF.

PATENT OWNER'S RESPONSE TO PETITION
FOR *INTER PARTES* REVIEW

The combination does not create PO's alleged problems



Dr. Saber

Q. By 2011, would a POSITA have been able to connect two cameras to a computer where the two camera pointed different directions?

A. ... Yes.

Q. By 2011, were there camera with multiple image sensors that can capture images in different directions?

A. I would assume yes ...

Q. What is the widest angle the fisheye camera would have been capturing back then?

A. Well, I mean, the specifications typically specify the angle that you can capture. And some of them have, like, two sensors on them that you can capture both the front and the back ...

Nielsen-372 does not require the camera to point at the television because the video sensor is optional

Claim 1 of Nielsen-372 recites determining a device's ON/OFF status without a camera:

1. A method to determine whether a presentation device is ON or OFF, the method comprising:
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; and
determining a result indicating whether the presentation device is ON or OFF based on whether the **audio signature** matches any of the set of reference signatures.

(19) United States

(12) Patent Application Publication
Nielsen et al.

(10) Pub. No.: US 2010/0274372 A1
(43) Pub. Date: Oct. 28, 2010

(54) DISPLAY DEVICE ON/OFF DETECTION

(76) Inventors: Christian V. Nielsen, Palm Harbor, FL (US); Dan Nelson, Tampa, FL (US)

Correspondence Address:
Husley, Flight & Zimmerman, LLC
150 S. Wacker Dr. Suite 2100
Chicago, IL 60606 (US)

(21) Appl. No.: 12/831,870

(22) Filed: Jul. 7, 2010

Related U.S. Application Data

(63) Continuation of application No. 11/576,328, filed on Mar. 29, 2007, filed as application No. PCT/US2006/031969 on Aug. 16, 2006.

(60) Provisional application No. 60/708,557, filed on Aug. 16, 2005; provisional application No. 60/761,678, filed on Jan. 24, 2006.

Publication Classification

(51) Int. Cl. G06F 7/00 (2006.01)

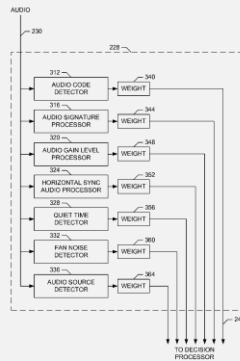
G06V 5/02 (2006.01)

(52) U.S. Cl. 709/94; 706/54

(57)

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, and determining a result indicating whether the presentation device is ON or OFF based on whether the audio signature matches any of the set of reference signatures.



TVision Insights, Inc. - Ex. 1006, Page 000001
IPR2023-01014 (TVision Insights, Inc. v. The Nielsen Company (US), LLC)

“Voluntary audience measurement” is not a principle of operation of Nielsen-372



PO: changing from voluntary audience measurement to automatic audience measurement would change the principle of operation

Nielsen-372:

[0048] The example local metering system 100 of FIG. 1 also includes a people meter 162 to capture information about the audience. The example people meter 162 may be configured to receive information from a people meter control device 164 having a set of input keys, each assigned to represent a single viewer. The people meter 162 may prompt the audience members to indicate that they are present in the viewing audience by pressing the appropriate input key on the people meter control device 164. The people meter 162 may

Petition:

accuracy of data. Ex-1006, ¶0048. Applying Steinberg’s method to Nielsen-372 would have allowed identification of the audience members without interruptions to the entertainment experience and with improved accuracy. A POSITA would have been motivated to do so because these benefits are consistent with Nielsen-372’s goal of collecting viewing information. Ex-1003, ¶201.

Both experts agree applying Steinberg's teaching to Nielsen-372 would improve the accuracy of audience measurement



Dr. Doermann



Dr. Saber

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

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

Remaining issues in dispute

Tian teaches “reduc[ing] a resolution”

Motivation to combine

Motivation to apply Tian’s teachings of resolution reduction and head orientation determination

Motivation to combine Nielsen-372 with Steinberg



Grounds 1 and 3 teach face recognition on corresponding region/portion of second image

Limitations at issue: face recognition on region/portion of second image corresponding to location/region of head in first image

Claim 5

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.

Claim 6

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

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

Nielsen seeks to improperly import a negative limitation into the claims



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, *wherein no determination of head location is performed on the second image*, and (ii) comparing the generated facial signature to a database of facial signatures.



POR:

Tian updates (“tracks”) the location of the head/face for each frame to perform face analysis for that frame. Regardless of whether *Tian* detects the head/face location in each frame or tracks the movement of the head/face in the current frame based on prior frames, *Tian* is locating and conducting facial analysis of a single frame (the “single-frame process”). This is in contrast to the claimed invention that determines a head/face location in the first frame and then uses that location in the second frame to perform facial recognition (the “two-frame process”).

Dr. Saber disparages PO's negative limitation that prohibits confirming head location in second image



Dr. Saber

73. Regarding drawback (3), a POSITA would understand performing facial recognition in a region of an image where the head was **in a prior frame** decreases the accuracy of facial recognition performance when compared to applying recognition in a region where the head **is located in the current frame**. For example, a POSITA would appreciate **a person could move their head between image captures in a sequence, and thereby render the previous location useless in** recognizing the face in the current image. Indeed, *Tian* discloses the accuracy of

Undisputed disclosures satisfy the limitations at issue

UNDISPUTED



Lu-577/Steinberg teaches performing face recognition on a region of an image corresponding to a face/head location.



Tian teaches determining a face/head location in a first image and using this location for subsequent images unless a change is detected.



Face/head location often does not change across a sequence of images taken in a short period of time.



Tian teaches performing face detection on a reduced-resolution image.

Obvious to determine a person's face/head location in a first image, whose resolution may be reduced, and perform face recognition on a region of this location in a second image.

Tian's face tracking confirms correspondence between region in second image and head location in first image

Tian

and absent. Given an image sequence, the region of the face and approximate location of individual face features are detected automatically in the initial frame. The contours of the face features and components then are adjusted manually in the initial frame. Then, all face feature changes are automatically tracked in the image sequence. Details of geometric feature extraction and representation can be found in paper [27].



Dr. Doermann

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, ... image sequence.”).* Under this approach, if no change is detected, the face is determined to be at the same location in subsequent images.

By using face location from a previous image, face tracking reduces processing resources for face detection

Ex. 1010 Bigioi

[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.



Dr. Doermann

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

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.

Rather than reading out “corresponding to,” Tian’s teachings confirm when the correspondence exists



PO: Petitioner’s interpretation reads out claims 5-6’s “corresponding to” limitations.



Dr. Doermann

Under [Tian’s] approach, **if no change is detected**, the face is determined to be at the same location in subsequent images.

[W]hen 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.



[C]ombinations of prior art that sometimes meet the claim elements are sufficient to show obviousness.

Unwired Planet, LLC v. Google Inc., 841 F.3d 995, 1002 (Fed. Cir. 2016)

Other issues

The field of the '243 patent is media analysis



Dr. Doermann

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).

Nielsen's proposed field of endeavor is contradicted by its own expert



Nielsen argues that the field of endeavor of the '243 patent is "audience measurement"

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



Dr. Saber

Q. Why do you believe you qualify as a POSITA regarding the '243 patent?

A. Well, as I mentioned, I've been working in this field for over 30 years. I've worked with colleagues. I've worked with students. I've worked with engineers during my time at the Xerox Corporation. So I've interacted with colleagues, students, engineers, other practitioners in the field. So I'm well aware of the knowledge and expertise of a POSITA.

Q. You mentioned you have worked in this field for over 30 years. **Which field are you referring to?**

A. The digital image -- digital -- well, typically we refer to it as signal processing as an umbrella, but basically the **digital image/video processing, computer vision, pattern recognition, machine learning**. I've been working in this field for -- or these fields for over 30 years now.

Audience measurement is an application of media analysis



Dr. Saber

Q. How is image processing related to the field of the '243 patent?

A. Well, obviously the '243 patent discusses an audience measurement system. So there is the -- you know, identifying people, performing facial recognition on audience members, counting people. So acquiring imagery and processing that imagery to -- for certain outcomes.

Q. Do you have any previous experience in audience measurement?

A. I haven't done direct work in audience measurement, because this is not one of the applications I have targeted. But all the image processing applications that go with what's being discussed in the '243 is work that I've been involved with for over 30 years.

Lu-577's first embodiment

Para. 39: Lu-000's face analysis

may be video camera, an IR camera, or the like. 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 (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.

Para. 40: Lert's audio signature extraction

ing digital television set **66**, analog television set **70**, and personal computer **80** contain television program-identifying codes, **characteristic 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.

UNDISPUTED

television set **66**, and by a detector **68** in the case of an analog television set **70**. The detector **64** may detect the audio portion of a program to which the digital television set **66** is tuned by non-intrusively detecting the sound provided by a speaker **72** of the digital television set **66** (in which case the detector **64** may act as a microphone) or by intrusively detecting the audio signal from a direct connection to the speaker, the audio output terminals, or the audio processing circuitry of the digital television set **66**. Similarly, the detector **68** may detect the audio portion of a program to which the analog television set **70** is tuned by non-intrusively detecting the sound provided by a speaker **74** of the digital television set **66** or by intrusively detecting the audio signal from a direct connection to the speaker, the audio output terminals, or the audio processing circuitry of the analog television set **70**. The analog television receiver **70** is

[0036] The audio portion of a television program may also be detected by a detector 78 when the television program is played by a personal computer 80 on a monitor 82 and speakers 84. The detector 78 may detect the audio portion of a program to which the personal computer 80 is tuned by the user 86. The detector 78 may also detect the audio portion 84 or by introspectively detecting the audio signal from a direct connection to the speaker, the audio output terminals, or the audio processing circuitry of the personal computer 80.

[0037] As has been taught by Thomas et al., in U.S. Pat. No. 5,629,739, a representation of the audio portion picked up by the detector 64 can be analyzed by a site unit (SU) 86 or a home unit (HU) 88 which may be provided with data storage and forwarding unit capability within the statistically selected monitoring site 62 in order to determine

The audio portion from the digital television set 86 contains television program identification codes. Alternatively, the representation of the audio portion picked up by the detector 84 can be analyzed at a remotely located data collection central office 90 in order to determine if the audio portion from the digital television set 86 corresponds to such television program identification code.

A computer based data collection central office 90 changes data with the home network 92, such as a public Internet, or the like.

[0038] Similarly, a repres

picked up by the detector 60, identify the persons watching the analog television set 70 at 94 and/or the home unit 88 is a portion from the analog television program-identifying codes, a portion picked up by the disc site unit 96 and/or the home unit 100 is the audio portion from the television program-identifying

[0039] Moreover, a person in order to identify the person coming on the digital television

computer identification of the audio portion of the computer-based data collection changes with the home use of the Internet, or the Internet, or the Internet.

[0403] Moreover, a representative example of the use of the computer-based data collection system is the use of the computer-based data collection system to identify the audio portion of the computer-based data collection system.

[0404] Instead of analyzing the representation of the audio portion of the computer-based data collection system to determine if the audio portion from the corresponding television set is the same as the audio portion of the computer-based data collection system, the computer-based data collection system 70 can contain television program-identifying codes, characteristic signatures can be extracted from the audio portion of the computer-based data collection system, statistically selected monitoring site 62 or in the television set, and the computer-based data collection system 70 can be used to identify the television program to which the digital television set 64 is the analog television set 62.

[0405] Moreover, a representative example of the use of the computer-based data collection system is the use of the computer-based data collection system to identify the audio portion of the computer-based data collection system.

100455 According to this second conventional 100, using data, and possibly demographic data, are collected from digital television reception equipment that includes processing the program preferably having adequate non-volatile memory-based memory available so that an algorithm, transmitter referred to in a software agent, can be stored thereon. The conventional 100 may include (a) a personal computer 100 that may be similar to the personal computer 40 of FIG. 1, and (b) a digital television 100 that may be similar to the digital converter 50 of FIG. 1. (c) an analog television 100 operating in response to the digital converter 106, and (d) digital television set 108.

SPUTN

Par

ing

digital television

Audio signals

nature extr

Reaction

Ex-1004, ¶¶0035-43, 0003; Reply, 25

Lu-577 does not disparage Lert



Dr. Doermann

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.

PO did not provide expert testimony or depose Dr. Doermann about its disparagement argument



No motivation to combine is required for teachings regarding the same embodiment



Material in a second document is considered incorporated by reference in a host document if the context “makes clear that the material is effectively part of the host document as it were explicitly contained therein.”

Virnetx Inc. v Apple Inc., No. 2022-1523, slip op. at 7 (Fed. Cir. Oct. 20, 2023)



A motivation to combine elements present together in a reference is not required.

Gen. Elec. Co. v. Raytheon Techs. Corp., 983 F.3d 1334, 1352 (Fed. Cir. 2020)

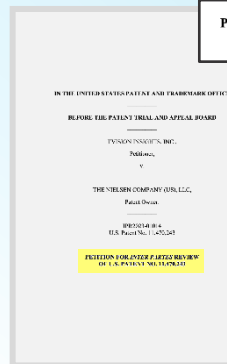
TVision never changed its obviousness theory regarding resolution reduction



Nielsen accuses TVision of introducing a new theory, that Tian renders obvious resolution reduction, in the Reply.

This argument mischaracterizes the Petition and Reply:

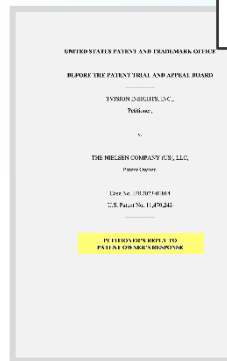
Petition: obvious for Lu-577's system to reduce the resolution in light of Tian's teachings



PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 11,470,243

more images. *Supra* §VI.A.2[1d-1e]. A POSITA would have found it obvious to apply Tian's teaching such that Lu-577's system reduces the resolution of a first image of the one or more of the images and performs face detection or head pose estimation on the reduced-resolution image. *Supra* §VI.A.1; Ex-1003, ¶120.

Reply: obviousness not negated even if Tian does not teach "experimental setup" for resolution reduction



PETITIONER'S REPLY TO PATENT OWNER'S RESPONSE

of processing.¹ Therefore, even under its own characterization, PO cannot deny Tian's explicit teaching of reducing image resolutions. PO's assertion that Tian's "experimental setup" or "structure" does not include a step of reducing a resolution (POR, 3, 5-6, 9), even if accepted,² does not negate the obviousness of doing so in light of Tian's teachings.

Sur-Reply, 7-8; Pet., 26-27; Reply, 4-5

Claims do not require resolution reduction during media content exposure



PO contends the claims require that “resolution reduction must occur during media content exposure.”

Claim 1

Claim 1 requires images to be obtained during media exposure.

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;

Claim 4

No claim requires resolution reduction to occur during media exposure.

4. The audience measurement system of claim 1, wherein the processor circuitry is to:
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
determine the orientation of the head with respect to the camera based on the reduced-resolution image.

Claims do not require resolution reduction during media content exposure



Dr. Saber

Q. Does Claim 1 require detecting a head in an image before capturing the next image?

A. I don't understand your question. I mean, Claim 1 states that you analyze a sequence of images of the media exposure environment to detect a head appearing in one or more of the images. So you're going to detect the head that would be appearing in one or more of the images of the sequence of images you just analyzed.

Q. If I take a sequence of images first before analyzing them and then analyze them to detect a head, will that satisfy this limitation?

A. Yes ...

Nielsen abandons its argument that “[a]pplying Tian’s teachings would not result in a useable ‘audience measurement system’ as claimed”

 Nielsen’s Argument in the POR	 Unchallenged Rebuttal in TVision’s Reply
A system relying on Tian’s teachings would not be able to activate the illumination source in order to perform analysis of succeeding frames. POR, 10-11	- Activating an illumination source is neither required by claims at issue nor part of the Petition’s grounds. Reply, 8 - Per Nielsen’s expert, activating illumination is unnecessary. Reply, 9
Applying Tian’s teachings would result in low-resolution, low-contrast images not suitable for analysis. POR, 11-12	- No claim require some level of contrast. Reply, 9 - Tian proves Nielsen wrong by reporting face detection and pose estimation being performed on down-sampled images with high accuracy. <i>Id.</i>
Down-sampling without antialiasing does not produce an image ready for analysis. POR, 12-13	- The ‘243 patent does not claim antialiasing and discloses a resolution-reduction method without a separate antialiasing step. Reply, 9-10 - Tian achieved high accuracy without performing antialiasing. Reply, 10
Down-sampling with antialiasing would delay capturing subsequent images. POR, 13	Dr. Saber admits that the claims encompass capturing a sequence of images (e.g., as frames of a video) before analyzing any of them. Reply, 10

ABANDONED

POR, 10-13; Reply, 8-10

PO's argument about conventional wisdom depends on a "magical system"



PO: Dr. Doermann agrees image analysis on higher-resolution images would "[o]f course" improve results.

The actual testimony:



Dr. Doermann

Q. But again, if you had a magical system, or FBI system that you alluded to earlier, required no computational resources, required -- had perfect accuracy, and required no processing time, that would be a no brainer, it would be desirable?

A. And it would improve my -- yes. And it would improve my results. I mean, there's an outside chance that it could reduce the results, but in general, I would try it. Of course.

PO's attorney argument about Ex. 1014 cannot rebut Dr. Doermann's expert testimony



PO: Ex. 1014 incurs no processing cost for down-sampling and uses full resolution images when performing face detection.



Dr. Doermann

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.

The Sur-Reply abandons other challenges to the motivation to apply Tian's resolution-reduction teachings



Nielsen's Argument in the POR

Given that the head location and the head orientation are known from face tracking, reducing the resolution has computational costs without any benefit, thus negating any benefit from reducing the resolution.

POR, 24-25

Tian's findings are not applicable to audience measurement, for which a face region has a lower resolution.

POR, 27-28



Unchallenged Rebuttal in TVision's Reply

- This argument misunderstands the combination and the claims, which include performing face detection and head pose estimation on a reduced-resolution image and performing face tracking on different, subsequent, images.

Reply, 19

- Dr. Saber contradicts PO by testifying that the camera does not need to capture the entire media exposure environment.

Reply, 19

- Examples relied on by PO are inapplicable because they are not images taken by a camera in a media exposure environment.

Reply, 20

- PO incorrectly assumes low original resolutions for images.

Id.

ABANDONED

The Sur-Reply abandons the argument that head orientation determination is redundant



POR: a pose estimation step would have been redundant in light of Lu-000's "brute force method."

Unchallenged rebuttal in TVision's Reply:



This argument ignores that:

1. there are head orientations not suitable for face recognition and
2. Lu-000 does not store signatures for all possible head orientations.



The head pose estimation step would prevent wasteful processing for comparing feature signatures extracted from unsuitable images with all stored signatures.



The head pose estimation step would also prevent misidentifications resulting from analyzing unsuitable images.

The Sur-Reply abandons attack on the motivation to modify Lu-577 based on Tian's face tracking teachings

 Nielsen's argument in the POR	 Unchallenged rebuttal in TVision's Reply
TVision fails to account for the increase in processing costs associated with performing the head tracking. POR, 36-37	Nielsen's argument misses the point because replacing face detection with face tracking reduces the number of pixels analyzed and the processing resources needed to do so. Reply, 23
Petitioner fails to account for the need to perform head detection regardless, in order to identify new audience members entering the frame. POR, 37-38	Neither the claims nor the prior-art combinations require locating new faces in any image, let alone an image in which face tracking is performed. Reply, 24
Face tracking harms the accuracy of face recognition when using out-of-date face locations. POR, 38-40	TVision's obviousness theory uses a face/head location from a previous image when no change is detected. This ensures no out-of-date face location is used. Reply, 24

ABANDONED

Dated: October 11, 2024

Respectfully submitted,

SHEPPARD, MULLIN, RICHTER &
HAMPTON LLP

/Harper Batts/

Harper S. Batts (Reg. No. 56,160)

Attorney for Petitioner

CERTIFICATE OF SERVICE

The undersigned hereby certifies that on October 11, 2024, a complete copy of the foregoing document was served via email to all parties to this proceeding at the addresses indicated:

Jennifer C. Bailey
jennifer.bailey@eriseip.com

Megan J. Redmond
megan.redmond@eriseip.com

Robin A. Snader
robin.snader@eriseip.com

PTAB@eriseip.com

Date: October 11, 2024

/Harper Batts/
Harper Batts (Reg. No. 56,160)
Attorney for Petitioner