UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

JUMIO CORPORATION,
Petitioner
v.
FACETEC, INC.,
Patent Owner.

IPR2025-00106; IPR2025-00107; IPR2025-00108; IPR2025-00109
U.S. Patent Nos. 10,776,471; 11,157,606; 11,693,938; 11,874,910

DECLARATION OF CHRIS DAFT, D. PHIL. IN SUPPORT OF PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NOS. 10,776,471; 11,157,606; 11,693,938; 11,874,910

TABLE OF CONTENTS

EXHIBIT LIST

Exhibit	Reference
1001	U.S. Patent No. 10,776,471 (“the ’471 Patent,” IPR2025-00106) U.S. Patent No. 11,157,606 (“the ’606 Patent,” IPR2025-00107) U.S. Patent No. 11,693,938 (“the ’938 Patent,” IPR2025-00108) U.S. Patent No. 11,874,910 (“the ’910 Patent,” IPR2025-00109)
1002	File History of the ’471 Patent (IPR2025-00106) File History of the ’606 Patent (IPR2025-00107) File History of the ’938 Patent (IPR2025-00108) File History of the ’910 Patent (IPR2025-00109)
1003	Declaration of Dr. Chris Daft
1004	Dr. Chris Daft Curriculum Vitae
1005	U.S. Patent 8,437,513 (“Derakhshani”)
1006	U.S. Patent Application Publication 2011/0299741 (“Zhang”)
1007	U.S. Patent Application Publication 2002/0113884 (“Tanii”)
1008	U.S. Patent Application Publication 2014/0028823 (“Tahk”)
1009	U.S. Patent Application Publication 2004/0239799 (“Suzuki”)
1010	U.S. Patent Application Publication 2010/0014720 (“Hoyos”)
1011	U.S. Patent 9,077,891
1012	U.S. Patent 8,965,064
1013	Zhengyou Zhang et al., *A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry*, Institut National De Recherche En Informatique Et En Automatique (May 1994)
1014	U.S. Patent Application Publication 2010/0158319
1015	U.S. Patent Application Publication 2007/0127787
1016	U.S. Patent Application Publication 2004/0036574
1017	Eugene Hect & Alfred Zajac, *Optics*, Addison-Wesley Publishing Co. (1974)
1018	Anil K. Jain et al., *Introduction to Biometrics*, Springer (2011)

Exhibit	Reference
1019	B. Honlinger & H.H. Nasse, *Distortion*, Zeiss (Oct. 2009)
1020	Hue and Hatchet, *What is the Focal Plane?*, https://hueandhatchet.com/what-is-the-focal-plane/
1021	*Introduction to Astronomy*, https://physics.weber.edu/palen/phsx1040/lectures/ldistcomp.html
1022	Daniel Baker, *Face distortion is not due to lens distortion*, (May 5, 2012), https://bakerdh.wordpress.com/2012/05/05/face-distortion-is-not-due-to-lens-distortion/
1023	Richard Hartley & Andrew Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge Univ. Press (2nd ed. 2011)
1024	Daniel Moreno & Gabriel Taubin, *Simple, Accurate, and Robust Projector-Camera Calibration*, Brown Univ., mesh.brown.edu/calibration/
1025	Serge Belongie, *CSE 252B: Computer Vision II, Lecture 4: Planar Scenes and Homography* (Apr. 7, 2004)
1026	Merriam-Webster's Collegiate Dictionary (11th ed. 2014)
1027	Laurenz Wiskott, *Face Recognition by Elastic Bunch Graph Matching* (1999)
1028	Axis Communications, *CCD and CMOS sensor technology: Technical White Paper* (2010)
1029	Jeff Meyer, *What is Depth of Field? How aperture, focal length and focus control sharpness*, Digital Camera World (July 17, 2013), https://prism.org.gg/wp-content/uploads/2016/07/Depth-of-Field.pdf
1030	Brian Klug, *Understanding Camera Optics & Smartphone Camera Trends, A Presentation by Brian Klug*, Anandtech (Feb. 22, 2013), https://www.anandtech.com/Show/Index/6777?cPage=2&all=False&sort=0&page=2&slug=understanding-camera-optics-smartphone-camera-trends
1031	*Portrait mode now available on iPhone 7 Plus with iOS 10.1*, Apple (Oct. 24, 2016)
1032	*The Timeline of Evolution of the Camera from the 1600s to 21st Century*, Capture.com (May 5, 2023),

Exhibit	Reference
	https://www.capture.com/blogs/insights/evolution-of-the-camera
1033	John Biggs, *A Nokia Camera Phone That's More Like a Camera With a Phone Attached*, The New York Times (Nov. 29, 2007), ttps://www.nytimes.com/2007/11/29/technology/personaltech/29phone.html
1034	U.S. Patent 7,412,081
1035	U.S. Patent 8,457,367
1036	U.S. Patent 8,260,008
1037	Zac Hall, *Over a decade of selfies, starting with iPhone 4*, 9to5Mac (July 1, 2021), https://9to5mac.com/2021/07/01/iphone-4-selfie-celebration/
1038	Duane C. Brown, *Decentering Distortion of Lenses*, Photogrammertic Engineering (1966)

I, Dr. Chris Daft, declare as follows:

I. INTRODUCTION

1. My name is Christopher Daft. I am currently a technical consultant for River Sonic Solutions specializing in biometric security and medical-device imaging.

2. I have been retained by Petitioner Jumio Corporation ("Jumio" or "Petitioner) to submit a Declaration in support of Jumio's Petitions for *Inter Partes* Review of each of the claims ("Challenged Claims") of U.S. Patent Nos. 10,776,471 ("the '741 Patent"); 11,157,606 ("the '606 Patent"); 11,693,938 ("the '938 Patent"); and 11,874,910 ("the '910 Patent") (collectively, "Challenged Patents").[1]

3. I have been asked to opine on whether the Challenged Claims are disclosed or rendered obvious by the prior art. My opinions are based on my years of education, research, and experience, as well as my investigation and study of relevant materials. The materials that I evaluated in support of this Declaration include all exhibits cited in this Declaration and in the Petitions.

4. I may rely upon these materials, my knowledge and experience, and/or additional materials to rebut arguments raised by the Patent Owner. Further, I may

[1] I understand that each Challenged Patent has been filed as Exhibit 1001 in each respective proceeding, and their prosecution history has been filed as Exhibit 1002. For this reason, I will cite each Challenged Patent by its shorthand title, and the prosecution history as [Shorthand Patent No.] *Prosecution History*.

also consider additional documents and information in forming any necessary opinions, including documents that may not yet have been provided to me.

5. My analysis of the materials produced in this matter is ongoing, and I will continue to review any new material as it is provided. This Declaration represents only those opinions I have formed to date. I reserve the right to revise, supplement, and/or amend my opinions stated herein based on new information and on my continuing analysis of the materials already provided.

6. I am being compensated for my work in this matter at my standard hourly rate of $550. I am also being reimbursed for reasonable and customary expenses associated with my work and testimony in this investigation. My compensation is not contingent on the outcome of this matter or the specifics of my testimony.

7. I make this declaration based upon my own personal knowledge and, if called upon to testify, would testify competently to the matters stated herein.

II. BACKGROUND AND QUALIFICATIONS

8. My qualifications for forming the opinions set forth in this report are summarized here and are presented in my curriculum vitae attached as Exhibit A to this declaration

A. Relevant Academic Experience

9. I hold Bachelors, Masters, and Doctoral degrees from the University of Oxford and have worked in the technology industry since 1990. I taught electrical and computer engineering for three years at the University of Illinois, and conduct research in physiological psychology with a group at the University of Arizona.

B. Relevant Professional Experience

10. I have published 27 peer-reviewed technical papers and hold 25 patents.

11. I am an engineer with over 30 years of product development experience in electronics and signal processing.

12. I am currently Principal at River Sonic Solutions LLC, a technical consulting firm that solves engineering problems in imaging and medical devices, sensors, electronics, and signal processing. I am currently involved in designing a products incorporating an optical system and a computer vision algorithm.

13. Since 2020, I have also served as Chief Technology Officer at a neurotechnology startup, Sanmai Technologies PBC. My responsibilities in this role include designing the wave-propagation aspects of the company's products. The mathematics of the sound propagation in this work mirrors the optics at issue in this case.

14. In this role I have also been involved in computer-vision work, which has resulted in several U.S. patent applications including US63/590,716 ("Non-

Imaging TFUS Systems”), US63/601,577 (“Wearable Closed Loop TFUS System”), US63/589,928 (“TFUS System Configured with Simplified Probes”), and US 63/554,004 (“Ultrasound Simulation Guided by Artificial Intelligence”).

15. I have received several awards for my work in the field. For instance, I received the Senior Key Expert designation as recognition of my technical contributions while employed by Siemens. And while employed at General Electric, I received the Whitney Technical Achievement Award for a software package enabling Six-Sigma product design. I also received the Dushman Award for my contributions to introducing GE’s first premium ultrasound imager.

16. I volunteer with the Institute of Electrical and Electronic Engineers (“IEEE”) and have served on the Technical Program Committee for its International Ultrasonics Symposium for over 10 years. I recently served for two years as Secretary of the Consultants Network of Silicon Valley, an organization which supports self-employed tech workers in the San Francisco Bay Area. For several years, I volunteered as the program manager of Keizai Silicon Valley, a non-profit working to improve links between Northern Californian companies and Japan.

17. I have trial and deposition experience and have been qualified by the courts as an expert on optical devices and biometric security.

18. All the opinions stated herein are based on my personal knowledge, professional judgment and more than 30 years of teaching, research, and work

experience in electrical and computer engineering, biosensing and signal processing, and an analysis of the materials and information that I have considered in preparing this report.

III. MATERIALS CONSIDERED

19. The materials that I considered and relied upon in preparing my Declaration and forming my opinions include all exhibits cited in this Declaration and the Petition, including each of the Challenged Patents, their file histories, and all of the relevant prior art.

20. I also have relied on my academic and professional experience in reaching the opinions expressed in this Declaration.

IV. LEGAL STANDARDS

21. I am not an attorney and offer no legal opinions, but Counsel has informed me of the legal principles that apply in determining whether patent claims are unpatentable for obviousness. Paragraphs in this section beginning with "I understand" or "I have been informed" capture my understanding of the law as provided to me by counsel. The other paragraphs are my opinions.

22. First, I understand that there are two versions of the patent laws, due to a change in the laws that occurred in 2012 under the "America Invents Act" ("AIA") I understand that, because the Challenged Patents were filed after the AIA took effect in March of 2013, the "post-AIA" patent laws apply here.

23. I understand that, under post-AIA law, a claim may be unpatentable as obvious under § 103(a) if the subject matter described in the claim as a whole would have been obvious to a hypothetical "person of ordinary skill in the art" ("POSITA") in view of a prior-art reference, or combination of prior-art references, before the "effective filing date" of the invention. I understand the "effective filing date" considers whether a patent claims priority to earlier patent applications or provisional patent applications.

24. I understand that, before considering the prior art, the claims are first construed to determine what subject matter the claims encompass. I understand that the claims are read from the perspective of a person of ordinary skill in the art ("POSITA") at the time of the alleged invention. I understand that the words of a claim are generally given their ordinary and customary meaning as understood by a POSITA in light of the entire patent specification in which they appear, as well as the prosecution history of the patent.

25. I understand, however, that the words of a claim may be given a meaning other than their plain and ordinary meaning if: (1) a patentee expressly and unambiguously defined a term in a way that is different from its plain meaning; or (2) a patentee "disclaims" the full scope of a claim term's plain meaning, for instance, by making clear and unambiguous statements during prosecution that a

claim term does or does not encompass certain subject matter that may otherwise be encompassed by the plain words of the claim.

26. I understand that, after construing the claims, the teachings of the prior art are considered. I understand that the prior art includes references, among other things, that were known or used by others in this country, or patented or described in a printed publication in this or a foreign country, before the date the invention was allegedly made by the inventor(s). For a printed publication to qualify as prior art, I understand that it must be demonstrated that the publication was disseminated or otherwise sufficiently accessible to the public. However, I also understand that a POSITA is assumed to be aware of all relevant prior art in the field of endeavor covered by the patent-at-issue and all analogous prior art.

27. I understand that although the ultimate question of the obviousness of a claimed invention is a legal determination, obviousness is based on several factual inquiries, including: the scope and content of the prior art, the differences between the claimed subject matter and the prior art, the level of ordinary skill in the art at the time of the invention, and any "objective indicia" or "secondary considerations" of non-obviousness, which must all be considered.

28. I understand that, in determining the scope and content of the prior art, a prior-art reference is considered relevant if it falls within the field of the inventor's endeavor, or if it is reasonably pertinent to the particular problem that the inventor

was trying to solve. A reference is reasonably pertinent if it logically would have commended itself to an inventor's attention in considering the problem sought to be solved.

29. I understand that, to assess the differences between prior art and the claimed subject matter, the claimed invention must be considered as a whole. I understand that this involves showing that a POSITA, confronted by the same problems as the inventor and with no knowledge of the claimed invention, would have identified the elements in the prior art and combined them in the claimed manner.

30. I understand that there are several rationales for combining prior-art references. For instance, I understand that it is considered obvious to:

- combine prior-art elements according to known methods to yield predictable results;
- substitute one known element for another to obtain predictable results;
- use the prior-art elements in a predictable way according to their established functions;
- apply a known technique to a known device (method or product) ready for improvement to yield predictable results;
- choose from a finite number of identified, predictable solutions, with a reasonable expectation of success; or

- if there is some teaching, suggestion, or motivation in the prior art that would have led a POSITA to modify a prior-art reference or combine prior-art teachings to arrive at the claimed invention.

31. I understand, however, that a POSITA does not need to have the same motivations or reasons to combine the prior art as the inventors, and can arrive at the claimed invention for completely different reasons than the inventors.

32. I understand that, when considering the prior art, the prior art cannot be modified or combined using "hindsight." I understand that hindsight refers to situations in which the patent-at-issue is used as a framework to pick and choose pieces from the prior art to arrive at the claims, or adopting rationales identified by the patent-at-issue that those in the art would not have recognized. Instead, I understand that the obviousness inquiry is evaluated from the perspective of a POSITA as of the critical date of the patent-at-issue, with all the prior art before them, to determine whether the POSITA would have independently arrived at the claimed invention without the benefit of the patent-at-issue for guidance.

33. I understand that "objective indicia" or "secondary considerations" of non-obviousness must also be considered in the obviousness analysis, and include whether: (1) there was a long-felt need for the claimed invention; (2) the claimed invention has achieved commercial success; (3) there was copying of the claimed invention by others; (4) there were failed attempts by others to make the alleged

invention; (5) there was praise of the invention in the field; and (6) there was skepticism the claimed invention could be achieved, among a few others.

34. As of the time that I prepared this Declaration, the Patent Owner had not identified any secondary considerations of non-obviousness. Therefore, I reserve the right to elaborate on my understanding of these secondary considerations and to address any secondary considerations that Patent Owner may subsequently identify.

35. Finally, I understand that, in an *inter partes* review proceeding, the obviousness of a claim must be demonstrated by "a preponderance of the evidence," and that this burden falls on the Petitioner. I understand that the preponderance of the evidence standard means that a reasonable factfinder would find a material fact more probable than the nonexistence of that fact. It does not allow for speculation regarding specific facts and is instead focused on whether the evidence more likely than not demonstrates the existence or non-existence of specific material facts. I understand that "preponderance of the evidence" is a lower standard than "clear and convincing evidence" (which requires a fact to be substantially more likely to be true than untrue), or "beyond a reasonable doubt" (which is an exceedingly high standard that I understand is generally reserved for criminal matters).

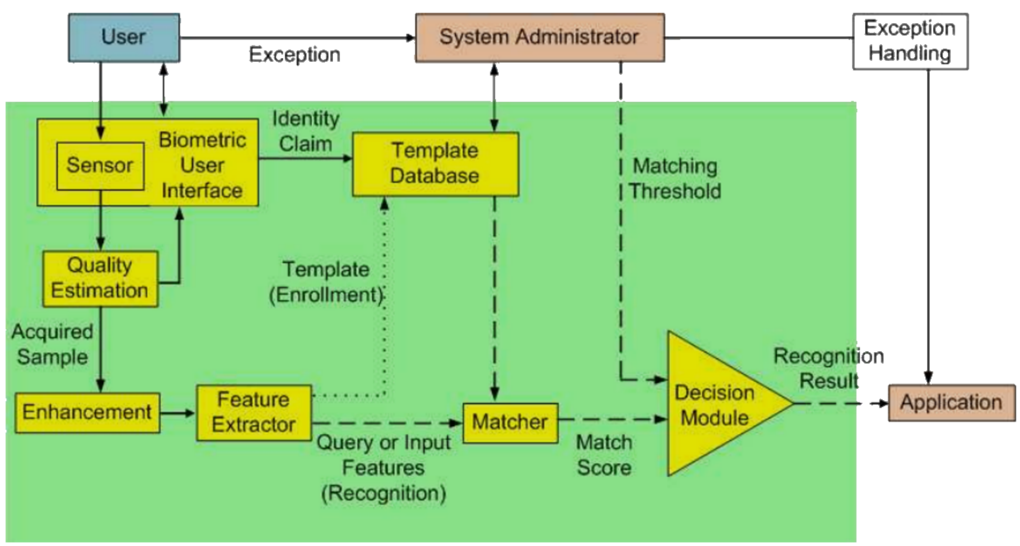
36. I have applied the "preponderance of the evidence" standard throughout my Declaration.

V. TECHNOLOGY BACKGROUND

37. The Challenged Patents relate to biometric security, and specifically the use of computer vision to enable such security for facial authentication systems.

A. Biometric Security

38. Biometric security refers to the use of unique biological information (such as the unique layout of facial features or fingerprints, characteristics of the voice, among many others) to authenticate the identity of an individual and serve as a "key" to provide access to some resource (such as a building, a car, or an electronic device). The following is a simplistic representation of a generic biometric security system:



Ex-1018, 6 (showing components of a biometric system).

39. Biometric security typically involves a two-step process. *Id.*, 4. First, a user presents their biological feature (e.g., their face, finger, voice, etc.) for

enrollment into the system, during which certain identifiers are extracted (such as the relationship between facial features, unique aspects of the fingerprint, or characteristics of the voice) and associated with the user's identity. *Id.* In other words, the enrollment process associates the unique aspects of the biological feature to the user (e.g., "this biometric feature is mine"). Second, when an already-enrolled user seeks to access a secured resource, the user re-presents the same biological feature, the identifiers are extracted again, and the identifiers are then compared to the identifiers previously captured during enrollment to determine whether there is sufficient match. *Id*. In other words, this "authentication" or "recognition" step is intended to *confirm* that the user has already been authorized to access the secure resource (e.g., "this biometric feature is the same as the one I previously enrolled as mine.").

40. Many different types of biometric-security systems were known in the art as of 2014 (which I understand to be the earliest possible priority date of the Challenged Patents) that used different biological features to identify users, including fingerprint, palmprint, iris (eye), face, walking gait, voice, and others. *Id.*, 30-34. The selection of a particular biological feature for authentication was largely a design choice, which considered factors such as accuracy, speed, ease of use, and risk of spoofing (e.g., an unauthorized user tricking the system). *Id*., 29-30.

B. Facial Recognition Systems

41. Facial-recognition systems were a well-known type of biometric-security system. *Id.*, 32, 98-103. They typically incorporate the use of "computer vision"—programming a computer to identify distinguishing features (in this case, a face) in an image. *Id.*, 103 ("A typical face recognition system is composed of three modules: (1) image acquisition, (b) face detection, and (c) face matching"). In other words, facial recognition is often performed by capturing two-dimensional images of a user's face, and then extracting the relevant facial features for identification and authentication from the captured images. *Id.*, 104-106 (acquisition), 109-111 (detection), 116-118 (feature extraction and matching).

42. However, biometric-security systems (including facial-recognition systems) were also prone to well-known "spoofing" attacks, such as by presenting a photograph of an enrolled user's face. *Id.*, 269. For this reason, biometric-security systems often employed countermeasures to prevent these types of spoofing attacks. *Id.*, 272. Fingerprint sensors, for instance, were designed to not only evaluate whether a fingerprint matched an already-enrolled fingerprint, but could also incorporate a capacitive element to ensure the fingerprint conducted electricity like a real finger would. *See, e.g.*, *id*., 272, 275. But because facial-recognition systems did not involve direct contact with the biometric sensor (because they instead typically operated based on images of the user's face), other spoofing

countermeasures were used, such as having users change their facial features, *id.*, 273, or even capturing multiple images of the face to verify the user is three-dimensional, rather than a two-dimensional picture.

43. As explained in further detail below, many facial-recognition systems based their spoofing countermeasures on well-known properties of camera optics to distinguish real, three-dimensional faces from pictures of a face.

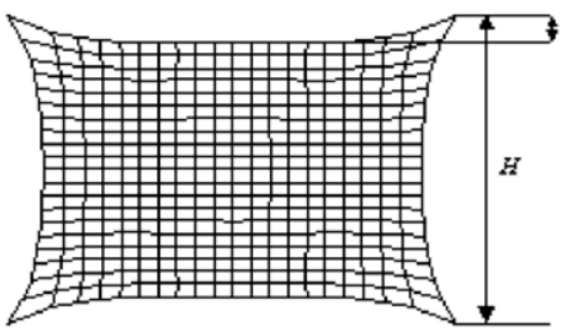
C. Camera Lens Systems

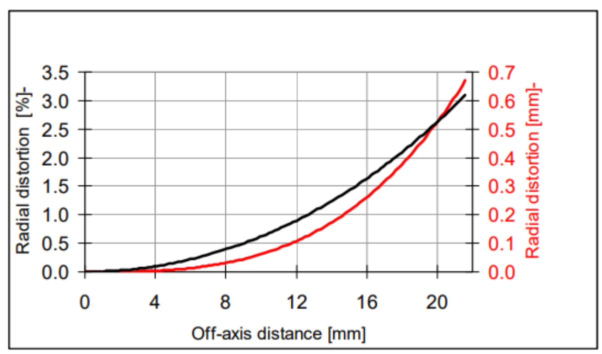
44. Many different optics principles had been utilized throughout the prior art to derive depth information from two-dimensional images. For instance, the use of focus distance or parallax principles—each of which I describe in more detail below in the context of the prior art—were well-known tools to derive depth information using a series of two-dimensional images of an object or scene. Rather than use these principles, however, the Challenged Patents here use well-understood image distortions that are also attributable to the depth of objects or scenes in captured images. These distortions were known to arise as a byproduct of the way the shapes of objects interact with the shape of the camera's lens when capturing images.

45. Specifically, camera lens systems exhibit numerous imperfections when capturing images of an object or scene, such as imperfections in color, spherical aberration, comatic aberration, astigmatism, field curvature and distortion.

These systems can also introduce ***distortion*** due to the shape of the lens, in which the ratio of the image height to the object height (called the transverse magnification) depends on the radial distance of the object from the optic axis. Ex-1017, 176-86. These types of distortions may be positive, where the magnification increases with the transverse distance, or negative.

46. Due to their effects on images, positive distortion is also known as "pincushion" distortion—which makes center of an image scene appear as if it being pulled away from the camera relative to the periphery. An example of "pincushion" distortion is provided below, with the left figure showing how a captured image distorts the scene being captured, and the right figure showing the different degrees of distortion that occurs based on the distance from the image center:





Ex-1019, 5. Telephoto lenses are more prone to pincushion distortion. For clarity, I distinguish here between the simple effect in systems with a lens, where magnification depends on the *axial distance* from the camera to the object, and pincushion and other radial distortions, where magnification depends on the *lateral extent* of the object.

47. The change in magnification with object distance can be easily quantified. Magnification is the ratio of image height (on the image sensor) h_i to the object height h_o. This is equal to the ratio of the distance from the lens to the image sensor d_i and the distance from the lens to the object d_o:

$$M = \frac{h_i}{h_o} = \frac{d_i}{d_o}$$

See Ex-1017, 112.

48. This relationship comes from the fact that when light rays pass through a lens, they form similar triangles between the object and the image, because the angles of incidence and refraction are the same for corresponding rays. With d_i approximately fixed (as is typical in cell phone cameras), the equation shows that the *magnification h_i/h_o becomes larger as the distance to the object d_o becomes smaller*.

49. We can also quantify the amount of radial distortion using an equation valid for both the pincushion and barrel types:

$$h_i = h_u\left(1 + C\frac{h_u^2}{f^2}\right)$$

This equation is a simplification of the general radial distortion model discussed in Brown which captures the essential physics. Ex-1017, 184-86; Ex-1038, 444-462.

Additional terms of higher order than h_u^2 may be introduced to increase accuracy if necessary, but do not change my conclusions. Here h_u is the undistorted height of the object on the image sensor, which would be produced by a perfect lens. As before h_i is the actual height of the object on the image sensor, which is distorted (or warped, in Tanii's definition of that term); f is the lens focal length, and C is a constant defined by details of the lens design. C is positive for pincushion distortion and negative for barrel distortion. This equation shows that, as Tanii explains, *the amount of radial distortion increases as the lens's focal length f decreases*.

50. Negative distortion is termed "barrel" distortion or "fish-eye" distortion—which makes the center of an image scene appear as if it is bulging towards the camera relative to the periphery. Ex-1019, 7. Thin lenses exhibit less distortion than thick lenses such as wide-angle devices, or those with short focal lengths (the distance between the camera's lenses and the image sensor).

51. Distortion can be reduced by adding more focusing surfaces to the lens, resulting in complex lens designs. For example, a rectilinear lens keeps all straight lines in the object as straight lines in the image. *See* Ex-1019, 3.

52. With this background we can understand the difference in imaging performance between a camera photographing either a flat object or a convex object, such as a face. Some distortion will be observed when a camera captures an image of a flat surface, due to the object being extended laterally from the optic axis.

However, an image of a convex three-dimensional object will show more distortion with the same camera than an image of a flat surface. This is because the parts of the three-dimensional object closer to the lens experience more magnification than those further away. If central features in the object are closer to the camera, this change in magnification with distance amplifies any barrel distortion effect. An example where this combination of effects occurs is a close-up image of a face where the nose is near the center of the image, and is the closest feature to the camera.

53. To summarize, image magnification is larger (a) for features in the object that are near the optic axis of a camera lens with barrel distortion, or (b) for features at shorter distances from the camera, regardless of lens shortcomings. Both effects may occur simultaneously in an imaging system.

VI. THE CHALLENGED PATENTS

A. The ’471 Patent

1. ’471 Specification

54. The ’471 Patent is titled “Facial Recognition Authentication System Including Path Parameters.” ’471 Patent, Cover.

55. Based on my review, I understand the ’471 Patent describes countermeasures to prevent “spoofing” attacks on facial-recognition systems in which an imposter presents a picture of a face to gain unauthorized access to a secure system by verifying that the user’s face is three dimensional. *Id.*, 1:64-67. The ’471 Patent looks for a well-known optical effect—the “fish-eye effect” or “fish-eye

distortion"—that is a byproduct of the camera's lens and a function of the distance between the camera and the object (in this case, a face) being captured in the image. *Id*., 3:52-59, 28:34-58.

56. To take advantage of this particular optical effect, the '471 Patent discloses that the facial-recognition system should capture two images of the user's face: one image where the user is "close" to the camera, and one image where the user is "far" from the camera. *Id*., 29:4-7. If the user's face is truly three-dimensional, the "close" image should exhibit expected fish-eye distortion, but the "far" image should have less distortion. *Id.*, 29:11-19. This is because, when the face is close to the camera, it occupies a larger portion of the field of view, and thus peripheral features of the face will be distorted more. But when the face is far from the camera, it occupies a smaller, more-centrally located portion of the field of view, and thus is less prone to fish-eye distortion.

2. '471 Prosecution History

57. I have reviewed the prosecution history of the '471 Patent and understand from my review that although the '471 Patent was filed March 18, 2019, it claims priority to a series of parent patents and provisional applications with an earliest possible priority date of August 28, 2014. '471 Patent, Cover. I understand, therefore, that August 28, 2014 is the earliest possible priority date for the '471 Patent, and have applied that date in my analysis of the prior art.

58. I also understand based on my review that the ’471 Patent was subject to one prior-art rejection. ’471 *Prosecution History*, 59-71. After a claim amendment that directed the claimed invention to verifying the three-dimensionality of a user’s face, however, the claims issued over the prior art. *Id.*, 44-52 (amendment), 8-16 (issuance).

59. The only prior-art reference that I rely upon below that appeared during prosecution is Hoyos. Specifically, the Hoyos *patent* is cited on the face of the ’471 Patent as being considered during prosecution, ’471 Patent, Cover, but that reference was never applied against the claims. Here, I rely on the earlier *application* publication of Hoyos for a narrow teaching related to a single dependent claim.

3. ’471 Claim Construction

60. As I noted above, I understand that unless claim terms are provided an express construction, the terms must be given their plain and ordinary meaning.

61. I understand that Petitioner is not advancing any constructions at this time. I have therefore applied the plain and ordinary meaning of each term in the claims throughout this Declaration as it relates to the ’471 Patent.

62. However, based on my review of the ’471 Patent, there are some terms which may be considered unclear. For instance, the independent claims recite that one of the two captured images exhibits “***expected differences***” or “***expected distortions***” resulting from movement of the camera or the user. Although I consider

this term ambiguous, I am not offering an opinion as to whether it is "definite" or "indefinite" as a legal matter. For purposes of this *inter partes* review, however, I have assumed that the term at least encompasses differences or distortions that would be expected from certain known optical principles—such as "fish-eye" distortion—that can be indicative of depth of a face, consistent with the '471 Patent specification. *See, e.g.*, '471 Patent, 3:52-59, 28:34-58. In assessing the prior art, I have considered whether the prior art teaches looking to these types of optical-based differences or distortions in an image to indicate whether the face has depth or not.

63. Moreover, claim 1 requires "[a] system" comprising a "computing device" that has a processor, screen, camera, and memory. But claim 1 later specifies that the memory is "configured to store machine readable instructions *that are stored on **the memory** of the authentication server*," which is then executed by the "computing device's" processor. By reciting "the memory" when referring to an authentication server—and not the "computing device"—it is unclear whether claim 1 is directed to a ***server*** computing device that must have a processor, screen, and camera, or a networked, client-computing device that merely receives instructions from an authentication server.

64. Although these terms may be unclear, in my opinion, that is immaterial because every claim element was taught by the prior art, as I explain in further detail below.

B. The ’606 Patent

1. ’606 Specification

65. The ’606 Patent is titled “Facial Recognition Authentication System Including Path Parameters.” ’606 Patent, Cover.

66. Based on my review, I understand the ’606 Patent to describe countermeasures to prevent “spoof[ing]” attacks on facial-recognition systems in which an imposter presents a picture of a face to gain unauthorized access to a secure system by verifying that the user’s face is three dimensional. *Id*., 1:67-2:3. The ’606 Patent looks for a well-known optical effect—“known as perspective distortion” or what “[s]ome texts may refer to as fish-eye type distortion”—that is a byproduct of the camera’s lens and a function of the distance between the camera and the object (in this case, a face) being captured in the image. *Id.*, 29:10-43.

67. To take advantage of this particular optical effect, the ’606 Patent discloses that the facial-recognition system should capture two images of the user’s face: one image where the user is “close” to the camera, and one image where the user is “far” from the camera. *Id*. 30:8-11. If the user’s face is truly three-dimensional, the “close” image should exhibit expected radial (e.g., perspective, fish-eye, or barrel) distortion, but the “far” image should have less distortion. *Id.*, 30:15-23. This is because, when the face is close to the camera, it occupies a larger portion of the field of view, and thus peripheral features of the face will be distorted

more. But when the face is far from the camera, it occupies a smaller, more-centrally located portion of the field of view, and thus is less prone to radial distortion.

2. ’606 Prosecution History

68. I have reviewed the prosecution history of the ’606 Patent and understand from my review that although the ’606 Patent was filed March 12, 2020, it claims priority to a series of parent patents and provisional applications with an earliest possible priority date of August 28, 2014. ’606 Patent, Cover. I understand, therefore, that August 28, 2014 is the earliest possible priority date for the ’606 Patent, and have applied that date in my analysis of the prior art.

69. I also understand based on my review that the ’606 Patent issued without any prior-art rejections. *See generally* ’606 *Prosecution History*. Instead, the Examiner appears to have allowed the claims after proposing a minor amendment: that, during the three-dimensional verification, “biometric data” derived from a first image is matched to “biometric data” derived from a second image, rather than to previously “stored” biometric data. *Id.*, 19-26. It appears, therefore, that this limitation was originally directed to a comparison of biometric data acquired during authentication with the biometric data acquired during enrollment, but the Examiner proposed changing it so that the matching related to the three-dimensional verification process instead.

70. Furthermore, none of the prior art I rely upon below appears to have been used by the Examiner to reject the claims, although one prior-art reference (Derakhshani) was cited during prosecution. *Id.*, 30; ’606 Patent, Cover.

3. ’606 Claim Construction

71. As I noted above, I understand that unless claim terms are provided an express construction, the terms must be given their plain and ordinary meaning.

72. I understand that Petitioner is not advancing any constructions at this time. I have therefore applied the plain and ordinary meaning of each term in the claims throughout this Declaration as it relates to the ’606 Patent.

73. However, based on my review of the ’606 Patent, there are some terms which may be considered unclear. For instance, the independent claims recite that one of the two captured images exhibits “***expected differences*** resulting from movement of the camera or the user.” Although I consider this term ambiguous, I am not offering an opinion as to whether it is “definite” or “indefinite” as a legal matter. For purposes of this *inter partes* review, however, I have assumed that the term at least encompasses differences that would be expected from certain known optical principles—such as “perspective” or “fish-eye” distortion—that can be indicative of depth of a face, consistent with the ’606 Patent specification. *See, e.g.*, ’606 Patent, 29:14-19, 30:15-23. In assessing the prior art, I have considered whether

the prior art teaches looking to these types of optical-based differences in an image to indicate whether the face has depth or not.

74. Furthermore, I have recognized that some claims contain some inconsistencies. For instance, independent claim 10 recites "[a] system" in the preamble but dependent claims 11-18—which depend from claim 10—recite "[t]he method" in their preambles. Moreover, claim 6 refers to a "hand-held" device, but claim 7—which depends from claim 6—recites devices such as a "desktop computer," which in my opinion a POSITA would not consider to be "hand-held." However, for purposes of this inter partes review, I have assumed that dependent claims 11-18 recite "[t]he system" to be consistent with independent claim 10, and that a desktop computer can be considered "hand-held."

75. Moreover, although these terms may be unclear, in my opinion, that is immaterial because every claim element was taught by the prior art, as I explain in further detail below.

C. The '938 Patent

1. '938 Specification

76. The '938 Patent is titled "Facial Recognition Authentication System Including Path Parameters." '938 Patent, Cover.

77. Based on my review, I understand the '938 Patent describes countermeasures to prevent "spoofing" attacks on facial-recognition systems in

which an imposter presents a picture of a face to gain unauthorized access to a secure system by verifying that the user's face is three dimensional. *Id.*, 1:66-2:2. The '938 Patent looks for a well-known optical effect—the "fish-eye effect" or "fish-eye distortion"—that is a byproduct of the camera's lens and a function of the distance between the camera and the object (in this case, a face) being captured in the image. *Id.*, 3:55-60, 28:37-61.

78. To take advantage of this particular optical effect, the '938 Patent discloses that the facial-recognition system should capture two images of the user's face: one image where the user is "close" to the camera, and one image where the user is "far" from the camera. *Id.*, 29:6-22. If the user's face is truly three-dimensional, the "close" image should exhibit expected fish-eye distortion, but the "far" image should have less distortion. *Id.*, 29:14-22. This is because, when the face is close to the camera, it occupies a larger portion of the field of view, and thus peripheral features of the face will be distorted more. But when the face is far from the camera, it occupies a smaller, more-centrally located portion of the field of view, and thus is less prone to fish-eye distortion.

2. '938 Prosecution History

79. I have reviewed the prosecution history of the '938 Patent and understand from my review that although the '938 Patent was filed August 27, 2020, it claims priority to a series of parent patents (including the '471 Patent above) and

provisional applications with an earliest possible priority date of August 28, 2014. ’938 Patent, Cover. I understand, therefore, that August 28, 2014 is the earliest possible priority date for the ’938 Patent, and have applied that date in my analysis of the prior art.

80. I also understand based on my review that the ’938 Patent issued after a single prior-art rejection. ’938 *Prosecution History*, 83-105. In response to that rejection, the Applicant amended one of the claims to specify that the differences between the two images would be attributable to this “fish-eye” distortion, (*id.*, 65-71), but argued for the remainder of the independent claims that none of the Examiner’s cited prior-art references taught comparing image-derived data to look for expected differences/distortions, but made other types of comparisons, (*id*., 72-77). The Examiner accepted those arguments, proposed a few minor amendments to the claims, and allowed the patent to issue. *Id.*, 18-28. None of the prior art presented here was before the Examiner during prosecution. ’938 Patent, Cover.

3. ’938 Claim Construction

81. As I noted above, I understand that unless claim terms are provided an express construction, the terms must be given their plain and ordinary meaning.

82. I understand that Petitioner is not advancing any constructions at this time. I have therefore applied the plain and ordinary meaning of each term in the claims throughout this Declaration as it relates to the ’938 Patent.

83. However, based on my review of the ’938 Patent, there are some terms which may be considered unclear. For instance, the independent claims recite that one of the two captured images exhibits “***expected differences***” resulting from movement of the camera or the user. Although I consider this term ambiguous, I am not offering an opinion as to whether it is “definite” or “indefinite” as a legal matter. For purposes of this *inter partes* review, however, I have assumed that the term at least encompasses differences or distortions that would be expected from certain known optical principles—such as “fish-eye” distortion—that can be indicative of depth of a face, consistent with the ’938 Patent specification. *See, e.g.*, ’938 Patent, 3:55-60, 28:37-61. In assessing the prior art, I have considered whether the prior art teaches looking to these types of optical-based differences or distortions in an image to indicate whether the face has depth or not.

84. Furthermore, claims 1 and 7 appear to present circular limitations that may render the claims unclear. Specifically, claim 1 states that biometric data is received or derived from images (which are made up of “image data” because these are computerized images), but claim 7—which depends from claim 1—states that the biometric data *comprises* image data. It is therefore somewhat unclear what constitutes “biometric data” and “image data,” or where the line between the two is. However, for purposes of offering my opinions, this circularity can be avoided by

treating image data that includes facial features as biometric data. This will be the assumption I make in the subsequent opinions.

85. Finally, claim 20 specifies evaluating data to determine the data does "not exhibit [first or second] characteristics." In my opinion, the term "characteristics" as used in the claim is ambiguous. However, I have again assumed that "characteristics" can at least include the types of optical-based differences—such as fish-eye distortion—the '938 Patent discloses.

86. Moreover, although these terms may be unclear, in my opinion, that is immaterial because every claim element was taught by the prior art, as I explain in further detail below.

D. The '910 Patent

1. '910 Specification

87. The '910 Patent is titled "Facial Recognition Authentication System Including Path Parameters." '910 Patent, Cover.

88. Based on my review, I understand the '910 Patent describes countermeasures to prevent "spoofing" attacks on facial-recognition systems in which an imposter presents a picture of a face to gain unauthorized access to a secure system by verifying that the user's face is three dimensional. *Id.*, 1:45-48. The '910 Patent looks for a well-known optical effect—known as "perspective" or "fish-eye" distortion—that is a byproduct of the camera's lens and a function of the distance

between the camera and the object (in this case, a face) being captured in the image. *Id*., 3:39-44, 28:53-29:20.

89. To take advantage of this particular optical effect, the ’910 Patent discloses that the facial-recognition system should capture two images of the user’s face: one image where the user is “close” to the camera, and one image where the user is “far” from the camera. *Id*., 29:51-55. If the user’s face is truly three-dimensional, the “close” image should exhibit expected fish-eye distortion, but the “far” image should have less distortion. *Id.*, 29:59-67. This is because, when the face is close to the camera, it occupies a larger portion of the field of view, and thus peripheral features of the face will be distorted more. But when the face is far from the camera, it occupies a smaller, more-centrally located portion of the field of view, and thus is less prone to fish-eye distortion.

2. ’910 Prosecution History

90. I have reviewed the prosecution history of the ’910 Patent and understand from my review that although the ’910 Patent was filed October 22, 2021, it claims priority to a series of parent patents (including the ’606 Patent above) and provisional applications with an earliest possible priority date of August 28, 2014. ’910 Patent, Cover. I understand, therefore, that August 28, 2014 is the earliest possible priority date for the ’910 Patent, and have applied that date in my analysis of the prior art.

91. I also understand based on my review that the ’910 Patent was never subject to a prior-art rejection. Instead, the only rejection ever issued was for obviousness-type double patenting, which I understand means the claims of the ’910 Patent application were just obvious variants of claims in other patents or patent applications filed by the Applicant. ’910 *Prosecution History*, 83-105. In response, I understand that the Applicant filed a series of terminal disclaimers and made a few minor claim amendments. *Id*., 59-63. The claims were subsequently allowed. *Id*., 18-29

3. ’910 Claim Construction

92. As I noted above, I understand that unless claim terms are provided an express construction, the terms must be given their plain and ordinary meaning.

93. I understand that Petitioner is not advancing any constructions at this time. I have therefore applied the plain and ordinary meaning of each term in the claims throughout this Declaration as it relates to the ’910 Patent.

94. However, based on my review of the ’910 Patent, there are some terms which may be considered unclear. For instance, the independent claims recite that one of the two captured images exhibits “***expected differences***” resulting from movement of the camera or the user. Although I consider this term ambiguous, I am not offering an opinion as to whether it is “definite” or “indefinite” as a legal matter. For purposes of this *inter partes* review, however, I have assumed that the term at

least encompasses differences or distortions that would be expected from certain known optical principles—such as "perspective" or "fish-eye" distortion—that can be indicative of depth of a face, consistent with the '471 Patent specification. *See, e.g.*, '910 Patent, 3:39-44, 28:53-29:20. In assessing the prior art, I have considered whether the prior art teaches looking to these types of optical-based differences or distortions in an image to indicate whether the face has depth or not.

95. Moreover, claim 1 requires "[a] system" comprising a "computing device" that has a processor, screen, camera, and memory. But claim 1 later specifies that the memory is "configured to store machine readable instructions *that are stored on* ***the memory*** *of the authentication server*," which is then executed by the "computing device's" processor. By reciting "the memory" when referring to an authentication server—and not the "computing device"—it is unclear whether claim 1 is directed to a ***server*** computing device that must have a processor, screen, and camera, or a networked, client-computing device that merely receives instructions from an authentication server.

96. There are also a handful of other errors (e.g., grammatical or antecedent basis) throughout the claims, some of which I identify below.

97. Although these terms may be unclear, in my opinion, that is immaterial because every claim element was taught by the prior art, as I explain in further detail below.

VII. PRIOR ART

98. Below is an overview of the primary prior-art references that I considered when analyzing the validity of the claims of the Challenged Patents.

A. Derakhshani (Ex-1005)

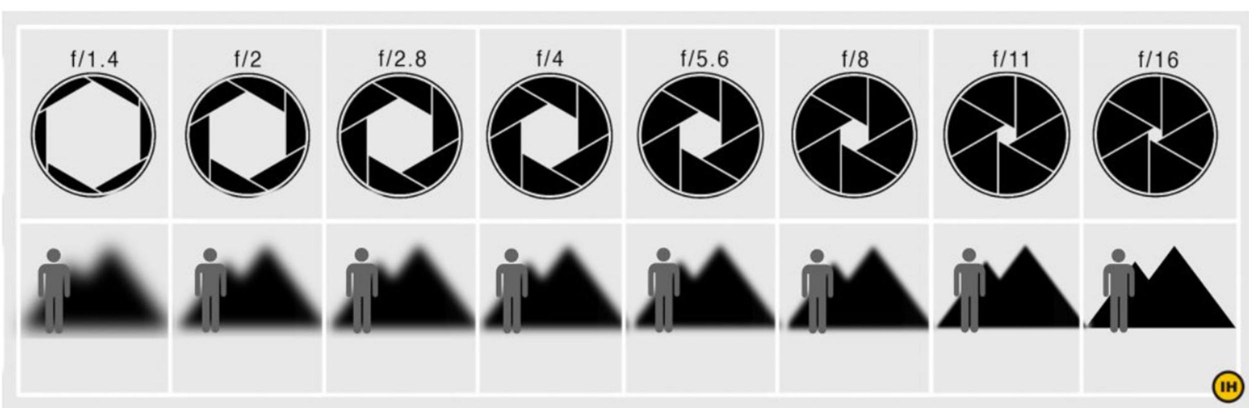
99. Derakhshani is a U.S. Patent (No. 8,437,513) that was filed on August 10, 2012, and issued May 7, 2013. I therefore understand that Derakhshani constitutes prior art under 35 U.S.C. §§102(a)(1) and (a)(2).

100. Derakhshani is titled "Spoof Detection for Biometric Authentication," and discloses a "biometric authentication" process using a device (e.g., a computer or smartphone) that captures images with a camera. *Derakhshani*, 1:11-25, 5:22-27, 6:3-5, 9:10-22, 18:1-3.

101. Although Derakhshani is primarily focused on a biometric-security system that uses the eyes for authentication, the authentication process also incorporates anti-spoofing countermeasures to ensure the user is presenting eyes from a real, three-dimensional face rather than a picture of a face. *Id.*, Abstract (ocular authentication), 16:44-18:4 (determining spatial metric of three-dimensional face verification). For this reason, I consider Derakhshani analogous prior art to the Challenged Patents because they are in the same field of endeavor (anti-spoofing countermeasures for biometric security).

102. To distinguish between three-dimensional faces and two-dimensional pictures of a face, Derakhshani discloses calculating a "spatial metric" for the entire face. *See, e.g.*, *id.*, 16:44-47. Similar to the Challenged Patents, Derakhshani calculates this spatial metric by exploiting one or more known optics principles.

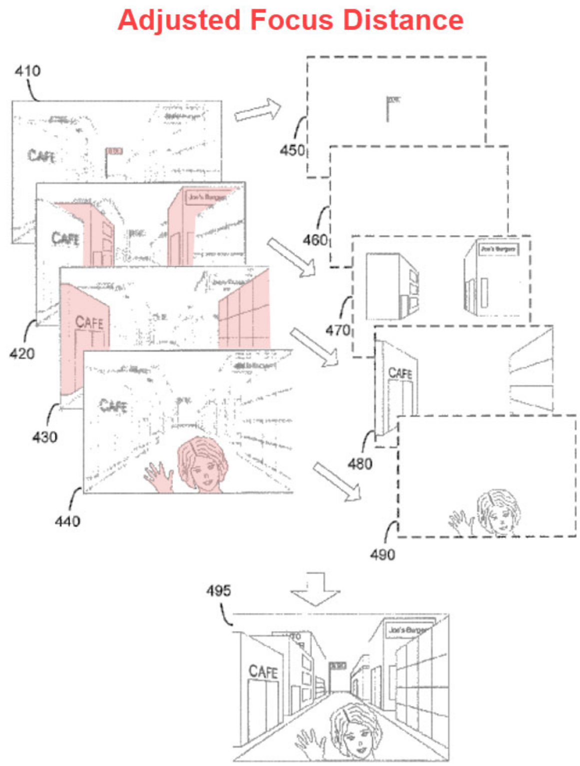
103. The ***first*** exemplary principle Derakhshani proposes exploiting to calculate a "spatial metric" is "focus distance." A POSITA would have understood that "focus distance" refers to the distance from the image sensor to the plane in the scene where the lens is focused. Objects at other distances are out-of-focus, resulting in a loss of spatial resolution. A camera with limited depth of field can illustrate focus distance by a technique used by portrait photographers. The "bokeh" effect makes the background of images appear blurry compared to the clear portrait of the individual being captured. If the lens is moved with respect to the image sensor , the focus distance can be changed while keeping the camera stationary.



See, e.g., Ex-1020, 4.[2]

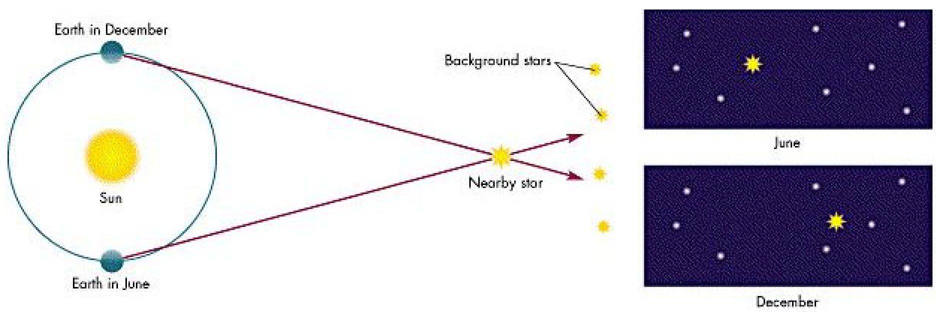
104. It was well-known at the time of the Challenged Patents that the distance and/or depth of objects could be derived by adjusting the focus distance in different images. When the object appears blurry, it can be determined that it exists at a distance *other than* the focus distance of the camera, and when it appears as clear as the camera is capable of producing, it can be determined that it exists at the focus distance of the camera. U.S. Patent No. 9,077,891 provides one example of this principle in practice by capturing a series of images at different focus distances, and determining the relative distances of the objects captured in the image by evaluating when they appear blurry and when they appear clear, as depicted below:

[2] Some of the exemplary figures used throughout this declaration post-date the Challenged Patents. However, the concepts being depicted had all been well-known prior to the ’606 Patent, and the figures are provided only for demonstration of those concepts.



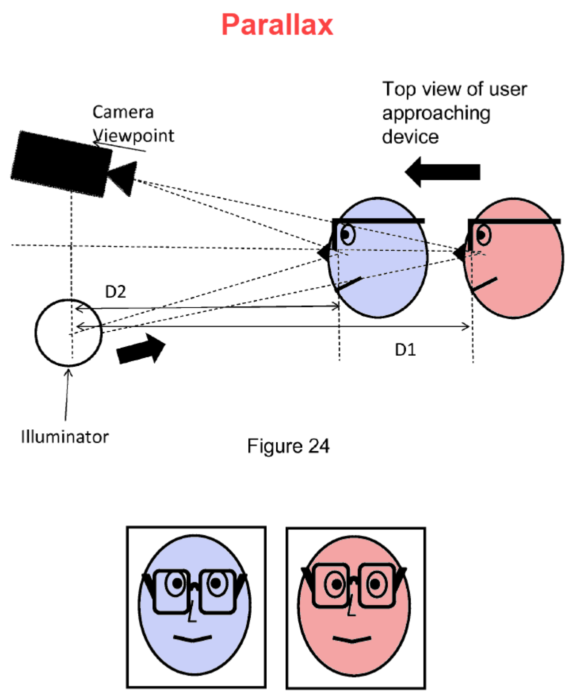
Ex-1011, Fig. 4B (annotated); *see also id.*, 5:38-7:43, Figs. 3A-3D, 4A. Derakhshani exploits this same "focus distance" principle and performs this same process to determine whether the face has three-dimensional depth. *Derakhshani*, 16:54-57 ("A landmark's representation in a particular image has a degree of focus that depends on how far the object corresponding to the landmark is from an in focus point in the field of view of the sensor. Degree of focus is a measure of the extent to the image of the landmark is blurred by optical effects"). This same effect can be exploited with fixed-focus distance cameras, however, by adjusting the *actual* distance of the camera as well.

105. The ***second*** exemplary principle Derakhshani proposes exploiting to calculate a "spatial metric" is "parallax." Parallax refers to a change in the relative positions of objects when viewed from different perspectives. Parallax is commonly associated with astronomy to determine the distance of stars (showing how the view of the "nearby star" changes relative to the "background stars" based on the position of Earth):



Ex-1021, 2.

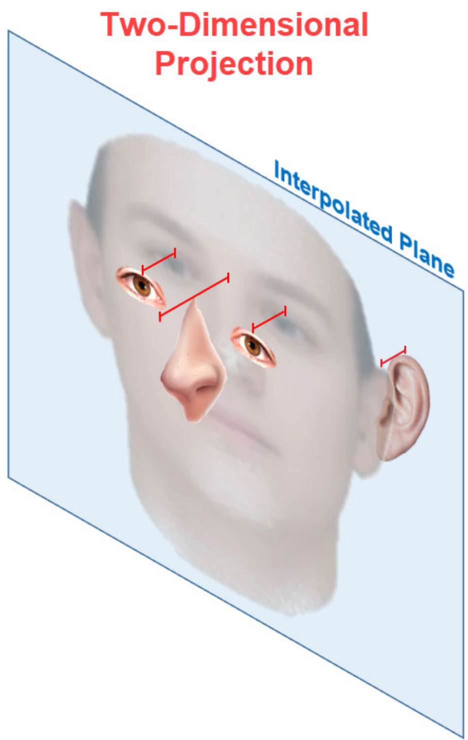
106. Parallax has applications far broader than astronomy, though. It was also well-known at the time of the Challenged Patents that parallax could be exploited to evaluate the depth of objects by evaluating how features are displaced relative to one another across different images taken from different positions. As one example, U.S. Patent No. 8,965,064 exploits parallax as a user approaches a camera (capturing images at two different distances) to ensure accurate images of the eyes are captured during biometric authentication.



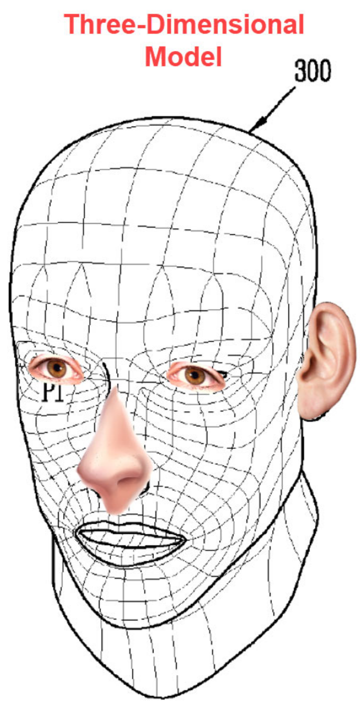
Ex-1012, Figs. 20, 9H (bottom), 24 (top) (annotated). Specifically, the parallax effect in the figure above is revealed by seeing, for instance, how the glasses appear to shift relative to the eyes between the two images. Derakhshani exploits this same parallax principle and performs this same process to determine whether the face has three-dimensional depth. *Derakhshani*, 17:49-52 ("A plurality of images taken from different perspectives on the subject may result in landmarks within the images appearing to move by different amounts because of differences in their distance from the sensor.").

107. Once the images are captured, Derakhshani discloses several different example processes to calculate the "spatial metric." For instance, Derakhshani discloses that the distance information derived by applying the "focus distance"

approach can be compared to an interpolated two-dimensional plane (which Derakhshani calls a “fit plane”), and an average distance of different features from that plane used to calculate the “spatial metric.” *Derakhshani*, 17:12-26. I have created an exemplary depiction of that process below:



108. Additionally or alternatively, Derakhshani discloses comparing the depth information derived from the series of images to a three-dimensional model—either a model specific to the user created during enrollment, or a generic model of a generic face—and determining whether the depth information derived from the series of images deviates from that model. *Derakhshani*, 17:27-44. I have created an exemplary depiction of that process below:



B. Tanii (Ex-1007)

109. Tanii is a U.S. Patent Application Publication (No. 2002/0113884) that was filed on February 15, 2002 and published August 22, 2002. I therefore understand that Tanii constitutes prior art under 35 U.S.C. §§102(a)(1) and (a)(2).

110. Tanii primarily concerns the identification and correction of a specific type of radial distortion when a camera—and particularly a camera with a wide-angle lens often found in mobile devices—captures an image of a face. *Tanii*, [0005], [0007], [0009]. Specifically, the image will exhibit "exaggeration warp" of facial features depending on the distance between the camera and face. *Id.*, [0007]. For this reason, I consider Tanii analogous prior art to the Challenged Patents because they

are both are concerned with identifying and accounting for distortions of three-dimensional faces in images captured using a camera.

111. Specifically, Tanii discloses that when an object such as a face is sufficiently close to the camera, the camera it will produce an "unnatural image … in which the perspective is exaggerated."

Fig.4A

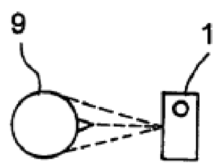
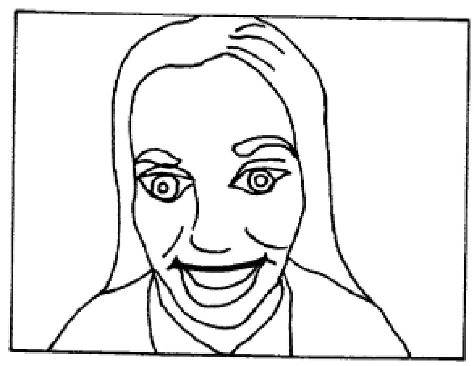


Fig.4B



Id., [0047], Figs. 4A, 4B. But when an object is further from the camera, Tanii discloses that "a natural image can be obtained."

Fig.3A

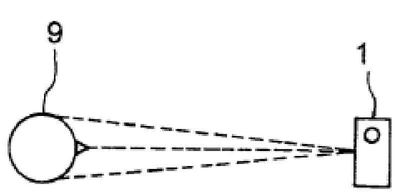
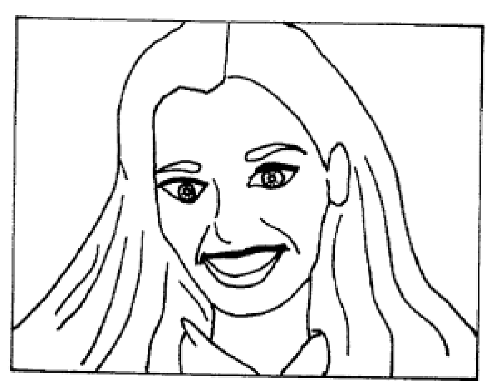


Fig.3B



Id., [0047], Figs. 3A, 3B. According to Tanii, this particular type of radial distortion arises because the face "has an essentially convex configuration that protrudes toward the [camera]," which causes the peripheral areas of the user to appear smaller relative to the center. *Id.*, [0048]. In other words, distortion is attributable in part to barrel distortion, the dependence on magnification of object distance, and possibly other factors. I refer to this distortion broadly throughout this declaration as "distance-induced distortion."

112. This type of distortion was widely understood in the art, and known to be accounted for when capturing images of the face.

Ex-1022. The images above, for instance, show a series of images with the face frames to be the same size in each. The only differences are: (1) the focal length of the lenses used (from a maximum of 85mm to a minimum of 8mm); and (2) distance between the camera and the face (from a maximum of 200cm to a minimum of 20cm).

113. Tanii's critical feature, however, is by providing a procedure to correct this type of distance-induced distortion by enlarging the image's peripheral areas relative to the center to produce an undistorted image. Ex-1007, [0056]

C. Zhang (Ex-1006)

114. Zhang is a U.S. Patent Application Publication (No. 2011/0299741) that was filed on June 8, 2010 and published December 8, 2011. I therefore understand that Tanii constitutes prior art under 35 U.S.C. §§102(a)(1) and (a)(2).

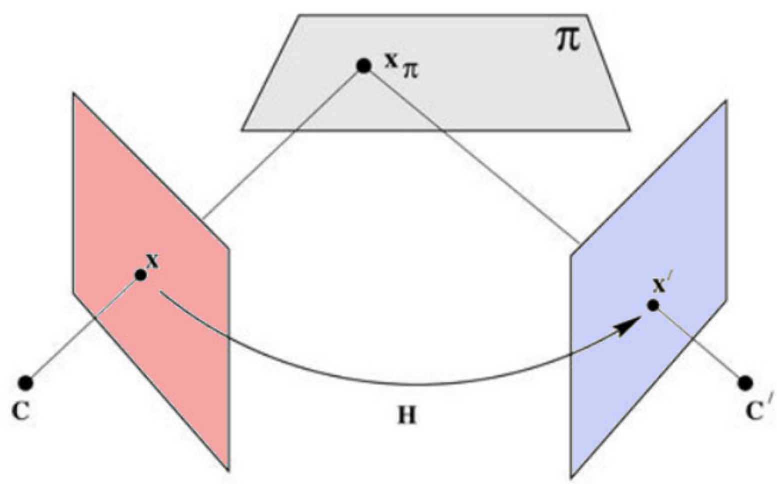
115. Zhang is titled "Distinguishing Live Faces from Flat Surfaces," and discloses a biometric authentication process using a device (e.g., a computer or smartphone) that captures images with a camera. *Zhang*, Title, Abstract, [0012], [0016].

116. Similar to Derakhshani, Zhang discloses that the biometric authentication process incorporates anti-spoofing countermeasures to ensure the user is presenting a real, three-dimensional face rather than a picture of a face for authentication. *See, e.g.*, *Zhang*, [0016]-[0017]. For this reason, I consider Zhang analogous prior art to the Challenged Patents because they are in the same field of endeavor (anti-spoofing countermeasures for biometric security).

117. One of the ways Zhang distinguishes three-dimensional faces from two-dimensional pictures of a face is by using a "homography based technique," which utilizes a well-known relationship that "two views of a flat (planar) surface

are related based on a homography matrix." *Zhang*, [0024]-[0025]. I will also refer to this as a "homography transformation."

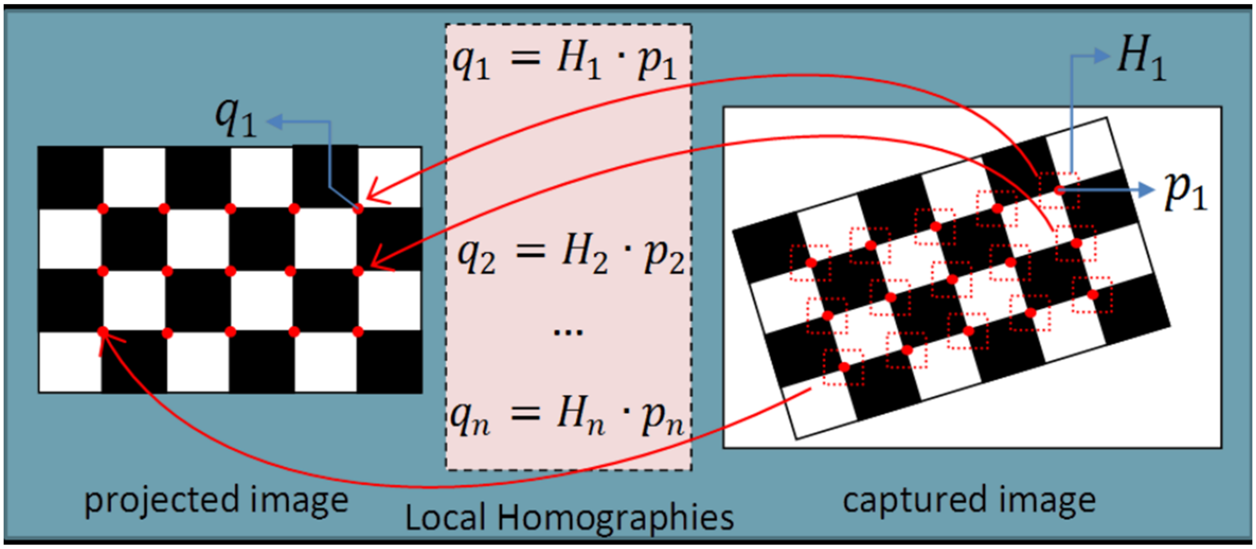
118. Specifically, when two images of *a planar* object are captured from different perspectives—e.g., when multiple cameras capture the same object from different positions, or a single camera captures the same object from different positions—a mathematical relationship exists between different points in the two images. *Id.* The geometrical construction resembles Derakhshani's "parallax" approach to determining depth, but the homography matrix *transforms* one perspective into another, as long as a planar object is being imaged:



Ex-1023, 40 (Fig. 13.1) (annotations added); *see also* Ex-1025. In this exemplary diagram depicting the homography relationship, "[t]he ray corresponding to a point x is extended to meet the plane π in a point x_π; this point is projected to a point x′ in the other image. The map from x to x′ is the homography induced by the plane π." Ex-1023, 40-41.

119. A homography matrix is a mathematical matrix that defines the relationship between these two perspectives, which enables transforming an image from one perspective (red, above) into another perspective (blue, above). The homography matrix only relates images of planar objects as shown below.[3] To perform a homography transformation, Zhang discloses that, first, facial features are extracted and matched between the first and second images to serve as reference points relating the two perspectives. *Zhang*, [0027]-[0028] ("This matching of the feature points across the first and second images refers to identifying the locations of the same feature points in each of the two images."). These matched feature points serve as inputs to generate the "homography matrix." *Zhang*, [0029]. For instance, in the example figure below, both the "captured image" points (p_1) and homography matrix (H) are both already known, which can be used to calculate "projected image" points (q_1). But if instead, like Zhang, two images from different perspectives are already known with known points (p_1 and q_1), you can use those points to calculate the homography matrix (H) to characterize how the two perspectives relate.

[3] For non-planar objects, a more general relation than homography is needed. This is discussed in Hartley's textbook (Ex-1023). The epipolar geometry which relates the pixel values of two camera images taken at different distances from an object such as a face is derived in chapter 9, at pages 4-39. Hartley's treatment of homography is in chapter 13 at pages 40-69.

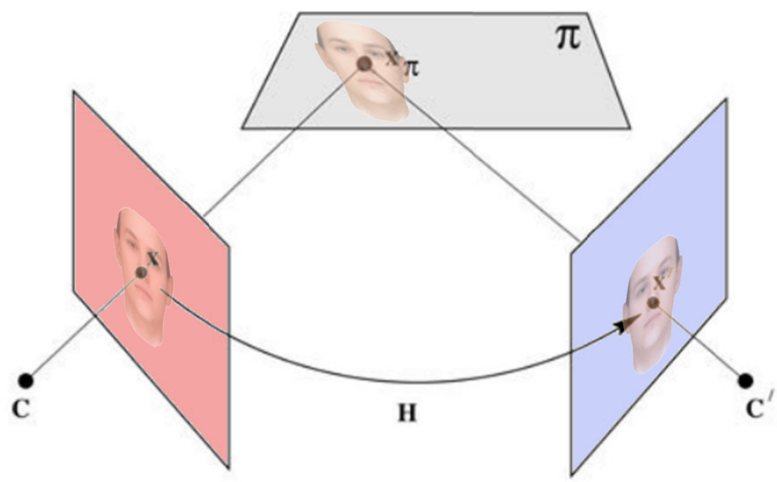


Ex-1024.

120. Once the homography matrix is calculated from the selected feature points, the matrix is then applied to every pixel in the first image to create "[a] warped image" that matches the perspective of the second image. *Zhang*, [0030]. The "warped" (transformed) image is compared to the second image. *Id.*, [0031]-[0032]. If the differences between the "warped" (transformed) image and second image meet a threshold, the face is determined to be a "live," three-dimensional face. *Id.*, [0034]. Otherwise, the face is determined to be an "imposter," two-dimensional image of a face. *Id.*

121. Ultimately, Zhang's homography transformation is looking for a two-dimensional picture because of the assumption that the object captured from different perspectives exists on a two-dimensional plane that lacks depth. A picture of a face will be consistent with this assumption, because it is restricted to the two-

dimensional plane of the paper that the face is printed on. Thus, performing a homography transformation on the picture of a face should produce a near-identical image no matter into which perspective it is transformed:



Ex-1023, 40 (Fig. 13.1) (annotated)

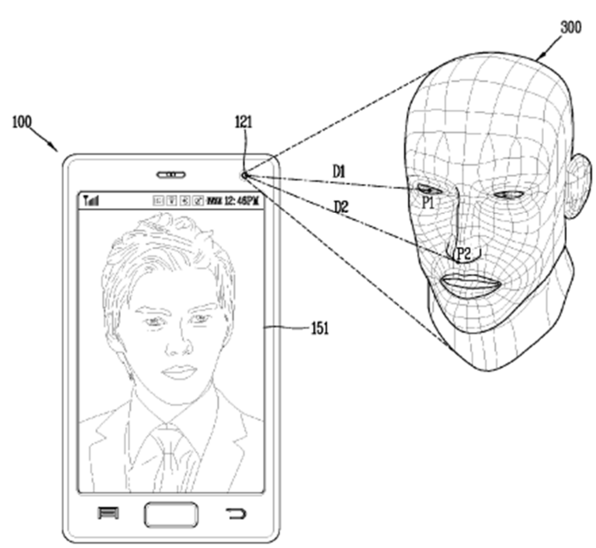
A three-dimensional face, however, does not exist on a single plane (e.g., the nose is closer to the camera than the ears), and therefore performing a homography transformation using a set of pictures of a three-dimensional face from different perspectives would result in various distortions (e.g., the ears would distort relative to the nose) when an image of a live face is transformed from one perspective to another. In sum, there would be a *mismatch* between Zhang's "warped" (transformed) first image and the second image, indicating the face has three-dimensionality.

D. Tahk (Ex-1008)

122. Tahk is a U.S. Patent Application Publication (No. 2014/0028823) that was filed on May 22, 2013 and published on January 30, 2014. I therefore understand that Tahk constitutes prior art under 35 U.S.C. §102(a)(2).

123. Tahk discloses a facial-recognition procedure for a mobile terminal (e.g., a cellphone) that captures at least two images of a user's face at different distances, and uses the "stereoscopic shape" of the user's face from those two images to distinguish between live, three-dimensional faces, and two-dimensional pictures of a face. *Id.*, Abstract, [0023], [0117], [0122], [0130]-[0131]. For this reason, I consider Tahk analogous prior art to the Challenged Patents because they are in the same field of endeavor (anti-spoofing countermeasures for biometric security).

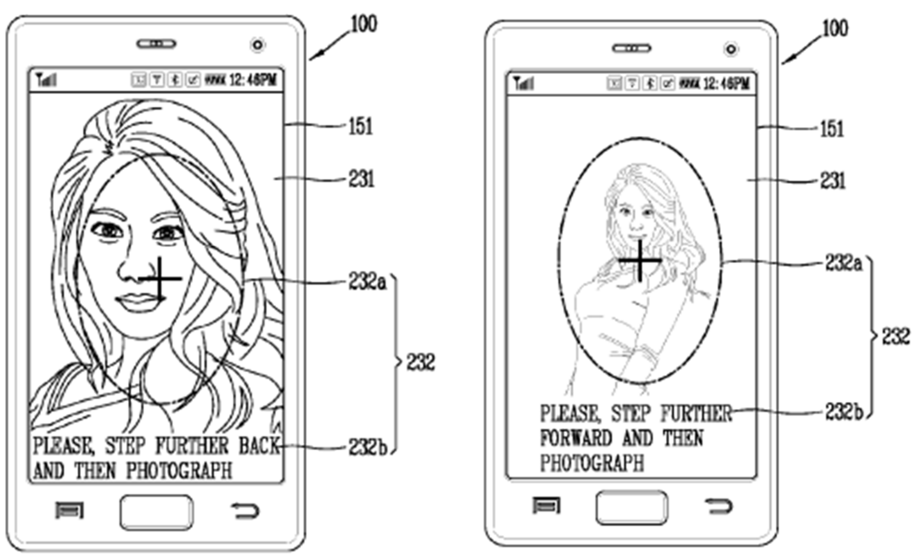
FIG. 5



Tahk, Fig. 5.

124. Tahk is unique in that it describes a process to walk a user through the facial authentication process in more detail than other prior art. For instance, Tahk discloses not only presenting a live image of the user so that they can preview what the image will look like before taking it, but also providing users different types of prompts—such as written instructions or oval-shaped overlays—to ensure the user's face is appropriately distanced from the camera for image capture.

FIG. 8A FIG. 8B



See, e.g., *id.*, Fig. 8A-B, [0118], [0129], [0135], [0139], [0143], [0144].

E. Suzuki (Ex-1009)

125. Suzuki is a U.S. Patent Application Publication (No. 2004/0239799) that was filed on May 25, 2004 and published on December 2, 2004. I therefore understand that Suzuki constitutes prior art under 35 U.S.C. §102(a)(1) and (a)(2).

126. Suzuki recognizes an available synergy in mobile devices that incorporate user-facing cameras. Specifically, cameras often utilize an illumination or "flash" component to provide additional lighting when capturing images of objects or scenes to improve the camera's performance. *See, e.g.*, *Suzuki*, [0005]. But Suzuki recognized that, because mobile devices must remain compact, when a user-facing camera is provided on a mobile device with a display, such as a cell phone, the mobile device's *display* could act as that light source. *See id.*, Abstract, [0005]-[0006]. Suzuki then describes a mobile device with a display and user-facing camera, in which the display "can be controllably switched between a display function and an illuminating function for illuminating a subject for use with the camera unit. *Id.*, [0009], [0019], [0021], [0024]-[0025], [0041]-[0045].

F. Hoyos (Ex-1010)

127. Hoyos is a U.S. Patent Application Publication (No. 2010/0014720) that was filed on October 2, 2007 and published on January 21, 2010. I therefore understand that Hoyos constitutes prior art under 35 U.S.C. §102(a)(1) and (a)(2).

128. Hoyos discloses a facial authentication system that includes a liveness check to ensure the user is a real person rather than a spoofer presenting a picture of an authorized user's face. *See Hoyos*, Abstract, [0009]-[0011]. To perform a liveness check, Hoyos discloses displaying one or more images on a device's screen—embodied as a particular black-and-white pattern—capturing an image of the user,

and evaluating the captured image to determine whether the image has a reflection of the displayed black-and-white image. *Id.*, [0018]-[0019], [0033]-[0036].

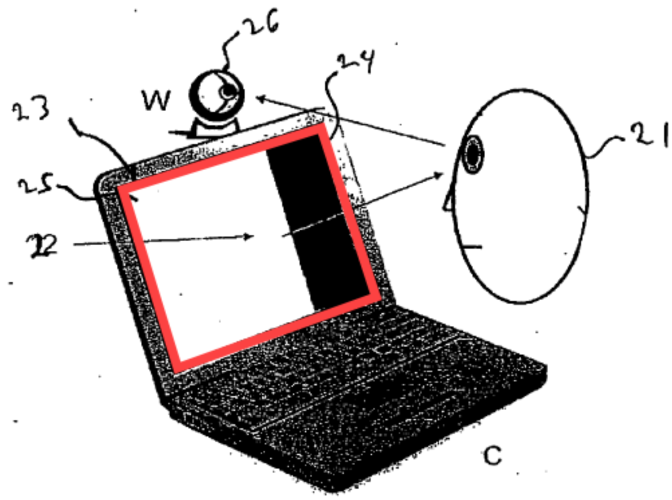


Fig. 2

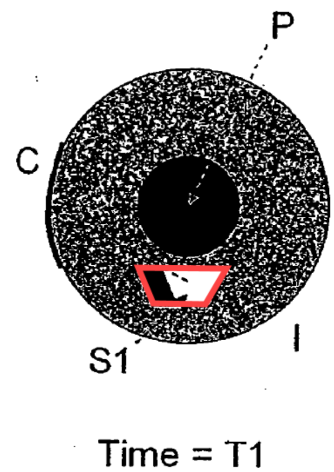


Fig. 3

Id., Figs. 2-3 (annotated). Figure 3 above, for instance, shows the reflection of the displayed image on the user's eye.

VIII. LEVEL OF ORDINARY SKILL

129. I understand that when both interpreting the claims and assessing the prior art, I must do so from the perspective of POSITA as of the effective filing date. My understanding is that the earliest-possible effective filing date of the Challenged Patents is August 28, 2014.

130. I understand there are multiple factors relevant to determining the level of ordinary skill in the pertinent art, including (1) the levels of education and

experience of persons working in the field at the time of the invention; (2) the sophistication of the technology; (3) the types of problems encountered in the field; and (4) the prior art solutions to those problems.

131. In my opinion, at that time, a POSITA in August 2014 would have had a Bachelor's degree in electrical engineering, computer engineering, computer science, physics, or a related field, and two years' work experience related to biometrics, facial authentication, computer vision, and/or optics, such that they would have had significant academic and/or work experience in both software development and optics. However, formal education would have been a substitute for work experience and relevant work experience could substitute for formal education.

132. I was at least a POSITA as of the earliest-possible effective filing date of the Challenged Patents, and I still am.

IX. SUMMARY OF OPINIONS

133. For the reasons I explain below, it is my opinion that the Challenged Claims of each of the Challenged Patents are disclosed in and rendered obvious by the prior art, according to the charts provided below:

’471 PATENT		
Ground (all under 35 U.S.C. §103)	**Claims**	**References**
1A	1-13, 15-17, 19-20	Derakhshani, Tanii
1B	14	Derakhshani, Tanii, Tahk
1C	18	Derakhshani, Tanii, Hoyos
2A	1-3, 5, 7-12, 16-17, 19-20	Zhang, Tanii
2B	4, 6, 13-15	Zhang, Tanii, Tahk
2C	18	Zhang, Tanii, Hoyos

’606 PATENT		
Ground (35 U.S.C. §103)	**Claims**	**References**
1A	1-4, 6-7, 9-16, 18-20	Derakhshani, Tanii
1B	5, 8, 17	Derakhshani, Tanii, Tahk
2A	1-3, 9-12, 14, 16, 18-20	Zhang, Tanii
2B	4-8, 13, 15, 17	Zhang, Tanii, Tahk

’938 PATENT		
Ground (all under 35 U.S.C. §103)	**Claims**	**References**
1A	1-10, 12-24	Derakhshani, Tanii
1B	11	Derakhshani, Tanii, Tahk
2A	1-3, 5-9, 12-14, 16-20, 22-24	Zhang, Tanii
2B	4, 10-11, 21	Zhang, Tanii, Tahk
2C	15	Zhang, Tanni, Suzuki

’910 PATENT		
Ground (35 U.S.C. §103)	**Claims**	**References**
1A	1-13, 15-24	Derakhshani, Tanii
1B	14	Derakhshani, Tanii, Tahk
2A	1-3, 5-12, 15-17, 19-24	Zhang, Tanii
2B	4, 13-14	Zhang, Tanii, Tahk
2C	18	Zhang, Tanii, Hoyos

X. ’471 PATENT: DETAILED EXPLANATION OF GROUNDS

A. Ground 1A: Derakhshani and Tanii (Claims 1-13, 15-17, 19-20)

1. Motivation to Combine

134. In my opinion, a POSITA would have been motivated to combine Derakhshani and Tanii because both concern identifying and accounting for the three-dimensional nature of a face when capturing an image. They differ, however, in what principles are used to account for the face’s three-dimensionality. Derakhshani, for instance, uses changes in focus distance (e.g., image resolution for structures imperfectly in focus) and/or parallax effect to determine whether a face has depth. *See* §VII.A (Derakhshani). And although Tanii is not expressly directed to *evaluating* whether a face has depth, Tanii exemplifies the well-known distortions caused by the interaction between the camera’s lens and the three-dimensional nature of the face at different distances. *See* §VII.B (Tanii). A POSITA would have appreciated, therefore, that Tanii recognizes another alternative to evaluating the depth of a face, consistent with Derakhshani’s existing two approaches.

135. A POSITA would have recognized, for instance, that Derakhshani’s focus-distance approach and Tanii’s evaluation of distance-induced distortions are both attributable to classical optical effects such as refraction and diffraction caused by (among other factors) different distances between the camera and the object(s) being captured. *Derakhshani*, 16:57-60 (“Degree of focus is a measure of the extent to the image of the landmark is blurred by optical effects … (e.g., due to *diffraction*

and convolution with the aperture shape."); *Tanii*, [0048] (noting the "unnatural image" is caused by the angles of the face relative to the angle of the camera lens).

136. Derakhshani and Tanii differ, however, in the type of effect that is occurring. Specifically, Derakhshani takes advantage of the blurring of objects that are at distances *other than* the camera's focal plane (referred to by photographers as a "bokeh effect"), which makes those objects appear unfocused. *Derakhshani*, 16:54-57; §VII.A (Derakhshani). By adjusting the focus distance (or position of the focal plane by moving the camera) and evaluating when objects (or features of an object) in an image are clear versus when they are blurry, distance information can be derived. *Derakhshani*, 16:51-63; §VII.A (Derakhshani).

137. Tanii is more specifically concerned with a type of radial distortion that arises due to the interaction of certain (e.g., wide-angle) lenses and the three-dimensional nature of the face. §VII.B (Tanii). As Tanii explains, the convex shape of a three-dimensional face, when placed near the lens, exacerbates this type of distortion. *Tanii*, [0048]; §VII.B (Tanii). Thus, particularly when a camera incorporates a wide-angle lens, images of a face close to the camera will exhibit significant radial distortion in-part because of the distances between different facial features and the lens, and in-part because the face occupies both the center and the periphery of the camera's field of view so differences in radial distortion are more apparent. *Tanii*, [0047]; §VII.B (Tanii). But when the face is further from the camera

and occupies less of the image, the distortion will be less apparent because the face is more centered on the region of the lens where radial distortion is not as severe, and there is sufficient distance for the light rays from the face to strike this central portion of the lens. *Tanii*, [0047]; §VII.B (Tanii).

138. In my opinion, a POSITA would have appreciated that when evaluating multiple images taken at either different *focus* distances or *actual* distances, these different effects serve to provide information about an object's depth. In other words, a POSITA would have understood that Tanii teaches another obvious alternative to Derakhshani's existing two approaches to evaluate whether a face being captured is three-dimensional or not.

139. That said, a POSITA would have also had specific reasons to substitute Derakhshani's existing approaches with Tanii's distance-induced distortion analysis in certain circumstances. A POSITA would have understood, for instance, that implementing Derakhshani's focus-distance approach requires a camera with a sufficiently sized sensor and lens that could provide enough sensitivity to distinguish small differences in depth on the scale of a few centimeters when trying to evaluate the depth of a face. *See Derakhshani*, 16:48-51; Ex-1029, 3 (A 200mm lens focused at 12ft will have a smaller depth of field compared to a 20mm lens focused at 12ft).

140. But a POSITA would have also understood that the cameras typically found in mobile devices—especially around the 2014 timeframe—do not have this

ability; mobile devices typically incorporate wide-angle lenses to capture a wide field of view, with a fixed focal length and a large depth of field because of their small size. *Tanii*, [0007]; Ex-1030 ("Other features of a smartphone are obvious but worth stating, they almost always are fixed focal length, fixed aperture, with no shutter, sometimes with an ND filter (neutral density) and generally not very low F-number. In addition to keep modules thin, focal length is usually very short, which results in wide angle images with lots of distortion."). With such limited-capability cameras, it was known that distortions would therefore largely be a product of the lens shape and distance between the object and the lens. *See* Ex-1017, 177 ("The amount of spherical aberration, when the aperture and focal length are fixed, varies with both the ***object distance*** and the lens shape."). In other words, there is not enough room in mobile devices to incorporate large image sensors with small F-numbers (a measure of light-gathering ability of the camera) to allow these cameras to fine-tune the focus distance and induce blurring of out-of-plane objects. That is why, for instance, the iPhone introduced its "Portrait Mode" (in 2016, a few years after the earliest possible effective date) as a software-based *simulation* of the blurring effect that can only be achieved by much larger cameras. Ex-1031 (noting how blurring backgrounds was "previously only capable on DSLR cameras" prior to the iPhone's software-based "bokeh" effect).

141. For this reason, in my opinion, a POSITA would have been motivated to modify Derakhshani to capture at least two images at different *actual* distances and evaluate whether one exhibits more distance-induced distortion than the other, as suggested by Tanii. A POSITA would have been especially motivated to make this change when implementing biometric authentication in a mobile device as Derakhshani already envisions. *Derakhshani*, 5:23-26. A POSITA would have found such a modification obvious because both techniques merely involve the application of different well-known optics principles relating camera design and object's distance from the camera, and would have had a reasonable expectation of success in doing so because Tanii already taught a mechanism to identify such distance-induced distortions and thus indicate when the face being captured has depth. *See, e.g.*, *Tanii*, [0056].

142. Although Derakhshani separately discloses a process to verify the three-dimensionality of a face using parallax, in my opinion, a POSITA would have understood that evaluating for distance-induced distortion consistent with Tanii would be easier for users on a mobile device. Specifically, a POSITA would have naturally understood that mobile devices such as phones or laptops typically capture images of users at arm's length distances because that is how these devices are used (at arm's length). Moreover, a POSITA would have appreciated that facial features do not have *significant* differences in their depth (on the order of a few centimeters,

as opposed to meters between the face and a background). Thus, to evaluate for parallax at hand-held distances with suitable accuracy, a POSITA would expect that the user would need to move their device around their head, or could simulate a parallax effect by rotating their head around a stationary camera to create substantial differences in perspective and thus more parallax to more accurately verify the face as three-dimensional. But to do so would have involved moving the device out of the user's line of sight, meaning the user could not see exactly what they are capturing or know if what they were capturing is sufficient.

143. Evaluating for distance-induced distortions when the camera is held at different distances consistent with Tanii, however, could be accomplished while keeping the device directly in the user's direct line of sight, and would therefore be easier for users to verify that their face is, in fact, three dimensional. But, in my opinion, a POSITA would have also appreciated that biometric security is always subject to spoofing, and thus would have known that evaluating for distance-induced distortion consistent with Tanii could be *supplemented* by also evaluating for any parallax.

2. Independent Claim 1

a. 1[pre]: A system for authenticating three-dimensionality of a user via a user's camera equipped computing device, the computing device comprising:

144. If the preamble is limiting, in my opinion, Derakhshani discloses or suggests it.

145. Derakhshani discloses systems and methods for using a camera-equipped computing device for "biometric authentication." *See Derakhshani*, 1:11-25, 2:4-30, 5:22-27, 6:3-5, 9:10-22, 18:1-3. Although Derakhshani uses the eye as the primary means of authentication, *see, e.g.*, *id.*, Abstract, as part of the ocular-authentication process, Derakhshani also verifies that the user's face is three-dimensional by capturing multiple images of a user's face at different focus distances or from different perspectives to calculate a "spatial metric" representing the face's three-dimensionality. *Id.*, 1:11-25, 3:14-15, 16:44-18:4.

b. 1[a]: a processor configured to execute machine executable code;

146. In my opinion, Derakhshani discloses or suggests 1[a].

147. Derakhshani discloses that the invention can be implemented in computing devices such as a "smart phone, a tablet computer, a television, a laptop computer, or a personal computer" (*Derakhshani*, 5:22-27), which incorporate a processor configured to execute machine-readable code (*see, e.g.*, *id.*, Fig. 9, 2:4-12, 2:31-38, 7:15-20, 22:12-44, 23:26-37, 24:49-25:8).

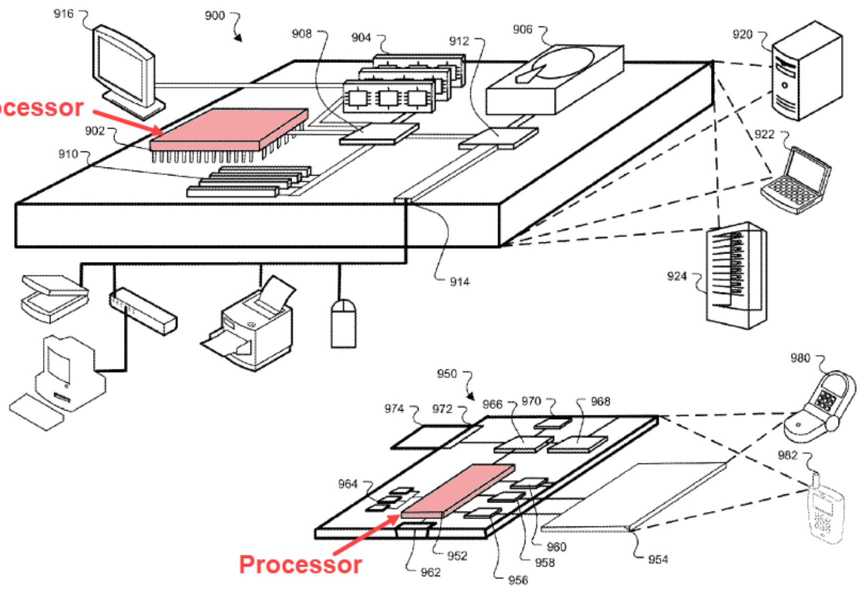


Fig. 9

Id., Fig. 9 (annotated). Derakhshani also discloses the computing device may be a server that also comprises a processor. *Id.*, 7:38-50, 8:48-9:4, 9:27-31, 10:16-19 ("the server system 514 is a data processing apparatus that includes one or more processors."), 23:14-44.

c. 1[b]: a screen configured to provide a user interface to the user;

148. In my opinion, Derakhshani discloses or suggests 1[b].

149. Derakhshani discloses that the computing device incorporates a screen to provide a user interface to the user. *See, e.g.*, *id.* 6:8-11, 9:22-24, 14:35-37, 22:33-38 23:48-52.

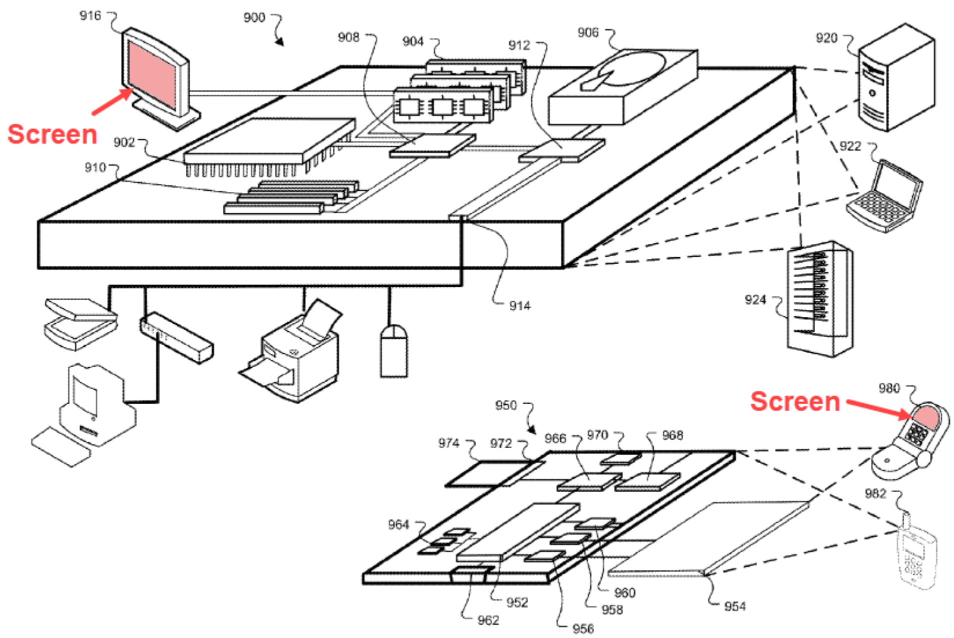


Fig. 9

Id., Fig. 9 (annotated). Derakhshani also discloses the computing device may be a server that also comprises a screen. *Id.*, 23:14-44, Fig. 9.

d. 1[c]: a camera configured to capture images;

150. In my opinion, Derakhshani discloses or suggests 1[c].

151. Derakhshani discloses that the computing device incorporates a camera configured to capture images. *See e.g.*, *id.*, 5:23-27, 6:3-10.

e. 1[d]: one or more memories configured to store machine readable instructions that are stored on **the memory of the authentication server which when executed by the processor, cause the computing device to:**

152. In my opinion, Derakhshani discloses or suggest 1[d].

153. As I mentioned previously, in my opinion, it is somewhat unclear which structure is intended to be the "computing device": (1) either a user-facing computing device that engages with a back-end authentication server; or (2) the authentication server itself. *See* §VI.A.3 (claim construction). However, in my opinion, Derakhshani discloses the claims under either interpretation.

154. First, Derakhshani discloses computing devices (e.g., a personal computer or phone) with memory that stores machine-readable instructions that are executed by the processor. *Derakhshani*, 2:4-12, 2:31-38, 22:26-44, 24:49-25:8.

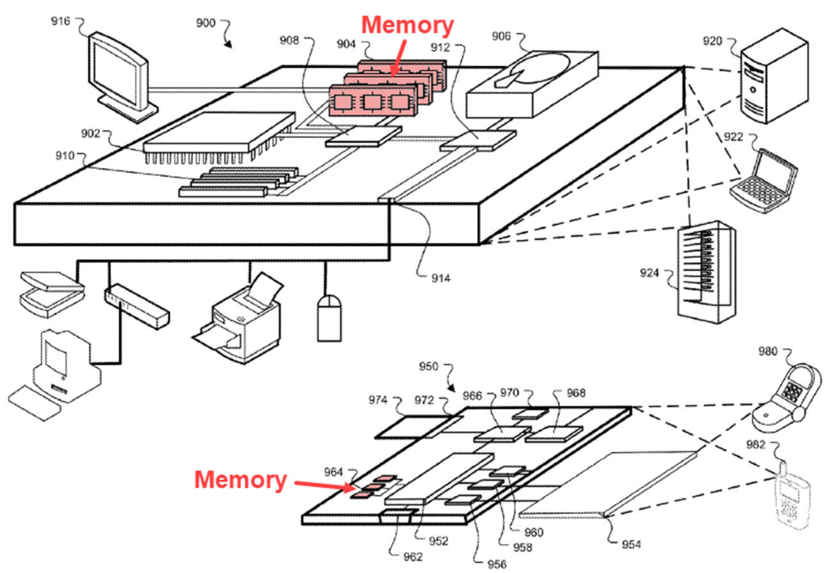
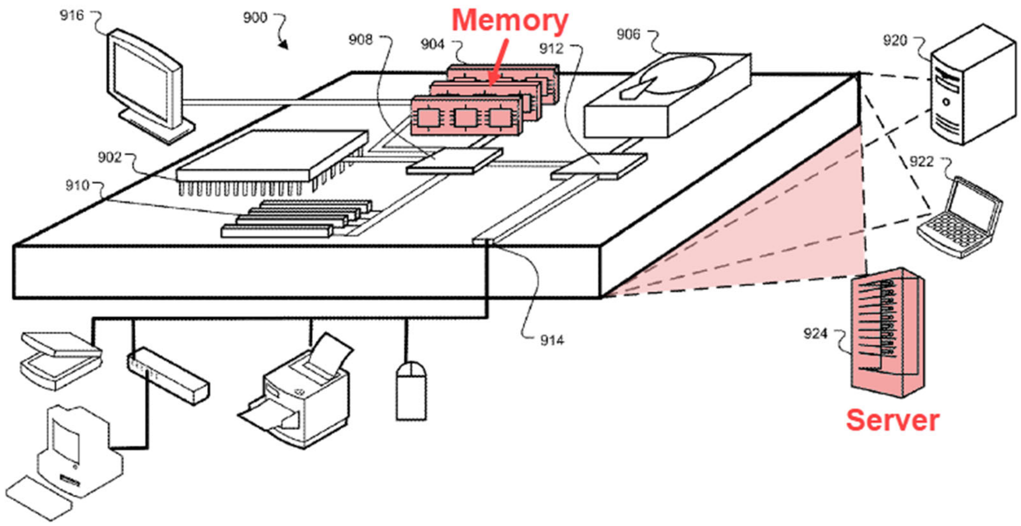


Fig. 9

Id., Fig. 9 (annotated). Derakhshani also discloses different embodiments of networked authentication systems, including: (1) a user-facing computing device that interacts with a "secure transaction service 523 hosted, e.g., by [a remote] server system" with "an authentication module 525 that coordinates authentication of users from the secured server's side of the interaction" (*Derakhshani*, 8:29-39); and (2) a user-facing computing device that hosts a local application that interacts with a remote authentication server (*id.*, 9:10-34). And Derakhshani discloses more generally that "authentication functions may be distributed between the client and the server side processes in a manner suited [to] a particular application." *Id*., 9:27-58, 10:1-24.

155. In my opinion, a POSITA would have understood that Derakhshani teaches different system configurations, including one in which a remote authentication server drives all aspects of the authentication process for a user-facing device such that the server directs the user-facing device to carry out certain aspects of Derakhshani's procedure (e.g., user image capture) using local memory that stores machine readable instructions originating from the memory of the authentication server.

156. Second, Derakhshani alternatively discloses that the computing device itself can be a server containing the components depicted in Figure 9 (such as the display 916 and processor 902). *Derakhshani*, 22:12-18, 23:14-25.



Id., Fig. 9 (annotated)

157. Although Derakhshani does not show a camera as part of the server system, a POSITA would have understood that Derakhshani at least contemplates the server itself needing biometric protection. A POSITA would have understood that, just like personal computers, servers were known to store sensitive information—from user profiles for websites, employment or medical records, and more. A POSITA seeking to prevent unauthorized access to reconfigure servers or access their files would have therefore understood that the server itself may be provided with biometric authentication—as Derakhshani at least suggests, *Derakhshani*, 22:12-18, 23:14-25—and that in such cases it would include a camera (1[c]) to carry that authentication out.

f. 1[d1]: capturing at least one first image of the user taken with the camera of the computing device at a first location which is a first distance from the user;

158. In my opinion, Derakhshani discloses or suggests 1[d1].

159. Derakhshani discloses that, as part of the process to verify that the face is in fact three-dimensional, "two or more images of a subject" are captured using the camera of the computing device. *Derakhshani*, 1:44-46, 16:44-17:11, 17:45-18:4.

160. In my opinion, a POSITA would have understood that Derakhshani captures an image at a first distance. *Derakhshani*, 16:44-17:11. Specifically, in my opinion, a POSITA would have understood that there *must* be some distance between the camera and the face to capture the first image—enough so that the camera's field of view encompasses the face. If there were no distance between the camera and the face (e.g., if the camera were pressed up against the user's skin), then the field of view would be limited to just that patch of skin and the face would not be captured in the image, which would be useless for the three-dimensional verification process.

g. 1[d2]: processing the at least one first image or a portion to create first data;

161. In my opinion, Derakhshani discloses or suggests limitation 1[d2].

162. Derakhshani discloses that, as part of the process to verify that the face is in fact three-dimensional, "a landmark (e.g., an iris, an eye corner, a nose, an ear,

or a background object) may be identified and located in the plurality of images." *Derakhshani*, 16:44-54 (focus-distance approach), 17:45-64 (parallax approach).

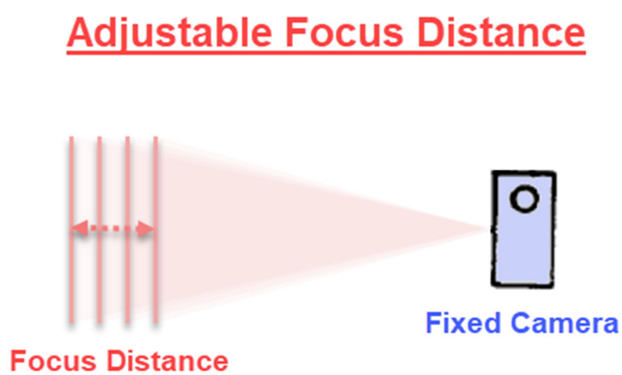
163. In my opinion, a POSITA would have understood that Derakhshani's identification of facial landmarks constitutes data—and more specifically biometric data—because the identification involves using a computer (which operates on data) to characterize the unique physical characteristics of an individual, which would include the positions of "landmarks" such as a user's eyes, nose, ears, and other such features. *See, e.g.*, Ex-1018, 2 ("biometric recognition can be defined as the science of establishing the identity of an individual based on the physical and/or behavioral characteristics of the person."), 100-103 (describing the types of biometric data about a face used for facial-recognition systems, "such as the structure of the face components (e.g., eyes), [and] the relationship between facial components").

h. 1[d3]: moving the camera from the first location to a second location, the second location being a second distance from the user, or the user moving from the first location to the second location to change the distance between the user and the camera from the first distance to a second distance;

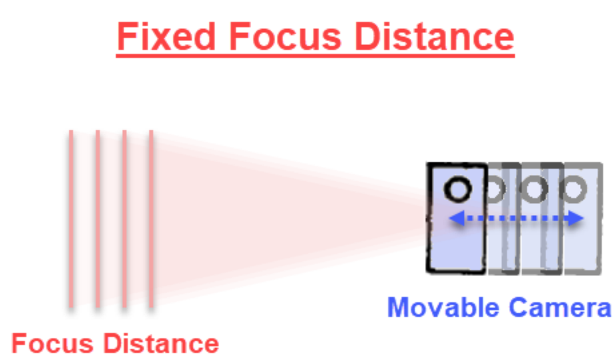
164. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 1[d3].

165. Derakhshani discloses capturing "two or more images of a subject" using the camera. *Derakhshani*, 1:44-46, 16:44-17:11, 17:45-18:4; §X.A.2.f (1[d1]). When utilizing Derakhshani's focus-distance approach to evaluate depth, however,

a POSITA would have understood that adjusting the *focus* distance of the camera does not require changing the *actual* distance between the camera if a stationary camera is capable of adjusting its lens position with respect to the image sensor. *See* §VII.A (Derakhshani explaining operation of the focus-distance approach).



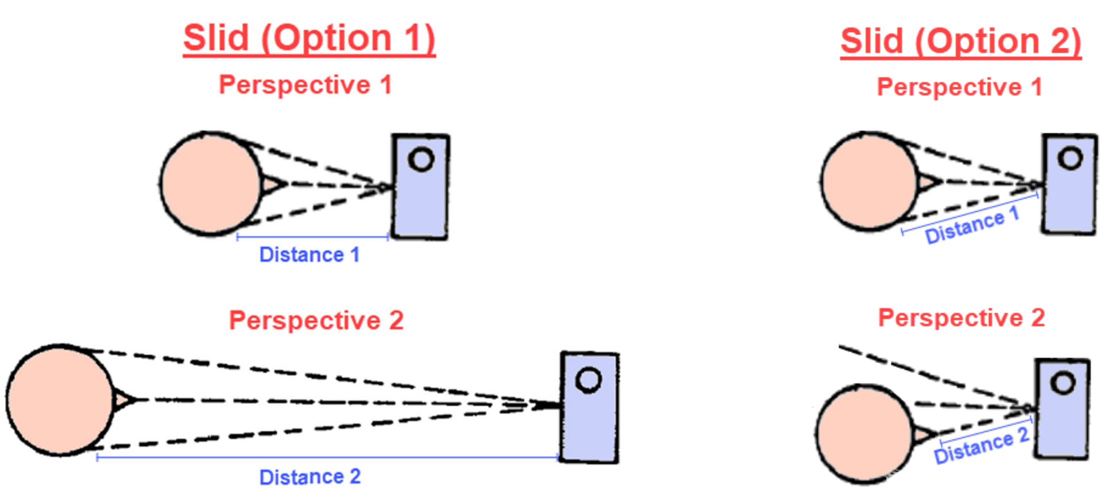
But if the camera has a fixed focus distance (i.e., position of the lens with respect to the image sensor), as is found in many mobile devices (*see* §X.A.1), a POSITA would have been motivated to instead implement Derakhshani's focus-distance approach by changing the *actual* distance to capture multiple images, as shown below:



In other words, even if the focus distance of the camera cannot be changed, the "slices" of a face at different depths can be evaluated by moving the camera.

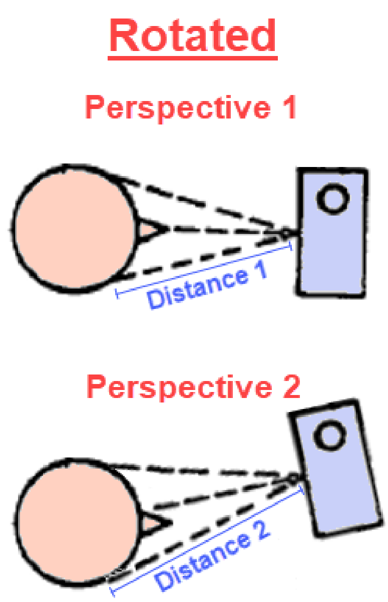
166. Regardless, in my opinion, a POSITA would have understood that Derakhshani's parallax approach captures multiple images from multiple distances, because Derakhshani discloses that "[a] plurality of images [are] taken from different perspectives on the subject," such as: (1) when "a single camera [is] rotated *or slid* slightly"; (2) "a user is prompted to move" between image captures; or (3) the sensor moves naturally, such as "where the sensor is a camera in a hand-held user device (e.g., a smartphone or tablet) [that] may naturally move relative to the users face due to involuntary haptic motion." *Derakhshani*, 17:45-18:4.

167. In my opinion, a POSITA would have understood that Derakhshani's use of the term "slid" means either of two things: (1) the camera is displaced front-to-back to increase or decrease the distance from the face; or (2) the camera is displaced side-to-side, both of which are depicted below:



In either case, a parallax effect would be evident if the face were three-dimensional because of the different perspectives of the face captured in each. For instance, a POSITA would have recognized that, with a front-to-back translation, more of the periphery of the face would be captured by the camera, and there may be other optical effects (e.g., distance-induced distortion) that are more apparent in the closer image than the further one. And with side-to-side translation, more features on the side of the face the camera favors would be captured, but features on the other side of the face may be obstructed due to the face's three-dimensionality.

168. Moreover, a POSITA would have understood that Derakhshani's use of the term "rotated" means the camera itself is rotated relative to the face. I have provided an example of rotation below that also includes some side-to-side translation to keep the face centered on the camera.



169. As these exemplary figures demonstrate, however, a POSITA would have understood that, regardless of whether the camera is "slid" or "rotated," distances between facial landmarks and the camera will change. In my opinion, a POSITA would have understood that any of these options results in "capturing at least one second image of the user … at a second distance from the user, the second distance being different than the first distance," as claimed, because there is no one single "distance" between the camera and a three-dimensional user when changing the position/perspective of the camera; some distances will always change. However, even if the claims were limited to a front-to-back translation to change the *overall* distance between the camera and the user, a POSITA would have understood that Derakhshani discloses or suggests as much.

170. But even if Derakhshani does not expressly disclose taking two images at different distances, in my opinion, a POSITA would have been motivated to look to the differences in degree of distance-induced distortions exemplified by Tanii as an alternative or additional evaluation of the three-dimensionality of the face besides Derakhshani's focus-distance and parallax approaches. §X.A.1 (motivation). When making this modification, a POSITA would have been motivated to modify Derakhshani in view of Tanii to expressly capture a second image at a second distance, and look for more distance-induced distortions in one image compared to the other to determine whether the face has depth. §X.A.1 (motivation). Moreover,

a POSITA would have had a reasonable expectation of success in making this modification because Tanii already taught a mechanism to identify such distance-induced distortions and thus indicate when the face being captured has depth. *See, e.g.*, *Tanii*, [0056].

171. Accordingly, because Derakhshani, whether alone or combined with Tanii, teaches capturing a plurality of images at different distances between the user and the camera, in my opinion, a POSITA would have understood that either the camera or user must move relative to the other in between image captures; there are no other ways to change the distance between the two. Although either option, in my opinion, would have been obvious, a POSITA would have understood that Derakhshani expressly recognizes moving the camera relative to the user when the camera is part of a portable computing device. *Derakhshani*, 17:45-18:4. In such instances, a POSITA would have understood that moving a camera in relation to the user's face is a convenient and obvious option for changing the distance between the user's face and the camera.

i. **1[d4]: capturing at least one second image of the user taken with the camera of the computing device** at **the second distance from the user, the second distance being different than the first distance;**

172. In my opinion, Derakhshani, alone or in combination with Tanii, teaches 1[d4]. *See* §X.A.2.h (1[d3]).

j. 1[d5]: processing the at least one second image or a portion thereof to create second data;

173. In my opinion, Derakhshani discloses or suggests 1[d5].

174. Derakhshani discloses processing the captured images to identify biometric "landmarks" in the face as part of the three-dimensional verification process. *Derakhshani*, 17:45-52; §X.A.2.g (1[d2]).

k. 1[d6]: comparing the first data to the second data to determine whether expected differences exist between the first data and the second data which indicated three-dimensionality of the user;

175. In my opinion, Derakhshani, alone or in combination with Tanii, teaches 1[d6].

176. Derakhshani discloses that, regardless of whether the focus-distance approach or parallax approach is used, biometric features are identified and compared across each of the images. *Derakhshani*, 16:66-17:2 ("comparing the degree of focus for a landmark in images with different focus distances."); 17:45-64 (evaluating relative displacement of identified landmarks across images). In my opinion, a POSITA would have also appreciated that, when modifying Derakhshani to evaluate for distance-induced distortions exemplified by Tanii, biometric data would also be compared across images to then determine whether they exhibit distance-induced distortion relative to each other, consistent with Derakhshani. In other words, a POSITA would have understood that each of the approaches to

evaluate an object's depth taught by Derakhshani (focus distance or parallax), whether alone or combined with Tanii (distance-induced distortion), would require comparing biometric data points across multiple images. For this reason, a POSITA would have known, or at least been motivated to, match the biometric data between each of the images (such as matching the ears, eyes, and nose in one image to those same features in another) to evaluate the differences between them in different images.

177. Derakhshani discloses that, when comparing the first biometric data to the second biometric data, a determination is made whether differences between the two exist. *Derakhshani*, 16:66-17:2 (for focus distance, "[b]y comparing the degree of focus for a landmark in images with different focus distances, the distance from the sensor to the landmark may be estimated."), 17:55-59 (for parallax, "[i]f all the landmarks in the image undergo the same apparent displacement due to the relative motion of the sensor…then the subject viewed by the camera has a likelihood of being a two-dimensional spoof attack.").

178. In my opinion, Derakhshani describes a comparison between images that looks for "expected" differences consistent with how the '471 Patent uses the term because one would *expect* that following either the focus-distance or parallax approaches Derakhshani discloses would produce specific differences: the focus-distance approach would capture some images where certain facial features are

blurred and others where those same features are clear, and the parallax approach would produce expected relative displacements of certain facial features depending on the change of perspective and distance between the specific features and the camera lens. Moreover, Derakhshani's focus-distance approach looks for expected differences in the blurriness or clearness of facial landmarks by changing the *actual* distance (for fixed-focus cameras), and Derakhshani's parallax approach looks for expected differences in the relative displacement of different facial landmarks by changing the *actual* distance alone. *See* §X.A.2.h (1[d3]).

179. Relatedly, a POSITA would have been particularly motivated to configure Derakhshani to capture images with specific, pre-defined configurations (e.g., a specific set of focus distances, or a specific position of the camera relative to the face) to minimize the variability between the images used for facial recognition and specifically tailor the system to look for expected changes between images. For example, Derakhshani's focus-distance approach (with its loss of spatial frequency) would improve its performance if images were acquired with the face at different distances from the camera. Doing this with two or more distances would remove range ambiguity and decrease the variance in estimates of the distance from the camera to particular features. *See, e.g.*, Ex-1018, 32 (noting how facial-recognition systems often require controlling conditions such as a "fixed and simple background with controlled illumination' because "systems … have difficulty in matching face

images captured from two different views, under different illumination conditions, and at different times."). In other words, rather than permitting users to change the focus distance or perspective of the camera any way they wish, which would require a system that could account for such variabilities, having the user follow a pre-determined protocol to capture images at set focus distances or perspectives would simplify the matching process.

180. In my opinion, however, a POSITA would have also understood that, when utilizing the distance-induced distortion approach exemplified by Tanii, the images captured from that process would also exhibit expected distortion based on the distance between the camera and the face. §X.A.1 (motivation). In my opinion, a POSITA would have been motivated to look for and utilize these expected differences in distortion as an alternative or supplemental verification of three-dimensionality of a face in Derakhshani, particularly in mobile devices that incorporate wide-angle lenses. §X.A.1 (motivation). In doing so, a POSITA would have understood that verifying a three-dimensional face using distance-induced distortion would be accomplished by matching the positions of biometric features across the first and second images—as Derakhshani already discloses—but rather than look for blurriness/clearness or parallax of those biometric features, the images would instead be evaluated for expected differences in the distortion of those features caused by the distance-induced distortion. §X.A.1 (motivation).

l. 1[d7]: authenticating the user when differences between the first data and the second data have expected distortion resulting from movement of the camera from the first location to the second location or movement of the user from the first location to the second location, which causes the change in distance between the user and the camera.

181. In my opinion, Derakhshani, alone or in combination with Tanii, teaches 1[d7].

182. Derakhshani discloses a process that determines a face is three-dimensional when expected distortions exist between the biometric landmarks (e.g., data) using at least the focus-distance approach. §§X.A.2.h (1[d3]), X.A.2.k (1[d6]). Specifically, the focus-distance approach looks for expected distortion by evaluating whether facial landmarks are blurry in one image and clear in another, indicating depth. §§X.A.2.h (1[d3]), X.A.2.k (1[d6]). Moreover, Derakhshani's focus-distance approach looks for expected distortions in the blurriness or clearness of facial landmarks by changing the *focus* distance (for adjustable-focus cameras) or *actual* distance (for fixed-focus cameras). *See* §X.A.2.h (1[d3]).

183. If Derakhshani's parallax approach for some reason cannot be considered to already disclose this limitation, however, in my opinion, Derakhshani combined with Tanii does. *See* §§X.A.2.i (1[d4]), X.A.2.k (1[d6]). Specifically, in my opinion, a POSITA would have appreciated that when modifying Derakhshani to specifically look for differences caused by distance-induced distortions

(consistent with Tanii), a three-dimensional face would be indicated when one of the two sets of biometric data exhibits expected distance-induced distortions due to the change in distance of the camera. *See* §§X.A.2.i (1[d4]), X.A.2.k (1[d6]). And a POSITA would have been motivated to utilize this expected distortion as an alternative or supplemental verification of three-dimensionality of a face in Derakhshani because it provided a user-friendly way of verifying three-dimensionality using well-understood optical effects common to widely used camera systems (e.g., wide-angle lens in mobile devices). §X.A.1 (motivation).

184. Finally, it is worth noting that although the '471 Patent uses the term "authentication" in other claims (such as claims 5 and 10, which recites "enrollment data" and an "authentication session," respectively) which a POSITA would have understood refers to comparisons to enrollment data to authenticate a user's identity, 1[d7] appears to use the term "authentication" to refer to authenticating the three-dimensionality of the face, because a user would not be broadly authenticated just because they demonstrated their face is three dimensional without also authenticating their identity. Accordingly, in my opinion, the '471 Patent does not use the term "authentication" consistently to refer to comparisons to enrollment data.

185. Regardless, Derakhshani discloses a computer-implemented authentication method for authenticating the three-dimensionality of a person's face during an authentication session. *See, e.g.*, *Derakhshani*, Abstract, 1:11-2:3, 9:59-

67, 11:17-26, 11:42-45, 13:62-66, 14:45-48 (use of "liveness score" and "match score" for overall authentication), 4:32-52 (calculating "spatial metric" to authenticate three-dimensionality as part of liveliness score).

3. Claim 2

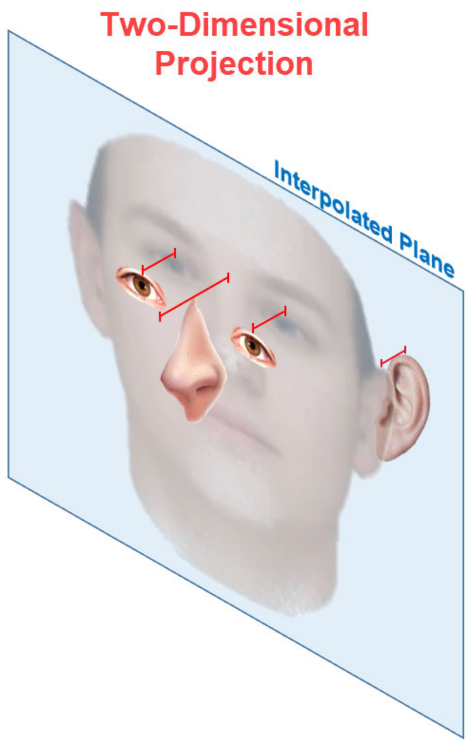
a. 2[a]: The system according to claim 1, further comprising: interpolating the first data and the second data to obtain estimated intermediate data;

186. In my opinion, Derakhshani, alone or combined with Tanii, teaches 2[a]'s additional limitation.

187. Derakhshani discloses that one optional way the three-dimensionality of a face can be verified is by fitting "the location of multiple landmarks…to the closest two dimensional plane and the average distance of the landmarks from this fit plane can be determined as the spatial metric." *Derakhshani*, 17:12-26.

188. In my opinion, a POSITA would have understood that Derakhshani identifies the position of this two-dimensional plane relative to the facial landmarks by matching up the landmarks that appear across different images, and then calculating an average distance between the various landmarks at their identified three-dimensional positions based on the series of images. If this average distance between the plane and landmarks is sufficiently large, the face is determined to be three dimensional. I have provided a graphic depiction of this process for demonstration purposes, with the plane being identified in blue, the position of the

facial landmarks determined by analyzing the series of images, and the red line between the facial landmarks and the plane representing the distance between the two:



A POSITA would have therefore understood that the plane constitutes "interpolated" biometric data, because the term "interpolated" is generally understood to mean "to insert between other things" or "estimate values of (data or a function) between two known values. Ex-1026, 654.

189. A POSITA would have further understood that a similar projection could be determined from Derakhshani's parallax process. However, rather than use a two-dimensional plane projection based on images taken from the same

perspective (like Derakhshani's focus-distance approach), images taken from two different perspectives to evaluate for parallax would be better suited by using a three-dimensional model approach. For instance, it was well-known that three-dimensional modeling of the perspective of a face could be estimated using images of a face. *See, e.g.*, Ex-1018, 117 (noting use of 2D and 3D modeling techniques to account for variations in perspective); Ex-1027, 8 (describing a method that involves estimating the position of a face to locate and match facial features); Ex-1015 (describing the generation of a three-dimensional model of a face based on two-dimensional images). In fact, Derakhshani expressly recognizes that the spatial metric can be determined by determining deviations between the images captured and a three-dimensional model of the face. *Derakhshani*, 17:27-44.

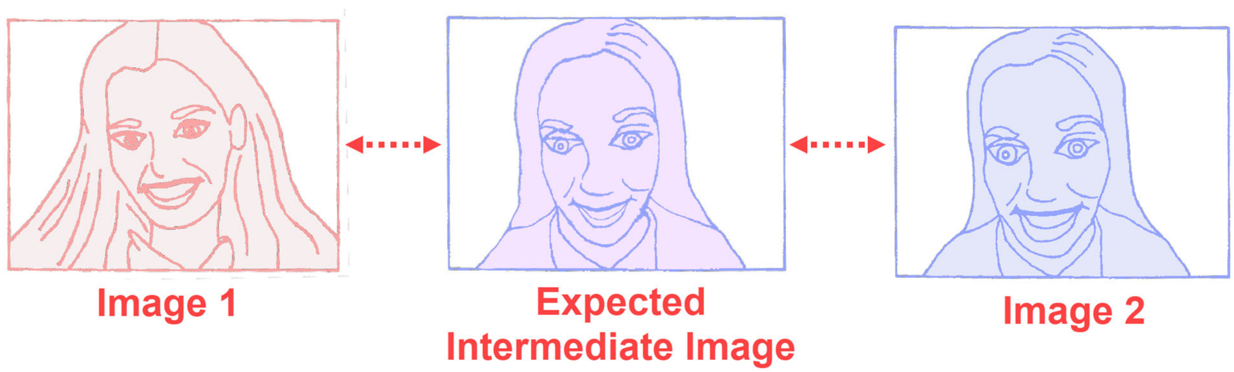
190. A POSITA would have understood that these three-dimensional models to which images are compared would be an "interpolation." In other words, a POSITA would have recognized, or at least been motivated to implement Derakhshani's parallax approach by constructing a three-dimensional, interpolated model based on the series of images captured to either: (1) compare it to an existing three-dimensional model generated during enrollment; or (2) determine whether the series of images can create a suitable three-dimensional model, which itself would indicate that the imaged face has three dimensions.

191. Based on a POSITA's understanding of Derakhshani, a POSITA would have further been motivated to derive interpolated data based on the combination of Derakhshani and Tanii. Specifically, Tanii discloses that distance-induced distortions increase as distance between the face and camera decreases. *See Tanii*, [0048]. A POSITA would have therefore understood that, all else being equal, distance-induced distortion depends on the distance between the user and the camera, and thus any set of images as the distance between the user and camera changes will have different degrees of facial warping, similar to the gradual changes in facial distortion that appears in the series of images below (although these images also altered the focal length of the camera to ensure the face remains a constant size in the frame, rather than just distance:

Ex-1022.

192. A POSITA reading Derakhshani—which discloses generating intermediate projections to evaluate depth—in view of Tanii therefore would have been motivated to interpolate intermediate data with an intermediate, interpolated

amount of distance-induced distortion based on the two images captured to create an array of potential distance-induced distortions that would indicate depth in a three-dimensional face, such a gradual projection from the lines of the facial landmarks shown in blue (the distorted image) to the lines of the facial landmarks shown in red (the distorted image).



Performing this type of interpolation between the distorted and undistorted images would allow for further comparison with additional images, for instance, to ensure the distance-induced distortion matches expectations across a range of distances if the face were truly three-dimensional, as depicted below:

In my opinion, a POSITA would have understood that this would be akin to Derakhshani's modeling approach, but rather than build a model based on a two-dimensional projected plane or three-dimensional model of a head, the model would be of various degrees of expected distance induced distortion with which the captured images could be compared.

b. 2[b]: capturing at least one third image of the user taken with the camera of the computing device at a third distance from the user, the third distance being between the first distance and the second distances;

193. In my opinion, Derakhshani, alone or combined with Tanii, teaches 2[b]'s additional limitation.

194. Derakhshani discloses that, as part of the three-dimensional verification process, "a plurality" of images may be captured. *Derakhshani*, 16:44-46 (focus-distance embodiment), 17:45-47 (parallax embodiment).

195. In my opinion, a POSITA would have also understood generally that capturing more images would provide increased accuracy in verifying a three-dimensionality of the face because there would be more samples to evaluate, with the trade-off being an increase in processing demands. For instance, taking four images using the focus-distance approach would enable precise depth information of at least four facial landmarks that sit on different planes, such as the ears, eyes, mouth, and nose. For the parallax approach, fewer images would likely be necessary depending on how significant the change of perspective is—e.g., rotating the camera

may reveal parallax in as little as two images, whereas sliding may benefit from an additional image—since that approach looks for displacement of facial landmarks due to the change in perspective, which does not depend on taking images at different "slices" of depth like the focus-distance approach.

196. A POSITA would have also appreciated as a general matter that, in any set of images with more than two images captured at different distances, *see* §X.A.2.i (1[d4]), one image would have a minimum distance and one would have a maximum distance, with the rest existing somewhere in between. For instance, in a set of distances of 10cm, 50cm, and 1m, 10cm would be the minimum, and 1m would be the maximum, with 50cm existing in between the two.

197. Moreover, when modifying Derakhshani in view of Tanii to interpolate intermediate data attributable to distance-induced distortions, a POSITA would have been further motivated to capture a third image at a distance that correlates to one of the interpolated data sets for further authentication of three-dimensional depth of the face in the captured images. *See* §§X.A.2.i (1[d4]), X.A.3.a (2[a]).

c. 2[c]: processing the at least one third image or a portion thereof to obtain third data; and

198. In my opinion, Derakhshani, alone or combined with Tanii, teaches 2[c]'s additional limitation.

199. Derakhshani discloses processing the images to identify feature landmarks in each of the images. *See* §§X.A.2.g (1[d2]), X.A.2.j (1[d5]). Therefore,

in my opinion, a POSITA would have found it obvious to process any images captured by the camera to derive biometric data so that the biometric data could be compared between images, consistent with Derakhshani.

200. Furthermore, when modifying Derakhshani in view of Tanii, in my opinion, a POSITA would have found it obvious to acquire a third image and extract data from the third image to compare it to the interpolated positions of the biometric data based on the first and second images. *See* §X.A.3.b (2[b]).

d. 2[d]: comparing the estimated intermediate data with the third data to determine whether the third data matches the estimated intermediate data.

201. In my opinion, Derakhshani, alone or combined with Tanii, teaches 2[d]'s additional limitation.

202. Derakhshani discloses comparing multiple images to the two-dimensional projection or three-dimensional model interpolated from those images. *Derakhshani*, 17:12-44; §X.A.3.a (2[a]). In my opinion, a POSITA would have therefore understood that once an interpolated projection or model of the face is generated consistent with Derakhshani, if a third image is captured, that too would be compared to the projection or model to estimate the distance or deviation of any facial landmarks in that image from the projection or model. §X.A.3.a (2[a]).

203. Furthermore, in my opinion, a POSITA modifying Derakhshani in view of Tanii would have considered it obvious to acquire a third image and extract data

from it to compare the data to the interpolated, expected positions of the data to determine if there is a match between the two. *See* §§X.A.3.a (2[a]), X.A.3.b (2[b]).

4. Claim 3: The system according to claim 1, further comprising verifying the presence of the user's ears in the at least one first image, and verifying the absence or reduced visibility of the user's ears in the at least one second image, wherein the first distance is larger than the second distance.

204. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 3's additional limitation.

205. Derakhshani discloses that "a landmark…an ear…may be identified and located." *Derakhshani*, 16:51-54; *see also* 17:14-19.

206. In my opinion, a POSITA would have appreciated that, when following Derakhshani's focus-distance approach, in some captured images, the ear would have reduced visibility (i.e., it is blurry) when it does not lie in the focal plane, and would be clear (e.g., a verified presence) when it does lie in the focal plane. *See* §X.A.2.h (1[d3]). A POSITA would have also appreciated that, in some circumstances, distances in which the ear would be clear would be greater than those with reduced visibility, such as when the focal plane is aligned behind the ears.

207. And when following Derakhshani's parallax approach, a POSITA would have also appreciated that some perspectives would obviously capture one or more ears (when both are exposed, such as a front-facing image from sufficient distance), and other perspectives would only capture one (when the other is

obstructed by the head), which would indicate that the user's face is three-dimensional. *See* §X.A.2.h (1[d3]) (providing an example figure in which a camera rotation would obfuscate one ear). If the object being captured were a two-dimensional picture of a face with ears, however, any perspective would capture both ears because the ears exist on a single plane of the picture.

208. Similarly, when modifying Derakhshani in view of Tanii to use distance-induced distortions to verify the three-dimensional nature of a face, Tanii teaches that the absence and presence of an ear is a natural result of the distance between the user and camera. Specifically, when a sufficient distance between the face and camera exists, the ears are captured because there is enough distance for the light rays from the ears to strike the camera's lens. *See Tanii*, [0048], Figs. 3A-3B.

Fig.3A

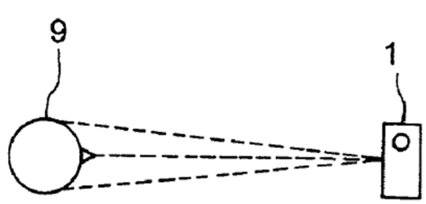
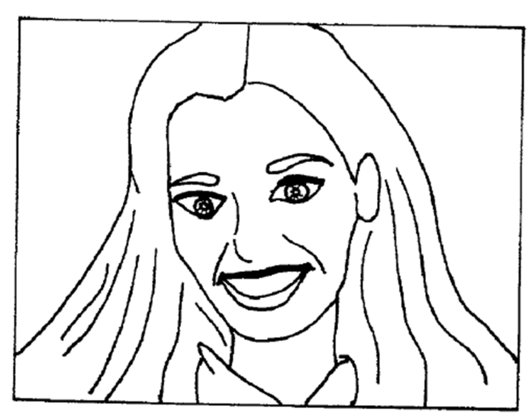


Fig.3B



But when a face is too close, the ears will not be captured because there is insufficient distance for the light rays from the ears to strike the camera's lens. *See Tanii*, [0048], Figs. 4A-4B.

Fig.4A

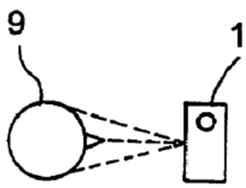
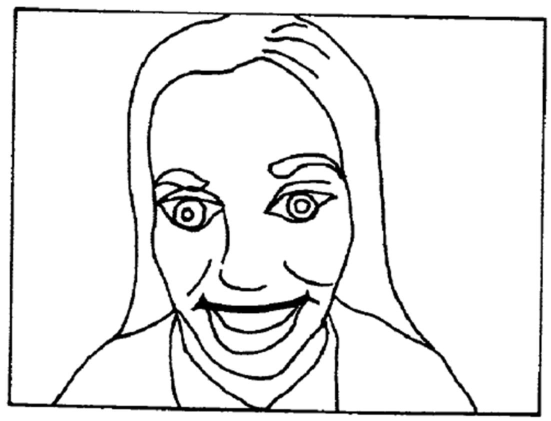
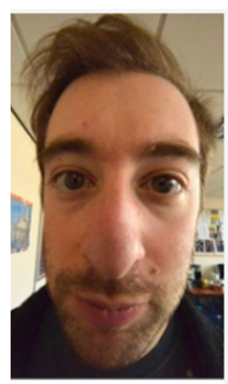


Fig.4B



This effect was well known and demonstrated in actual applications, as shown below.

85mm @ 200cm



8mm @ 20cm

209. In my opinion, therefore, a POSITA would have appreciated based on at least Tanii that the presence of a user's ears in one image at a sufficient distance, but absence of a user's ears in another image at a closer distance would be indicative

of a three-dimensional face, and would have been motivated to modify Derakhshani to verify the presence and absence of the ears between images as yet another indicator of the three-dimensional nature of the face.

5. Claim 4: The system according to claim 1, wherein the computing device is configured to display one or more prompts on a screen of the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

210. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 4's additional limitation.

211. Derakhshani discloses that the invention can be implemented in computing devices such as a "smart phone, a tablet computer, a television, a laptop computer, or a personal computer," *Derakhshani*, 5:22-27, which incorporate a camera, *id*., 5:23-27, 6:3-10, and a display. *Id*., 6:8-11, 9:22-24, 14:35-37, 22:33-38, 25:9-15. Derakhshani discloses displaying prompts to a user to guide the user to capture images of the user's face for authentication, *Derakhshani*, 5:23-32, 6:8-16, 9:22-26, including at more than once distance, *id.*, 17:64-66; §X.A.2.i (1[d4]).

212. But even if Derakhshani does not expressly disclose taking two images at different distances, doing so would have been obvious in view of Tanii to identify distance-induced distortions that indicate depth of a three-dimensional face. §X.A.2.i (1[d4]). When modifying Derakhshani to look for distance-induced distortions by capturing images at different distances consistent with Tanii, in my

opinion, a POSITA would have been motivated to provide prompts to a user to ensure the images are captured at the correct distances because Derakhshani already discloses providing prompts to correctly orient the user relative to the camera.

6. Claim 5

a. 5[a]: The system according to claim 1, further comprising comparing the first data, second data, or both to enrollment data derived from an enrollment image, the enrollment image captured and stored prior to an authenticating; and

213. In my opinion, Derakhshani discloses or suggests 5[a]'s additional limitation.

214. Derakhshani discloses capturing and analyzing multiple images of a user and comparing the user's features to a previously stored "reference record" to authenticate the user. *Derakhshani*, 4:19-24; 7:20-34; 8:60-64; 9:31-34. In my opinion, a POSITA would have understood the "reference record" to be "enrollment data" because the process Derakhshani describes to generate and then use the "reference record" for authentication is consistent with a typical biometric-authentication enrollment procedure. *See* §V.A (biometric security overview). Specifically, Derakhshani discloses that the system captures one or more initial reference images of the user during a registration process, extracts features from the reference images, stores the extracted features as the reference record, and then subsequently compares later-captured images to the reference record. *Derakhshani*,

7:19-34 ("To create a reference record *for a new user* and enrollment or registration process may be carried out."); 9:31-34 ("The collection of image data from user may also facilitate authentication against a reference record for a user identity."); 13:62-14:9 (describing authentication matching against a reference record). Then, during the authentication process, Derakhshani compares the extracted features from the captured images (i.e., portions of the first data, second data, or both) to the user's enrollment reference record to determine a match score. *Id*. 9:59-67; 13:62-14:9; 17:32-36. This is consistent with a conventional biometric enrollment and authentication process. *See, e.g.*, Ex-1018, 4-11 (providing overview of biometric authentication and verification).

b. 5[b]: only authenticating the user when the first data, the second data, or both match the enrollment data within a predetermined threshold.

215. In my opinion, Derakhshani discloses or suggests 5[b]'s additional limitation.

216. Derakhshani discloses calculating a match score during the authentication process based on the comparison of features extracted from the first and second image to the corresponding features in an enrollment reference record. *Derakhshani*, 13:62-14:9. Derakhshani also discloses that, only when the match score is above a threshold—because the first or second data, or both, sufficiently correspond to the enrollment data—it is determined the user is authenticated. *Id*.,

14:25-35. This is consistent with conventional biometric-authentication processes. *See, e.g.*, Ex-1018, 17 ("[B]iometric systems mostly decide on a person's identity based on a *close* match between the template and the query, where the strength of the match (or the degree of similarity) is represented by the match score."), 18 ("a verification system makes a decision by comparing the match score *s* to a threshold η").

7. Claim 6: The system according to claim 1, wherein the computing device is a hand-held device, and the user holds the device at the first and second distance to capture the at least one first image and the at least one second image.

217. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 6's additional limitation.

218. Derakhshani discloses that the biometric-authentication process can be implemented on a variety of different type of hand-held computing devices, such as "a laptop computer, a handheld computer…, a tablet computing device, a personal digital assistant (PDA), a cellular telephone…, a camera, a smart phone," and more. *See, e.g.*, *Derakhshani*, 8:11-28, 18:1-4. Derakhshani also recognizes that, to verify three-dimensionality of the face, "a single camera may be rotated or slide slightly," or that, when the device is hand-held, "the [camera] sensor may naturally move relative to the users face due to involuntary haptic motion" that may sufficiently capture a parallax effect. *Id.*, 17:59-18:4. Similarly, Tanii recognizes that distance-induced distortions often occur in mobile devices that have incorporated wide-angle

lenses, and the amount of distortion is dictated by the distance between the user and the camera. *Tanii*, [0007], [0047]-[0048], Figs. 3A-B, 4A-B.

219. When implementing a three-dimensional verification process on a handheld mobile computing device consistent with Derakhshani, alone or in combination with Tanii (*see* §§X.A.2.h (1[d3]), X.A.5 (claim 4)), it is my opinion that a POSITA would have further understood that the user would hold the computing device at a first distance for the first image, and a second distance for the second image (e.g., by extending and retracting the user's arm), because that is a convenient and obvious way of changing the distance between a hand-held device and the user's face, and because Derakhshani already envisions evaluating depth based on the displacement of the user's arm while holding the device. §VII.A (Derakhshani); *Derakhshani*, 16:44-11, 17:45-18:4.

8. Claim 7: The system according to claim 1, wherein the first data and the second data comprise biometric data.

220. In my opinion, Derakhshani discloses or suggests claim 7's additional limitation for the reasons discussed in §§X.A.2.g (1[d2]), X.A.2.j (1[d5]).

221. In my opinion, a POSITA would have also understood that data created from an image (or portion of an image) of a user's face would comprise biometric data. *Id.*

9. Claim 8: The system according to claim 1, wherein the first data and the second data comprise a mapping of facial features.

222. In my opinion, Derakhshani discloses or suggests claim 8's additional limitation.

223. Derakhshani discloses processing the captured images to identify and locate facial biometric "landmarks" (e.g., an iris, an eye corner, a nose, a mouth, an ear) in a three-dimensional verification process. *Derakhshani*, 16:44-54. In my opinion, a POSITA would have understood that the identification of facial landmarks would include their locations relative to one another, thus constituting a mapping of facial features. In fact, processing image data to map facial features was a conventional aspect of facial-recognition systems. *See, e.g.*, Ex-1018, 103 (Fig. 3.5(b) describing how "Level 2 features require detailed processing for face recognition. Information regarding the structure and the specific shape and texture of local regions in a face is used to make an accurate determination of the subject's identity."). Moreover, a POSITA would have understood that the process of converting facial features in an image to computer-readable data conventionally involves mapping those features to data. *See, e.g.*, *id.*, 116-17 (noting how "appearance-based techniques generate a compact representation of the entire face region in the acquired image by *mapping* the high-dimensional face image into lower dimensional sub-space.").

10. Claim 9: The method according to claim 1, wherein the first image and the second image is of the user's face and the user's face is held steady and without movement during capture of the first image and the second image.

224. Derakhshani, alone or combined with Tanii, teaches claim 9's additional limitation.

225. Derakhshani and Tanii both teach or suggest moving the camera to capture images at two different distances. *See* §X.A.2.hX.A.2.i (1[d3]).

226. In my opinion, a POSITA would have understood that, when moving the camera to capture images from different distances, the user's face would be stationary (e.g., steady) both during each image capture (to ensure each image is not blurry due to the camera's exposure time) and during movement of the camera (to isolate any differences between images to those attributable to the change of distance). Moreover, a POSITA would have appreciated that holding the user's face steady and moving the camera closer and further away would be more user friendly than requiring the user to move their head closer and further from the camera while holding the camera steady.

11. Independent Claim 10

a. 10[pre]: A method for authenticating three-dimensionality of a user via a user's camera equipped computing device, the method, during an authentication session comprising:

227. If the preamble is limiting, in my opinion, Derakhshani discloses or suggests it.

228. Derakhshani discloses systems and methods for using a camera-equipped computing device for "biometric authentication." *See Derakhshani*, 1:11-25, 5:22-27, 6:3-5, 9:10-22, 18:1-3. Although Derakhshani uses the eye as the primary means of authentication, *see, e.g.*, *id.*, Abstract, as part of the ocular-authentication process, Derakhshani also verifies that the user's face is three-dimensional by capturing multiple images of a user's face at different focus distances or from different perspectives to calculate a "spatial metric" representing the face's three-dimensionality. *Id.*, 1:11-25, 3:14-15, 16:44-18:4.

b. 10[a]: capturing at least one first image of the user taken with the camera of the computing device at a first location which is a first distance from the user;

229. In my opinion, Derakhshani discloses or suggests limitation 10[a] for the reasons discussed in §X.A.2.f (1[d1]).

c. 10[b]: processing the at least one first image or a portion to create first data;

230. In my opinion, Derakhshani discloses or suggests limitation 10[b] for the reasons discussed in §X.A.2.g (1[d2]).

d. 10[c]: moving the camera from the first location to a second location, the second location being a second distance from the user, or the user moving from the first location to the second location to change the distance between the user and the camera from the first distance to a second distance;

231. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 10[c]. *See* §X.A.2.h (1[d3]).

e. 10[d]: capturing at least one second image of the user taken with the camera of the computing device at the second distance from the user, the second distance being different than the first distance;

232. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 10[d] for the reasons discussed in §X.A.2.i (1[d4]).

f. 10[e]: processing the at least one second image or a portion thereof to create second data;

233. In my opinion, Derakhshani discloses or suggests limitation 10[e] for the reasons discussed in §X.A.2.j (1[d5]).

g. 10[f]: comparing the first data to the second data to determine whether expected distortion exist between the first data and the second data which indicated three-dimensionality of the user;

234. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 10[f]. *See* §§X.A.2.k (1[d6]; describing comparison to look for expected differences), X.A.2.l (1[d7]; describing expected differences as distorting changes from Derakhshani's focus-distance approach and Derakhshani-Tanii's distance-induced distortion approach).

h. 10[g]: authenticating the user when the differences between the first data and the second data have expected distortion resulting from movement of the camera from the first location to the second location or movement of the user from the first location to the second location, which causes the change in distance between the user and the camera.

235. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 10[g] for the reasons discussed in §X.A.2.l (1[d7]).

12. Claim 11

a. 11[a]: The method according to claim 10, further comprising: interpolating the first data and the second data to obtain estimated intermediate data;

236. In my opinion, Derakhshani, alone or in combination with Tanii, discloses or suggest limitation 11[a] for the reasons discussed in §X.A.3.a (2[a]).

b. 11[b]: capturing at least one third image of the user taken with the camera of the computing device at a third distance from the user, the third distance being between the first distance and the second distances;

237. In my opinion, Derakhshani, alone or in combination with Tanii, discloses or suggest limitation 11[b] for the reasons discussed in §X.A.3.b (2[b]).

c. 11[c]: processing the at least one third image or a portion thereof to obtain third data; and

238. In my opinion, Derakhshani, alone or in combination with Tanii, discloses or suggest limitation 11[c] for the reasons discussed in §X.A.3.c (2[c]).

d. 11[d]: comparing the estimated intermediate data with the third data to determine whether the third data matches the estimated intermediate data.

239. In my opinion, Derakhshani, alone or in combination with Tanii, discloses or suggest limitation 11[d] for the reasons discussed in §X.A.3.d (2[d]).

13. Claim 12: The method according to claim 10, further comprising verifying the presence of the user's ears in the at least one first image, and verifying the absence or reduced visibility of the user's ears in the at least one second image, wherein the first distance is larger than the second distance.

240. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 12's additional limitation for the reasons discussed in §X.A.4 (claim 3).

14. Claim 13: The method according to claim 10, wherein the computing device is configured to display one or more prompts on a screen of the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

241. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 13's additional limitation for the reasons discussed in §X.A.5 (claim 4).

15. Claim 15: The method according to claim 10, wherein the computing device is a hand-held device, and the user holds the device at the first and second distances to capture the at least one first image and the at least one second image.

242. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 15's additional limitation for the reasons discussed in §X.A.7 (claim 6).

16. Claim 16: The method according to claim 10, wherein the first data and the second data comprise biometric data.

243. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 16's additional limitation for the reasons discussed in §X.A.8 (claim 7).

17. Claim 17: The method according to claim 10, wherein the first data and the second data comprise a mapping of facial features.

244. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 17's additional limitation for the reasons discussed in §X.A.9 (claim 8).

18. Claim 19: The method according to claim 10, wherein the user's face is held steady and the camera moves from the first location to the second location.

245. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 19's additional limitation for the reasons discussed in §X.A.10 (claim 9).

19. Claim 20: The method according to claim 10, wherein the first data and the second data are maintained on the computing device.

246. In my opinion, Derakhshani discloses or suggests claim 20's additional limitation.

247. Derakhshani discloses that the biometric-authentication process can be performed locally on the device, on a server, or split between the two. *See, e.g.*, *Derakhshani*, 9:27-58, 10:1-24. In my opinion, a POSITA would have understood that Derakhshani's three-dimensional verification—which is part of the biometric-authentication process—would be configured in some circumstances to perform locally on the device. In such instances, in my opinion, a POSITA would have understood that the biometric data would be maintained on the computing device so that all processes would be performed using a single device. If a single, local device is being used for all authentication procedures, there would be no need to store any data on a separate device. Moreover, a POSITA would have understood that biometric data is extremely sensitive, and transmitting that data over networks presents a security risk of being intercepted. A POSITA would have understood that there would be no need to take any security risks by transmitting such data if the single, local device is handling the entire biometric-authentication procedure.

B. Ground 1B: Derakhshani, Tanii, and Tahk (Claim 14)

1. Motivation to Combine

248. In my opinion, a POSITA would have been motivated to modify Derakhshani, with or without Tanii, in view of Tahk because Tahk provides a user-friendly way of ensuring that a face presented for facial authentication is properly framed. In my opinion, a POSITA would have understood that providing a real-time preview of what an image would look like prior to capturing the image, as well as providing express prompts—such as written instructions or oval shapes on the live-preview screen—as taught by Tahk, §VII.D (Tahk), would have been particularly useful for authentication procedures that require capturing multiple images of a face, as taught by Derakhshani and Tanii.

249. For instance, providing a user a real-time preview would allow the user to actively adjust the position of the camera and/or their orientation of their face to properly frame their face for the image capture. Moreover, providing express prompts and oval shapes sized to guide the user to properly position their face would ensure images best suited for facial recognition can be captured, and that the faces would be captured from sufficiently different perspectives and/or distances to ensure the three-dimensional verification taught by Derakhshani, alone or in combination with Tanii, could be performed. This is particularly important for facial-recognition

systems, which generally are known to have difficulty matching faces across different views. *See, e.g.*, Ex-1018, 32.

250. Moreover, a POSITA would have known that providing real-time image feedback, written instructions, and oval shapes to frame a face during a facial authentication process were all well-known and conventional techniques to provide user feedback during image capture as of the time of the invention. *See, e.g.*, Ex-1034, 7:16-8:7, Figs. 6B-7C (providing "an example of an interface used upon registering a facial image of a person to be authenticated" in which "the image of this person is displayed on a monitor" during registration, and oval-shaped prompts to indicate the size/distance of the user from the face authentication sensor); Ex-1035, 5:31-32 ("The computing device may present prompts that instruct the user to perform one or more liveness gestures"), 6:3-4 (same).

2. Claim 14: The method according to claim 13, wherein the one or more prompts are ovals on the screen within which the face of the user is placed to capture the at least one first image and the at least one second image.

251. In my opinion, Derakhshani, combined with Tanii and/or Tahk, teaches claim 14's additional limitation.

252. As I have previously explained, Derakhshani, alone or in combination with Tanii, teaches providing prompts to user to properly frame themselves at different distances to capture images for biometric authentication. *See* §X.A.14

(claim 13). But Derakhshani and Tanii do not expressly describe using oval-shaped prompts to guide a user during the facial-authentication process.

253. In my opinion, however, a POSITA would have been motivated to provide such oval-shaped prompts (as well as express written instructions) in view of Tahk. *See, e.g.*, *Tahk*, Figs. 8A-B ("Please step further back" and "Please step further forward," and presenting an oval to frame the face at the correct distance). A POSITA would have been motivated to modify Derakhshani, alone or in combination with Tanii, to provide such oval-shaped prompts because they are a natural shape to appropriately size and frame a face at different distances. *See* §VII.D (Takh); X.B.1 (motivation).

C. Ground 1C: Derakhshani, Tanii, and Hoyos (Claim 18)

1. Motivation to Combine

254. Derakhshani discloses implementing a process to verify the three-dimensionality of a user's face by capturing a series of images of a user using a mobile computing device, such as a phone or laptop. *See, e.g.*, *Derakhshani*, 5:22-27, 6:3-5, 8:11-28, 9:10-22, 16:44-18:4. As part of the authentication process, Derakhshani also discloses determining a "reflectance metric"—a measure of the reflection patterns on the surface of the user's eye—based on first and second images to further support the liveliness/three-dimensionality determination. *Derakhshani*, 18:8-29, Fig. 7. Derakhshani discloses that the light source use to calculate the

reflectance metric can be, for instance, an "LCD screen." *Id.*, 18:13-22. Derakhshani does not expressly disclose displaying an image on an LCD screen and evaluating the reflectance of that particular image.

255. Evaluating the reflectance of a particular image off of a user's face as part of a facial-authentication liveness determination was well known, and Hoyos provides one example of this process. *Hoyos*, [0018]-[0019]; [0033]-[0035]. Specifically, rather than simply illuminate a device's screen and measure the reflectance of light, generally, Hoyos discloses displaying a specific patterned image on a device's screen, capturing an image of the user, and evaluating the captured image to determine whether the image has a reflection of the specific patterned image. *Id.*, [0018]-[0019], [0033]-[0036].

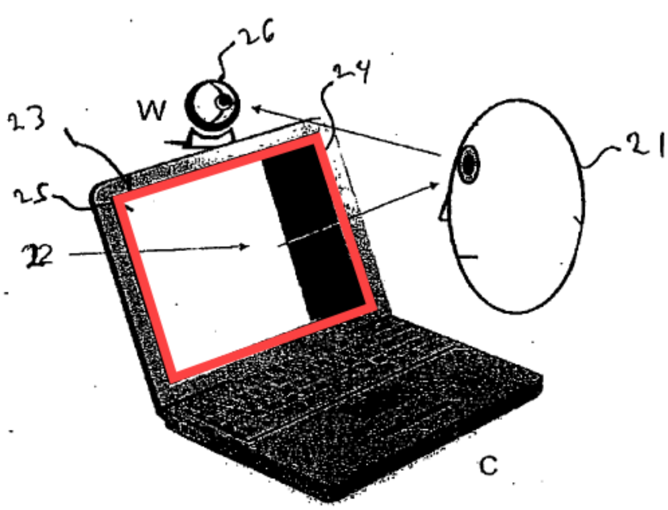


Fig. 2

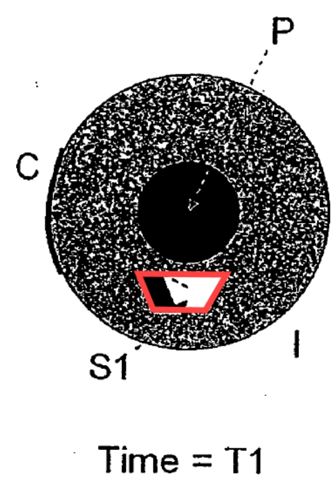


Fig. 3

Id., Figs. 2-3 (annotated).

256. In my opinion, a POSITA would have been motivated to implement Derakhshani's existing "reflectance metric" by using an image with specific pre-defined patterns and evaluating resulting images of the user for reflections of those patterns, as taught by Hoyos. That way, Derakhshani's reflectance metric would produce a "liveness" result based on *any* light reflectance appearing in a captured image of the user—which could be simulated by presenting a picture of a face with some light reflectance already incorporated. In other words, by providing a specific pattern that the "reflectance metric" looks for in captured images of the user, the pattern serves as a unique identifier to ensure the image of the user was captured *at the time of image capture*, and not just a picture of a user's face captured at a different time. This would further prevent spoof attacks that use pictures with light reflectance already in the picture.

257. Moreover, in my opinion, a POSITA would have had a reasonable expectation of success in making this modification because Derakhshani already discloses using the screen as the light source that induces reflectance in a real face. The only change required would be having the screen display a uniquely patterned image, which would involve minimal computer coding to display the image on the screen at the time of image capture.

2. Claim 18: The method according to claim 10, further comprising displaying an image on a screen of the computing device while capturing the at least one first image and/or the at least one second image, and processing the at least one first image and/or the at least one second image to detect a reflection of the displayed image off of the user's face.

258. In my opinion, Derakhshani in view of Tanii and/or Hoyos teach claim 18's additional limitation.

259. Derakhshani discloses using a reflectance metric to check for liveness in addition to a spatial metric to check for three-dimensionality of the face, and using a devices screen to provide the light source that would reflect off of the user's face. *See, e.g.*, *Derakhshani*, 5:22-27, 6:3-5, 8:11-28, 9:10-22, 16:44-18:29, Fig. 7. Although Derakhshani does not expressly disclose displaying an *image* on the device's screen when capturing images, or detecting a reflection of the displayed image off of the user's face in the captured images, Hoyos teaches this is a well-known method to verify the liveness of the user. *Hoyos*, [0018]-[0019]; [0033]-[0035]. In my opinion, a POSITA would have been motivated to modify Derakhshani to incorporate Hoyos's reflectance detection to provide an additional verification that the user is presenting a real, three-dimensional face. *See* §X.C.1 (motivation).

D. Ground 2A: Zhang and Tanii (Claims 1-3, 5-12, 15-17, 19-20)

1. Motivation to Combine

260. In my opinion, a POSITA would have been motivated to combine Zhang and Tanii because both concern identifying and accounting for the three-dimensional nature of a face when capturing an image. They differ, however, in what principles are used to account for the face's three-dimensionality. Zhang, for instance, looks to dissimilarities in two images after one undergoes a mathematical homography. *See* §VII.C (Zhang). And although Tanii is not expressly directed to *evaluating* whether a face has depth like Zhang, Tanii exemplifies the well-known distortions caused by the interaction between the camera's lens and the three-dimensional nature of the face, *see* §VII.B (Tanii). A POSITA would have appreciated, therefore, that Tanii recognizes another clear alternative to evaluating the depth of a face, consistent with Zhang's existing homography transformation.

261. A POSITA would have recognized, as Tanii does, that distance-induced distortions occur because of the interactions between the shape of the camera lens and shape of the face, and the distortion in part depends on the distance between the face and the camera. §VII.B (Tanii); *Tanii*, [0048]. Accordingly, a POSITA would have understood from Tanii that, by taking two images from two different distances, a larger amount of distortion in the closer of the two images indicates whether a face is three-dimensional or not.

262. In my opinion, a POSITA would have therefore appreciated from Tanii that images captured by Zhang—without any modification—may exhibit distance-induced distortions based on the particular camera used to perform Zhang's process (e.g., particularly when a wide-angle lens with significant barrel distortion is used, as is common in computers and mobile devices). However, a POSITA would have also appreciated that any distance-induced distortions would further enhance Zhang's homography-transformation process because a homography transformation cannot correct for these distortions.

263. For instance, if a homography transformation were applied to Tanii's Figure 4B (serving as Zhang's "first image") to compare to Figure 3B (serving as Zhang's "second image"), the transformation would not account for differences between the images caused by the distance-induced distortion.

Fig.4B

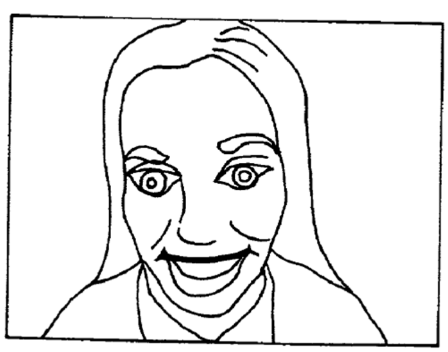
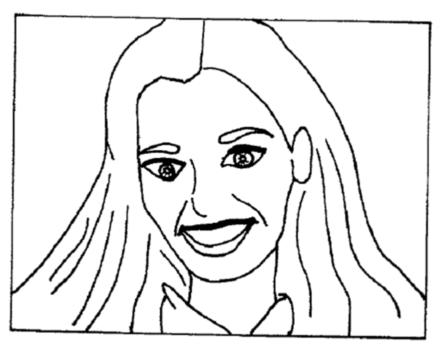


Fig.3B



Tanii, Figs. 3B, 4B. That is because Zhang relies on a mathematical principle that enables transforming the *perspective* of a planar object, such as a photograph being used to spoof the authentication procedure to a different *perspective*, §VII.C (Zhang), whereas the distortion identified by Tanii is *radial* and a byproduct of the lens' imperfections and the change in magnification with distance. A homography transform does not account for such radial distortions, but would instead transform the perspective of Tanii with its distortions intact. In other words, in a transformation of perspective with a three-dimensional object such as a real face, Tanii's distance-induced distortions would remain. Ultimately, however, when comparing the two images once one is transformed into the perspective of the other, there would remain differences attributable to the distance-induced distortion which, in my opinion, a POSITA would have understood would result in Zhang identifying the face as three-dimensional.

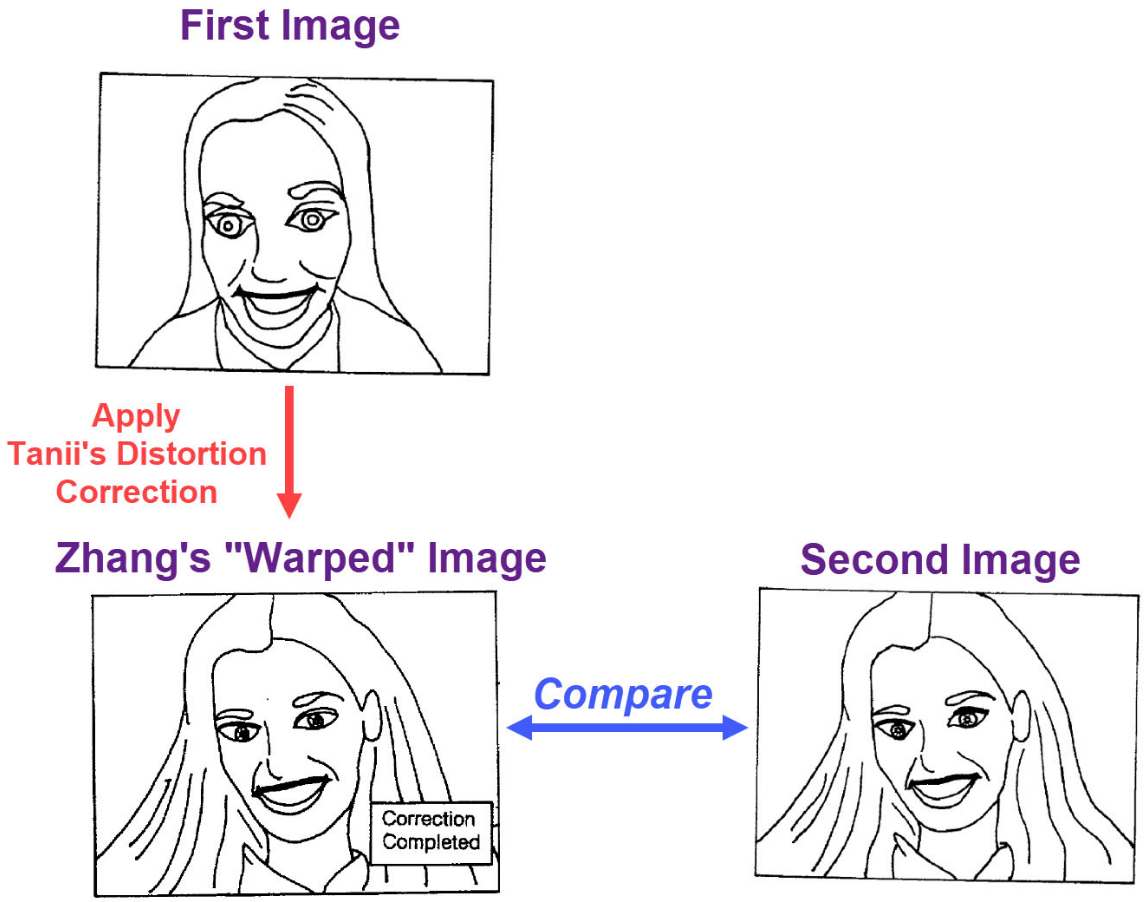
264. In my opinion, a POSITA would have therefore recognized that Zhang's existing process would be *enhanced* by prompting a user to capture two images and two distances—one of which would have increased distance-induced distortion—because if the face were three-dimensional, Zhang's existing procedure would identify the two images as different and indicate a three-dimensional face. The lack of a match between the two images would likely be enhanced by changes in radial distortion: it makes them even less like data from two planar objects which would produce a match.

265. However, in my opinion, a POSITA would have also been motivated to modify Zhang's process in view of Tanii in either of two additional ways.

266. First, in my opinion, a POSITA would have been motivated to follow Zhang to verify the three dimensionality of a face during a facial authentication procedure by taking two or more images, applying mathematics to one of the images, and comparing the mathematically altered image to a second (unaltered) image. But instead of the mathematics applied being a homography transformation, in my opinion, a POSITA would have been motivated to *substitute* Zhang's mathematics for those taught by Tanii to correct for distance-induced distortion. In other words, rather than change the perspective of one image to match the second image, a POSITA would correct the distortion of one image (to create what Zhang refers to as its "warped" image[4]) and compare the result to another image taken further away

[4] Zhang and Tanii both use the term "warped" to refer to different effects, but they are not inconsistent with one another. Specifically, Zhang uses the term "warped" to refer to the resulting image that has undergone homography transformation because the original relationship between the pixels in the image are modified. Tanii uses the term "warped" to refer to the distortions in an image of a face induced by the image-capture conditions (e.g., distance and lens geometry). When I refer to Zhang's

that does not exhibit the same degree of distance-induced distortions.



Tanii, Figs. 3B, 4B, 9.

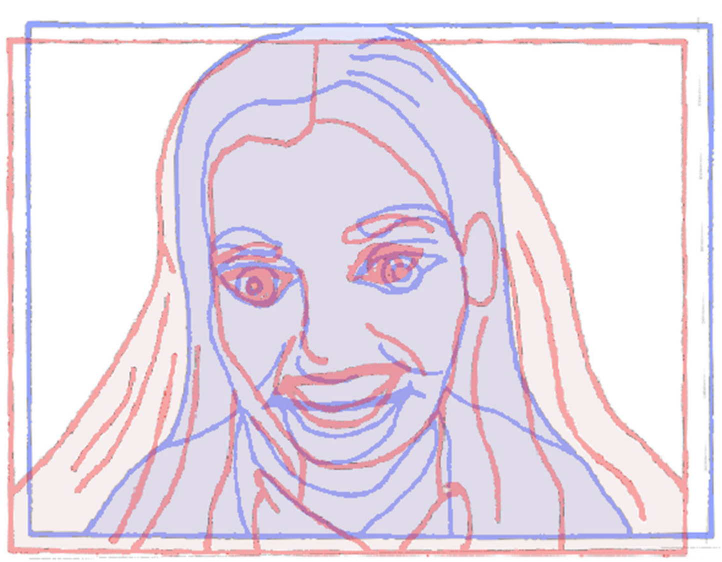
267. A POSITA would have appreciated that if the "warped" (distortion-corrected) image and second image are sufficiently similar, that indicates a three-dimensional face because Tanii is correcting for distortions attributable the three-

---

"warping," I am referring to the result of a mathematical application to an image; and when I refer to Tanii's warping, I am referring to distance-induced distortion.

dimensionality of the user's face. By following this approach, a POSITA would have recognized that the only difference (besides the mathematics) is that the comparison between the Zhang-Tanii "warped" (distortion-corrected) image would look for a match with the second image.

268. Alternatively, a POSITA would have appreciated that Zhang and Tanii could be modified to eliminate the mathematical transformation of a first image entirely. Once again, a POSITA would have been motivated to follow Zhang to verify the three dimensionality of a face during a facial authentication procedure by taking two or more images, but rather than apply mathematics to "warp" one of the images (e.g., using either a homography transform or distortion-correction procedure), the facial features would be mapped in each image, matched between the two images, and evaluated to determine whether differences attributable to distance-induced distortion appear (e.g., does the shape of the nose, size of the mouth or forehead, or do facial features shift by expected degrees relative to one another?). For instance, I have overlayed Tanii's two images to show how one (in blue) exhibits expected distortions while the other (in red) does not, resulting in various misalignments in facial features (assuming the faces are normalized in size):



In such circumstances, a POSITA would understand that two images would still be required, rather than just evaluating one image for distance-induced distortion. Otherwise, an imposter could provide a picture of a user with distance-induced distortion already applied to spoof the system; the need for a more-distance, undistorted image of the user for comparison would still be required.

269. In my opinion, a POSITA would have been motivated to make either of these two modifications for two reasons. First, a POSITA would have appreciated that Zhang's homography-transformation process may be spoofed by presenting a non-planar picture of a face, because Zhang's homography transformation process specifically looks for a *planar* structure in images that can be transformed nearly identically from one perspective to another. Thus, Zhang's system could possibly be subverted by bending the picture in a way to trick the system, or applying the picture to a three-dimensional shape. For this reason, in my opinion, a POSITA would have

therefore been motivated to look for other methods to ensure the user's face is from the user, and not a spoofer. A POSITA would have also appreciated that distance-induced distortion is more difficult to spoof, because it is induced by the interactions of geometries between the user's face and the camera's lens, and therefore could not be circumvented as easily. Second, a POSITA would have appreciated that either of the processes suggested by Tanii offers a potentially less computationally demanding than the homography mathematics proposed by Zhang, which may be more suitable for a low-power portable device.

270. In my opinion, a POSITA would have had a reasonable expectation of success in making this modification to Zhang because Tanii already taught a mechanism to identify (and correct) distance-induced distortions, *see, e.g.*, *Tanii*, [0056], and it was already well-known to use depth information about a face derived from a series of images to distinguish between live faces and two-dimensional images of faces. *See, e.g.*, Ex-1014, Abstract, [0031], [0036].

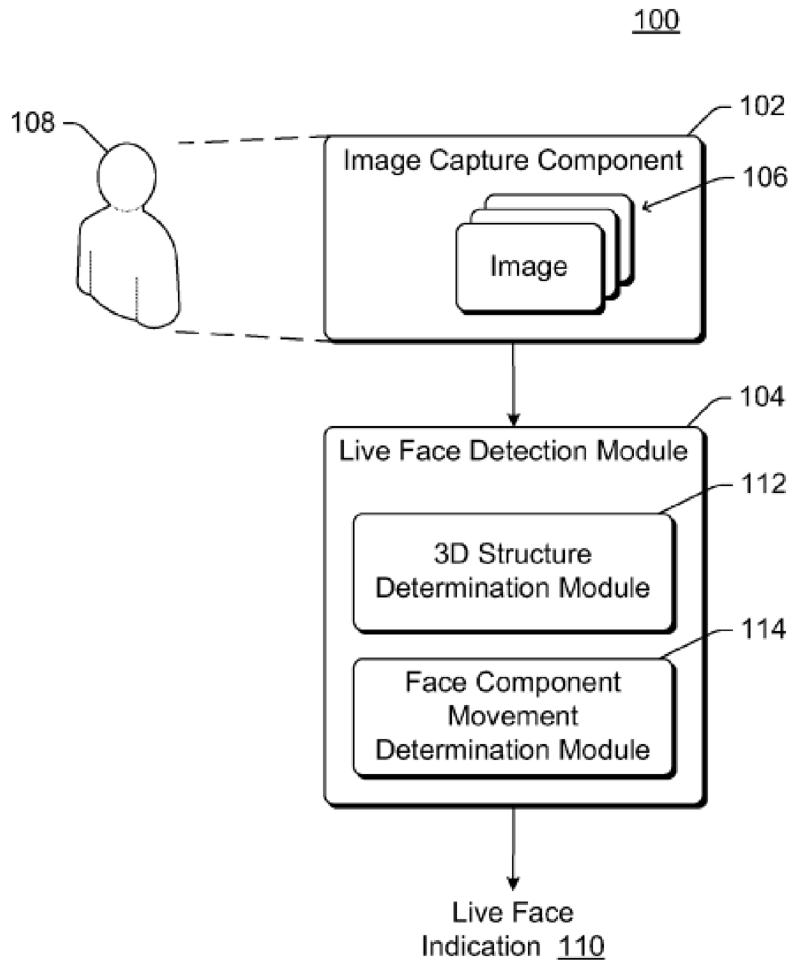
2. Independent Claim 1

a. 1[pre]: A system for authenticating three-dimensionality of a user via a user's camera equipped computing device, the computing device comprising:

271. If the preamble is limiting, in my opinion, Zhang discloses or suggests it.

272. Zhang discloses a system "to determine whether a face in multiple images is a 3D structure or a flat surface" (*Zhang*, [0026], Figs. 1-3; *see also, e.g., id.*, Abstract, [0003], [0013]) to "authenticate a user for particular access" (*id.*, [0012]). To accomplish this, Zhang captures and analyzes multiple images of a user's face using the image capture component 102 implemented in a computing device (e.g., "a desktop computer, a laptop or notebook computer…[or] a cellular or other wireless phone"). *Zhang*, [0012]-[0013], [0016].

273. In my opinion, a POSITA would have understood that the "image capture component 102" would be a camera, because cameras are conventionally used to capture images, especially in computing devices. In fact, the "CCDs" and "CMOS" sensors Zhang references are the types of sensors commonly used in cameras. *Zhang*, [0016]; *see also, e.g.*, *Suzuki*, [0019] ("The camera unit includes solid-state image pickup elements such as CCD or CMOS"); Ex-1028, 3 ("Presently, there are two main technologies that can be used for the image sensor ***in a camera***, i.e., CCD (Charge-coupled Device) and CMOS (Complementary Metal-oxide Semiconductor).

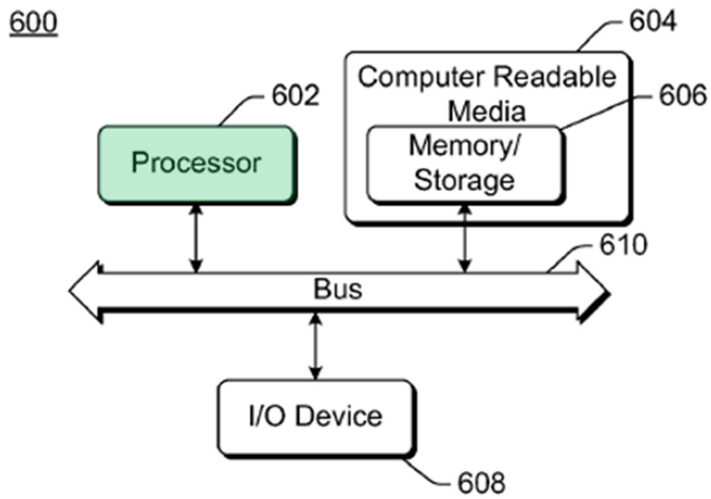


Zhang, Fig. 1.

b. 1[a]: a processor configured to execute machine executable code;

274. In my opinion, Zhang discloses or suggests limitation 1[a].

275. Zhang discloses a computing device that contains a processor and computer-readable media (e.g., memory) storing software instructions.



Zhang, Fig. 6 (annotated), [0063]-[0067].

c. 1[b]: a screen configured to provide a user interface to the user;

276. In my opinion, Zhang discloses or suggests limitation 1[b].

277. Zhang discloses a computing device that contains a display that allows a user to interact with the device and presents information to the user. *Zhang*, [0067].

d. 1[c]: a camera configured to capture images;

278. In my opinion, Zhang discloses or suggests limitation 1[c].

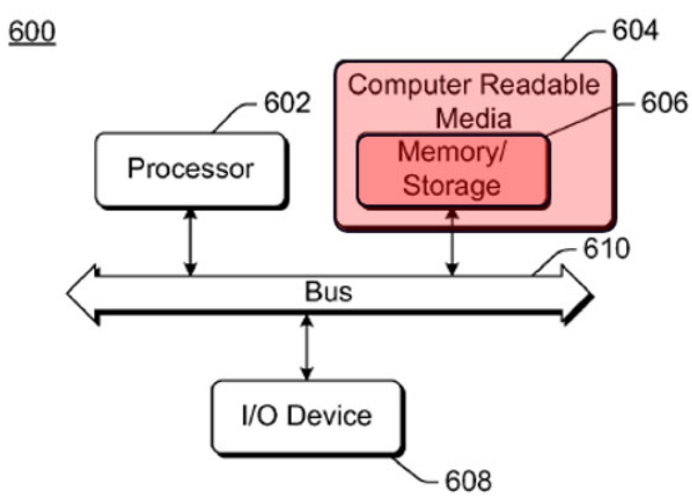
279. Zhang discloses a computing device that contains an image capture component. *Zhang*, [0012]-[0013]. As I explained previously, a POSITA would have understood that Zhang's "image capture component" is a camera. §X.D.2.a (1[pre]).

e. 1[d]: one or more memories configured to store machine readable instructions that are stored on the memory of the authentication server which when executed by the processor, cause the computing device to:

280. In my opinion, Zhang discloses or suggests limitation 1[d].

281. As I mentioned previously, in my opinion, it is somewhat unclear which structure is intended to be the "computing device": (1) either a user-facing computing device that engages with a back-end authentication server; or (2) the authentication server itself. *See* §VI.A.3. However, in my opinion, at a minimum, Zhang discloses the former, and the latter would have generally been obvious to a POSITA.

282. Zhang discloses a computing device that contains computer-readable media (e.g., memory) storing software instructions. *Zhang*, Fig. 6 (annotated), [0063]-[0067].



Zhang also discloses that the image capture component and live face detection module (104) can be separate computing devices that communicate and send data, including biometric facial feature data, over a variety of different networks, such as the Internet, a local area network (LAN), an intranet, etc. *Zhang*, [0014].

283. Although Zhang does not expressly mention that the data is sent to a "server," in my opinion a POSITA would have found it obvious that Zhang's separate computing device would be a server because servers were well-known networking infrastructure, and servers were known to be used for back-end processing of biometric data. *See, e.g.*, *Derakhshani*, 9:27-58, 10:1-24; Ex-1016, Abstract, [0040]-[0043]; Ex-1012, Fig. 1A, 5:24-50. Furthermore, in my opinion, a POSITA would have further understood that, when using a server as the separate computing device running live face detection module (104), the server would store the machine-readable instructions to carry out Zhang's disclosed process and would send instructions to the separate image-capture component that would be stored (even if temporarily) in memory provided in the image-capture component to process and execute those instructions to carry out facial recognition.

284. Alternatively, a POSITA would have considered it obvious that a server itself may utilize biometric protection. A POSITA would have understood that, just like personal computers, servers were known to store sensitive information—from user profiles for websites, employment or medical records, and more. A POSITA

seeking to prevent unauthorized access to reconfigure servers or access their files would have therefore understood that the server itself may be provided with biometric authentication and that in such cases it would include a camera (X.D.2.d (1[c])) to carry that authentication out.

f. 1[d1]: capturing at least one first image of the user taken with the camera of the computing device at a first location which is a first distance from the user;

285. In my opinion, Zhang discloses or suggests limitation 1[d1].

286. Zhang discloses capturing a first image of a user as part of the authentication method. *Zhang*, [0016] ("user 108 presents himself or herself to image capture component 102, allowing component 102 to capture images 106 of user 108."), [0021].

287. In my opinion, a POSITA would have understood that Zhang's process captures an image at a first distance between the user and image capture component 102 in order to capture a picture of the user's face. *Zhang*, [0016]. Specifically, a POSITA would have understood that there *must* be some distance between the camera and the face to capture the first image—enough so that the camera's field of view encompasses the face. If there were no distance between the camera and the face (e.g., if the camera were pressed up against the user's skin), then the field of view would be limited to just that patch of skin and the face would not be captured in the image, which would be useless for the three-dimensional verification process.

g. 1[d2]: processing the at least one first image or a portion to create first data;

288. In my opinion, Zhang discloses or suggests limitation 1[d2].

289. Zhang discloses processing the first image to extract "feature points" from the image. *Zhang*, [0027] ("[O]ne or more feature points are extracted from two images… A variety of different feature points can be extracted, such as a corner of an eye, a corner of a mouth, a tip of a nose, and so forth."), [0026] (disclosing "software, firmware, hardwire, or combin[ed]" implementations).

290. In my opinion, a POSITA would have understood that Zhang's extracted feature points constitute data—and specifically biometric data—because the identification involves using a computer (which operates on data) to characterize the unique physical characteristics of an individual, which would include the positions of "feature points" such as a user's eyes, nose, mouth, and other such features. *See, e.g.*, Ex-1018, 2 ("biometric recognition can be defined as the science of establishing the identity of an individual based on the physical and/or behavioral characteristics of the person."), 100-103 (describing the types of biometric data about a face used for facial-recognition systems, "such as the structure of the face components (e.g., eyes), [and] the relationship between facial components").

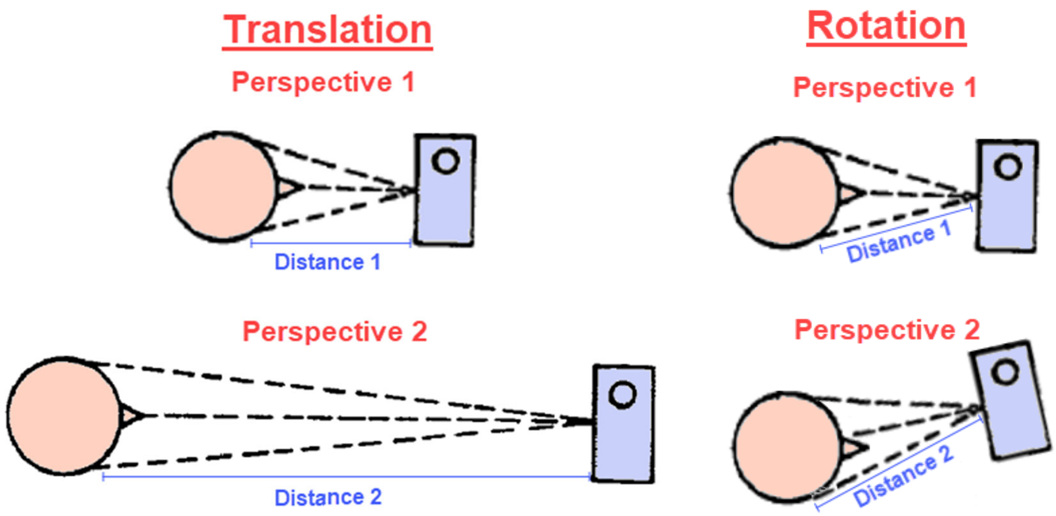
h. 1[d3]: moving the camera from the first location to a second location, the second location being a second distance from the user, or the user moving from the first location to the second location to change the distance between the user and the camera from the first distance to a second distance;

291. Zhang, alone or in combination with Tanii, teaches limitation 1[d3].

292. Zhang discloses capturing a second image of a user as part of the authentication method. *Zhang*, [0016] ("Image capture component 102 captures multiple images").

293. Zhang does not expressly disclose that the second image is captured at a second distance different from the first distance of the first image. But, in my opinion, a POSITA would have understood that Zhang at least implicitly requires *some* change of distance. §X.D.1 (motivation). Specifically, Zhang discloses a "3D structure determination module 112" that uses a "homography" technique to distinguish between a real face and a picture of a face by, *inter alia*, transforming a first image to the perspective of a second image and comparing the two. *Zhang*, [0024], [0026]-[0035]; §VII.C (Zhang). In my opinion, a POSITA would have understood from Zhang that—like Derakhshani's parallax approach—the distances between the camera and at least some facial landmarks would change in order to obtain an image from a different perspective than the first, and would obviously also encompass changing the overall distance between the camera and face as well. *See,*

e.g., §X.A.2.i (in the context of Derakhshani, discussing changes of distance for parallax).



Moreover, a POSITA would have not only understood that providing images at different distances allows for a greater understanding of depth between objects in the scene, as exemplified in the paper Zhang references; Ex-1013, 22-25, but that taking pictures at different distances may induce distance-based distortion that would *enhance* the accuracy of Zhang's homography transformation to detect a three-dimensional face. §X.D.1 (motivation).

294. Even if Zhang cannot be considered to disclose or suggest taking two images at different distances, however, a POSITA would have been motivated to do so in view of other prior art. For instance, a POSITA would have understood that distortions caused by camera lenses can indicate depth in the object being captured, as exemplified by Tanii. §X.D.1 (motivation). Thus, even if Zhang does not already disclose this limitation, a POSITA would have been motivated to modify Zhang in

view of Tanii to capture a second image at a second distance and evaluating the images for different degrees of distance-induced distortions to distinguish between live, three-dimensional faces and two-dimensional pictures of a face. §X.D.1 (motivation).

295. Accordingly, because Zhang, whether alone or combined with Tanii, teaches capturing a plurality of images at different distances between the user and the camera, in my opinion, a POSITA would have understood that either the camera or user must move relative to the other in between image captures; there are no other ways to change the distance between the two. In such instances, a POSITA would have understood that moving a camera in relation to the user's face is a convenient and obvious option for changing the distance between the user's face and the camera.

i. 1[d4]: capturing at least one second image of the user taken with the camera of the computing device at the second distance from the user, the second distance being different than the first distance;

296. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 1[d4]. *See* §X.D.2.h (1[d3]).

j. 1[d5]: processing the at least one second image or a portion thereof to create second data;

297. In my opinion, Zhang discloses or suggests limitation 1[d5].

298. Zhang discloses processing the second image to obtain second feature-point biometric data from the image. *Zhang*, [0026]-[0027]; §X.D.2.g (1[d2]).

k. 1[d6]: comparing the first data to the second data to determine whether expected differences exist between the first data and the second data which indicated three-dimensionality of the user;

299. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 1[d6].

300. Zhang discloses that "[t]he feature points extracted…are matched across the first and second images (act 304)" and, in my opinion, those feature points constitute data, and specifically biometric data. *Zhang*, [0028]; §X.D.2.g (1[d2]). Zhang also discloses that the matching process may also "determine[] whether the first and second images include the same face," including "during the matching of feature points in act 304, if all (or at least a threshold number) of the feature points cannot be matched then it is determined that the first and second images are of different faces." *Zhang*, [0038].

301. Zhang discloses that, after calculating a homography matrix between the first and second image, a "warped" version of the first image is created and then compared to the second image to determine whether expected differences exist. *Zhang*, [0025], [0031]. Zhang also discloses that, as part of the comparison, "any of a variety of conventional face detection algorithms or face recognition algorithms can be used to detect the face within each image, and the selected locations are the locations that are part of a face within at least one of the warped and second images." *Zhang*, [0032].

302. In my opinion, a POSITA would have understood that Zhang discloses comparing a first (biometric) data (e.g., the facial-feature locations in the first "warped" (transformed) image) and second (biometric) data (e.g., the facial-feature locations in the second image) to determine whether differences between the two exist, in which it would be expected that a live face would have sufficient differences between the two images due to movement of the image capture component 102 (camera).

303. However, a POSITA would have also been aware that differences between two images—one with lens-induced distortions and one without—can also be used to distinguish between live, three-dimensional faces, and two-dimensional pictures of a face, as exemplified by Tanii. §X.D.1 (motivation). And, in my opinion, a POSITA would have been motivated to look for these expected distortions as either a supplemental or alternative verification of three-dimensionality of a face. *Id*. A POSITA would have appreciated that verifying the three-dimensional nature of the face using distance-induced distortion would be accomplished by matching the positions of biometric facial features across the first and second images, consistent with Zhang. But rather than using that comparison to calculate a homography matrix, the comparison would evaluate whether one of the images exhibits the distance-induced distortion that would be expected when the user's face is captured at a close distance to the camera, and the other image does not exhibit similar lens-induced

distortion when captured further from the camera. *Id*. In my opinion, a POSITA would have appreciated that, when modifying Zhang to evaluate differences caused by distance-induced distortions, a three-dimensional face would be indicated when one of the two sets of data exhibits expected distance-induced distortions due to the change in distance of the camera. *Id.*, [0025], [0034].

l. 1[d7]: authenticating the user when differences between the first data and the second data have expected distortion resulting from movement of the camera from the first location to the second location or movement of the user from the first location to the second location, which causes the change in distance between the user and the camera.

304. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 1[d7].

305. Zhang discloses that captured images are determined to be of a live, three-dimensional face when differences in the image data exist after undergoing a homography transformation, *Zhang*, [0031], including when first biometric data (the position of facial features in the first "warped" (transformed) image) does not match the second biometric data, *Zhang*, [0032]-[0034]. These differences would be expected distortions due to a change in perspective (rotation and/or distance) of the camera between the two images. *See* §§X.D.2.h-X.D.2.i (1[d3]-[d4]), X.D.2.k (1[d6]). By matching data across the first and second images, Zhang also discloses determining whether the first and second images include the same face for the

purpose of authenticating the user. *Zhang*, Fig. 2, [0017], [0038]. Conversely, if "all (or at least a threshold number) of the feature points cannot be matched then it is determined that he first and second images are of different faces," and the user is not authenticated. *Zhang*, Fig. 2, [0017], [0038].

306. Moreover, to the extent Zhang does not disclose this limitation, Zhang combined with Tanii does. *See* §§X.D.2.h-X.D.2.i (1[d3]-[d4]), X.D.2.k (1[d6]). Specifically, a POSITA would have appreciated that, when modifying Zhang to evaluate differences arising from distance-induced distortions in Tanii, a three-dimensional face would be indicated when one of the two sets of biometric data exhibits expected distance-induced distortions due to the change in distance of the camera, allowing Zhang's authentication process to proceed. *See* §§X.D.2.h-X.D.2.i (1[d3]-[d4]), X.D.2.k (1[d6]).

307. Finally, it is worth noting that although the '471 Patent uses the term "authentication" in other claims (such as claims 5 and 10, which recites "enrollment data" and an "authentication session," respectively) which a POSITA would have understood refers to comparisons to enrollment data to authenticate a user's identity, 1[d7] appears to use the term "authentication" to refer to authenticating the three-dimensionality of the face, because a user would not be broadly authenticated just because they demonstrated their face is three dimensional without also

authenticating their identity. Accordingly, in my opinion, the ’471 Patent does not use the term “authentication” consistently to refer to comparisons to enrollment data.

308. Regardless, Zhang discloses that, as part of the authentication process, a user’s identity can be authenticated “by comparing one or more of the images 106 to previously captured images of the user 108.” *Zhang*, [0017].

3. Claim 2

a. 2[a]: The system according to claim 1, further comprising: interpolating the first data and the second data to obtain estimated intermediate data;

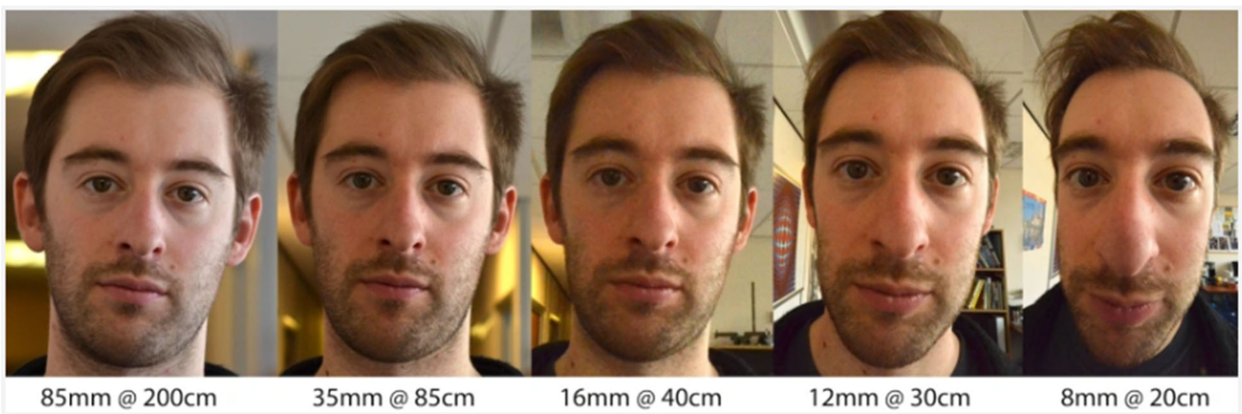
309. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 2[a].

310. Zhang discloses that, as part of the authentication process, the two images being compared may be “non-adjacent.” *Zhang*, [0035]. Zhang explains that images are “non-adjacent” when additional images exist between the two images being compared for authentication. *Id.* In such instances, Zhang discloses performing some of the processes, such as “feature point extraction and feature point matching” using the intermediate images to “facilitate the feature matching process when matching features across two images with one or more intervening images.” *Id.*, [0036] Zhang also discloses that the homography-transformation process can be applied to multiple pairs of images, whether the images are adjacent or non-adjacent. *Id.*, [0037].

311. When a set of intermediate images exist *between* the first and second images, as Zhang discloses, in my opinion, a POSITA would have been motivated to generate predictions (i.e., interpolations) of what those intermediate images should look like based on Zhang's first and second images because using static images to build models or predictions of the face as a means of identifying a user was well-known in the art. Ex-1015, Abstract; *Derakhshani*, 17:24-44 (interpolating two-dimensional and three-dimensional models for comparison to acquired biometric data); Ex-1036, 8:19-27 (describing capturing one or more biometric features and calculating "change parameters" to evaluate whether the changes match expectations, or predictions of what the biometric features should look like). And a POSITA would have understood that building models or predictions of what Zhang's intermediate images *should* look like would further ensure against spoofing because a spoofer could not rely on artificial differences between the first and second images to have Zhang's system authenticate a face; the differences would also have to match what is expected *between* the two images.

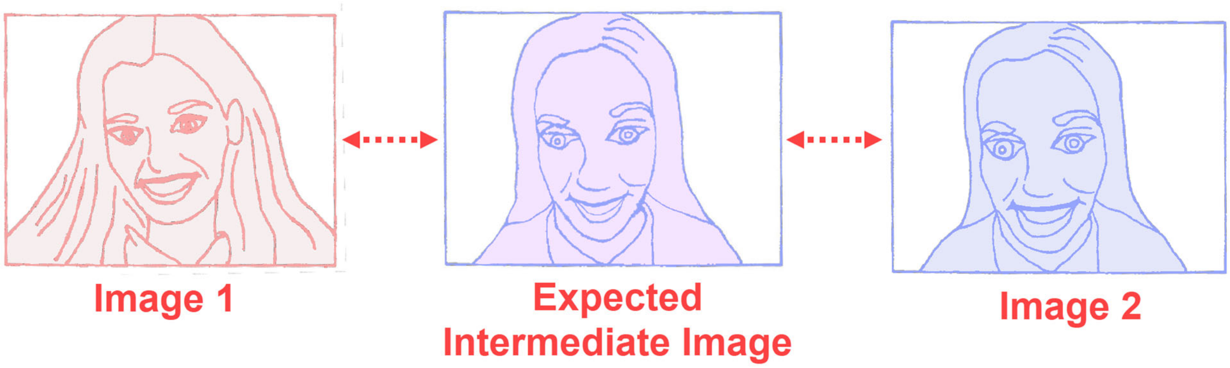
312. Based on a POSITA's understanding of Zhang, a POSITA would have further been motivated to derive interpolated biometric data based on the combination of Zhang and Tanii. Specifically, Tanii discloses that distance-induced distortions increase as distance between the face and camera decreases. *See Tanii*, [0048]. A POSITA would have therefore understood that, all else being equal,

distance-induced distortion depends on the distance between the user and the camera, and thus any set of images as the distance between the user and camera changes will have different degrees of facial warping, similar to the gradual changes in facial distortion that appears in the series of images below (although these images also altered the focal length of the camera to ensure the face remains a constant size in the frame, rather than just distance:



Ex-1022.

313. In my opinion, a POSITA reading Zhang—which discloses processing, interpolating, and evaluating intermediate images—in view of Tanii therefore would have been motivated to interpolate intermediate biometric data with an intermediate, interpolated amount of distance-induced distortion based on the two non-adjacent images to create an array of intermediate distance-induced distortions that would indicate depth in a three-dimensional face, such a gradual projection from the lines of the facial landmarks shown in blue (the distorted image) to the lines of the facial landmarks shown in red (the distorted image).



Performing this type of interpolation between the distorted and undistorted images would allow for further comparison with additional images, for instance, to ensure the distance-induced distortion matches expectations if the face were truly three-dimensional, as depicted below:

b. 2[b]: capturing at least one third image of the user taken with the camera of the computing device at a third distance from the user, the third distance being between the first distance and the second distances;

314. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 2[b].

315. Zhang discloses capturing a series of intermediate images between two non-adjacent images. *Zhang*, [0035]-[0037]; *see* §X.D.3.a (2[a]). A POSITA would have understood that these intermediate images would provide images at different positions (e.g., rotation or translation) of the camera relative to the first and second images. *See* §X.D.3.a (2[a]).

316. When modifying Zhang in view of Tanii to interpolate intermediate biometric data attributable to distance-induced distortions, a POSITA would have been further motivated to capture a third image at a distance that correlates to one of the interpolated data sets for further authentication of three-dimensional depth of the face in the captured images. *See* §X.D.3.a (2[a]).

c. 2[c]: processing the at least one third image or a portion thereof to obtain third data; and

317. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 2[c].

318. Zhang discloses processing a third (intermediate) image to obtain third biometric data. *Zhang*, [0036] ("the feature point extraction and feature point matching in acts 302 and 304 can be generated for each adjacent pair of images in the sequence, which can facilitate the feature matching process when matching features across two images with one or more intervening images.").

319. Moreover, as discussed previously, when modifying Zhang in view of Tanii, a POSITA would have found it obvious to acquire a third image and extract

biometric data from the third image to compare it to the interpolated positions of the biometric data based on the first and second images. *See* §§X.D.3.a (2[a]), X.D.3.b (2[b]).

d. 2[d]: comparing the estimated intermediate data with the third data to determine whether the third data matches the estimated intermediate data.

320. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 2[d].

321. Zhang discloses tracking and comparing biometric features between the non-adjacent and intermediate images. *See Zhang*, [0036]-[0037]; §X.D.3.a (2[a]).

322. In my opinion, a POSITA would have understood that, when interpolating what the intermediate images *should* look like based on the first and second images, a POSITA would have understood that the estimated, interpolated biometric data would be compared to the intermediate images to determine whether the intermediate images match what was predicted. *See* §X.D.3.b (2[b]).

323. Furthermore, a POSITA would have been motivated to acquire a third image and extract biometric data from the third image to compare it to the interpolated positions of the biometric data based on the first and second images to determine if there is a match between the two. *See* §§X.D.3.a (2[a]), X.D.3.b (2[b]).

4. **Claim 3: The system according to claim 1, further comprising verifying the presence of the user's ears in the at least one first image, and verifying the absence or reduced visibility of the user's ears in the at least one second image, wherein the first distance is larger than the second distance.**

324. In my opinion, Zhang combined with Tanii teaches claim 3's additional limitation.

325. Zhang does not expressly disclose a process of verifying the presence in one image and absence in another of a user's ears.

326. However, when modifying Zhang in view of Tanii to use distance-induced distortions to verify the three-dimensional nature of a face, Tanii teaches that the absence and presence of an ear is a natural result of the distance between the user and camera. Specifically, when a sufficient distance between the face and camera exists, the ears are captured because there is enough distance for the light rays from the ears to strike the camera's lens. *See Tanii*, [0048], Figs. 3A-3B.

Fig.3A

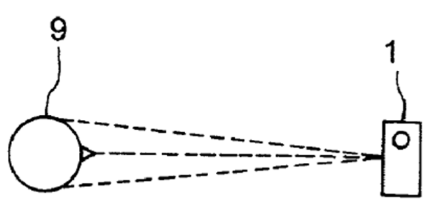
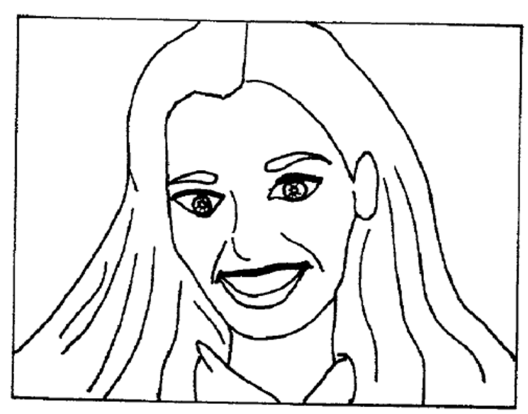


Fig.3B



But when a face is too close, the ears will not be captured because there is insufficient distance for the light rays from the ears to strike the camera's lens. *See Tanii*, [0047]-[0048], Figs. 4A-4B.

Fig.4A

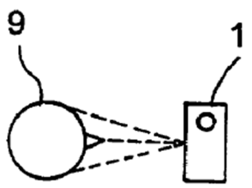
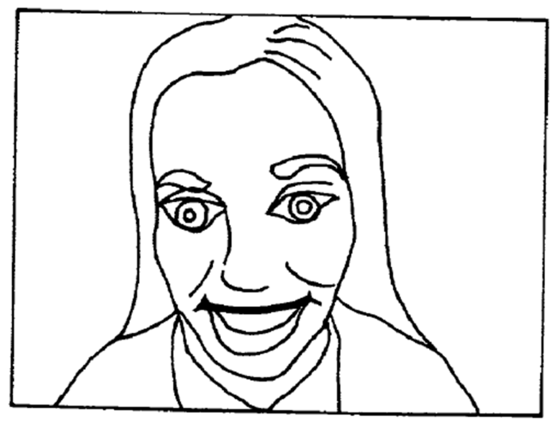
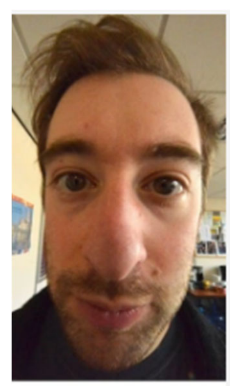


Fig.4B



This effect was well-known and demonstrated in actual applications, as shown below.

85mm @ 200cm



8mm @ 20cm

In my opinion, therefore, a POSITA would have appreciated based on at least Tanii that the presence of a user's ears in one image at a sufficient distance, but absence of a user's ears in another image at a closer distance would be indicative of a three-dimensional face, and would have been motivated to modify Zhang to verify the presence and absence of the ears between images as yet another indicator of the three-dimensional nature of the face.

5. Claim 5

a. 5[a]: The system according to claim 1, further comprising comparing the first data, second data, or both to enrollment data derived from an enrollment image, the enrollment image captured and stored prior to an authenticating; and

327. In my opinion, Zhang discloses or suggests 5[a]'s additional limitation.

328. Zhang discloses comparing at least portions of the first image, second image, or both to enrollment data captured and stored prior to the authentication session. *Zhang*, [0017] ("The authentication of a user 108 can be performed…by comparing one or more of images 106 to previously captured images of user 108.").

329. In my opinion, a POSITA would have understood that Zhang's "previously captured images" would be taken during an enrollment session, as is conventional for biometric-authentication systems. *See* Ex-1018, 4-11 (providing overview of biometric authentication and verification).

b. 5[b]: only authenticating the user when the first data, the second data, or both match the enrollment data within a predetermined threshold.

330. In my opinion, Zhang discloses or suggests 5[b]'s additional limitation.

331. Although Zhang does not provide significant details about the overall authentication process—but instead states "a variety of different manners" can be used—Zhang's description of comparing biometric features to "previously captured

images" is consistent with a conventional biometric-authentication procedure that requires a sufficient "match" above a threshold. *See Zhang*, [0017] ("The authentication of a user 108 can be performed…by comparing one or more of images 106 to previously captured images of user 108."), [0038] (disclosing inter-picture matching); *see also, e.g.*, Ex-1018, 17 ("[B]iometric systems mostly decide on a person's identity based on a *close* match between the template and the query, where the strength of the match (or the degree of similarity) is represented by the match score."), 18 ("a verification system makes a decision by comparing the match score *s* to a threshold η"). In fact, thresholds were used because facial authentication must account for many different conditions that cause differences between image captures of a face, such as the perspective of the face in the image, lighting, facial accessories, facial hair, and more. *See, e.g.*, *id.*, 99, Fig. 3.1 (noting "[t]he problem of intra-class (i.e., intra-user) variation is quite pronounced in the context of face recognition. The face image of an individual can exhibit a wide variety of changes that make automated face recognition a challenging task" such as differences in "pose, illumination, and expression…aging," and facial accessories). In other words, facial authentication looks for matches to a prescribed certainty (e.g., the threshold), rather than an exact match.

332. For these reasons, in my opinion, a POSITA would have understood Zhang as disclosing a conventional facial-authentication procedure in which the first

or second data (or both) must match "previously captured" enrollment data within a predetermined threshold to authenticate the identity of the user. In fact, authenticating a user's *identity* is a central aspect of facial authentication systems, and not just evaluating whether the face is three-dimensional or not. *Zhang*, [0001] (noting the purpose of the invention is to prevent unauthorized users from accessing secure resources); *see* Ex-1018, 259 (noting "[l]iveness detection"—like Zhang—is just one aspect of biometric authentication systems to mitigate spoofers).

6. Claim 6: The system according to claim 1, wherein the computing device is a hand-held device, and the user holds the device at the first and second distance to capture the at least one first image and the at least one second image.

333. In my opinion, Zhang combined with Tanii teaches claim 6's additional limitation.

334. Zhang discloses that the face authentication process can be implemented on a variety of different types of hand-held computing devices, such as a cellular or other wireless phone, a digital camera or video camera. *Zhang*, [0013]. Moreover, Tanii notes that distance-induced distortions often occur in mobile devices that have incorporated wide-angle lenses, and the amount of distortion is dictated by the distance between the user and the camera. *Tanii*, [0007], [0047]-[0048], Figs. 3A-B, 4A-B.

335. In my opinion, a POSITA would have also understood that, when performing a three-dimensional verification of the face on a mobile computing

device, *see* §X.D.2.i (1[d4]), the user would adjust the distance by holding the mobile device and extending and retracting their arm (holding the mobile device at a first distance, then a second distance). A POSITA would have appreciated that mobile devices are routinely held to capture images, and holding the mobile device and adjusting distance would be a convenient and obvious way of changing the distance.

7. Claim 7: The system according to claim 1, wherein the first data and the second data comprise biometric data.

336. In my opinion, Zhang discloses or suggests claim 7's additional limitation for the reasons discussed in §§X.D.2.g (1[d2]), X.D.2.j (1[d5]).

337. In my opinion, a POSITA would have also understood that data created from an image (or portion of an image) of a user's face would comprise biometric data. *Id.*

8. Claim 8: The system according to claim 1, wherein the first data and the second data comprise a mapping of facial features.

338. In my opinion, Zhang discloses or suggests claim 8.

339. Zhang discloses processing the multiple images to extract "feature points" from the image that correspond to characteristics of a user's face. *Zhang*, [0027], ("[O]ne or more feature points are extracted from two images… A variety of different feature points can be extracted, such as a corner of an eye, a corner of a mouth, a tip of a nose, and so forth."). In my opinion, a POSITA would have

understood that the identification of facial landmarks would include their locations relative to one another, thus constituting a mapping of facial features. In fact, processing image data to map facial features was a conventional aspect of facial-recognition systems. *See, e.g.*, Ex-1018, 103 (Fig. 3.5(b) describing how "Level 2 features require detailed processing for face recognition. Information regarding the structure and the specific shape and texture of local regions in a face is used to make an accurate determination of the subject's identity."). Moreover, a POSITA would have understood that the process of converting facial features in an image to computer-readable data conventionally involves mapping those features to data. *See, e.g.*, *id.*, 116-17 (noting how "appearance-based techniques generate a compact representation of the entire face region in the acquired image by *mapping* the high-dimensional face image into lower dimensional sub-space.").

9. Claim 9: The method according to claim 1, wherein the first image and the second image is of the user's face and the user's face is held steady and without movement during capture of the first image and the second image.

340. In my opinion, Zhang, alone or combined with Tanii, teaches claim 9's additional limitation.

341. Zhang and Tanii both teach or suggest moving the camera to capture images at two different distances. *See* §X.D.2.h (1[d3]).

342. In my opinion, a POSITA would have understood that, when moving the camera to capture images from different distances, the user's face would be

stationary (e.g., steady) both during each image capture (to ensure each image is not blurry due to the camera's exposure time) and during movement of the camera (to isolate any differences between images to those attributable to the change of distance). Moreover, a POSITA would have appreciated that holding the user's face steady and moving the camera closer and further away would be more user friendly than requiring the user to move their head closer and further from the camera while holding the camera steady.

10. Independent Claim 10

a. 10[pre]: A method for authenticating three-dimensionality of a user via a user's camera equipped computing device, the method, during an authentication session comprising:

343. If the preamble is limiting, in my opinion, Zhang discloses or suggests it.

344. Specifically, Zhang discloses a method "to determine whether a face in multiple images is a 3D structure or a flat surface," *Zhang*, [0026], Figs 2-3; *see also, e.g.*, *id.*, Abstract, [0003], to "authenticate a user for particular access," *id.* [0012].

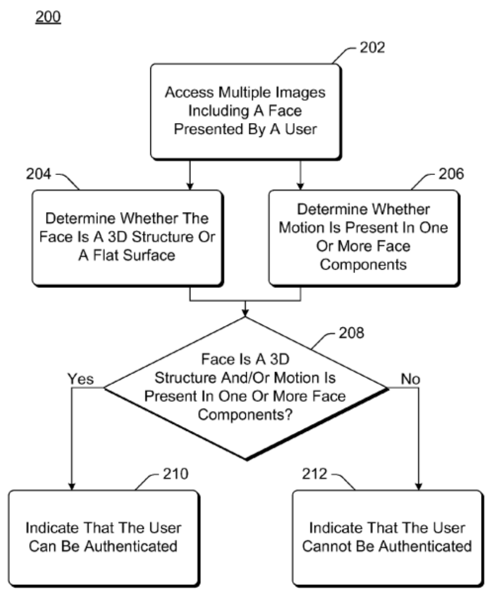


Fig. 2

Id., Fig. 2.

345. To accomplish this, Zhang captures and analyzes multiple images of a user's face using the image capture component 102 implemented in a computing device (e.g., "a desktop computer, a laptop or notebook computer…[or] a cellular or other wireless phone"). *Zhang*, [0012]-[0013], [0016]. As I explained previously, a POSITA would have understood that Zhang's "image capture component" is a camera. *See* §X.D.2.a (1[pre]).

b. 10[a]: capturing at least one first image of the user taken with the camera of the computing device at a first location which is a first distance from the user;

346. In my opinion, Zhang discloses or suggests 10[a] for the reasons discussed in §X.D.2.f (1[d1]).

c. 10[b]: processing the at least one first image or a portion to create first data;

347. In my opinion, Zhang discloses or suggests 10[b] for the reasons discussed in §X.D.2.g (1[d2]).

d. 10[c]: moving the camera from the first location to a second location, the second location being a second distance from the user, or the user moving from the first location to the second location to change the distance between the user and the camera from the first distance to a second distance;

348. In my opinion, Zhang, alone or in combination with Tanii, teaches 10[c]. *See* §X.D.2.h (1[d3]).

e. 10[d]: capturing at least one second image of the user taken with the camera of the computing device at the second distance from the user, the second distance being different than the first distance;

349. In my opinion, Zhang, alone or in combination with Tanii, teaches 10[d] for the reasons discussed in §X.D.2.i (1[d4]).

f. 10[e]: processing the at least one second image or a portion thereof to create second data;

350. In my opinion, Zhang teaches limitation 10[e] for the reasons discussed in §X.D.2.j (1[d5]).

g. 10[f]: comparing the first data to the second data to determine whether expected distortion exist between the first data and the second data which indicated three-dimensionality of the user;

351. In my opinion, Zhang, alone or in combination with Tanii, teaches 10[f]. *See* §§X.D.2.k (1[d6]; describing comparison to look for expected differences), X.D.2.l (1[d7]; describing expected differences as distorting changes from Zhang's homography transformation and Zhang-Tanii's distance-induced distortion approach).

h. 10[g]: authenticating the user when the differences between the first data and the second data have expected distortion resulting from movement of the camera from the first location to the second location or movement of the user from the first location to the second location, which causes the change in distance between the user and the camera.

352. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 10[g] for the reasons discussed in §X.D.2.l (1[d7]).

11. Claim 11

a. 11[a]: The method according to claim 10, further comprising: interpolating the first data and the second data to obtain estimated intermediate data;

353. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 11[a] for the reasons discussed in §X.D.3.a (2[a]).

b. 11[b]: capturing at least one third image of the user taken with the camera of the computing device at a third distance from the user, the third distance being between the first distance and the second distances;

354. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 11[b] for the reasons discussed in §X.D.3.b (2[b]).

c. 11[c]: processing the at least one third image or a portion thereof to obtain third data; and

355. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 11[c] for the reasons discussed in §X.D.3.c (2[c]).

d. 11[d]: comparing the estimated intermediate data with the third data to determine whether the third data matches the estimated intermediate data.

356. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 11[d] for the reasons discussed in §X.D.3.d (2[d]).

12. Claim 12: The method according to claim 10, further comprising verifying the presence of the user's ears in the at least one first image, and verifying the absence or reduced visibility of the user's ears in the at least one second image, wherein the first distance is larger than the second distance.

357. In my opinion, Zhang, alone or in combination with Tanii, teaches claim 12's additional limitation for the reasons discussed in §X.D.4 (claim 3).

13. Claim 15: The method according to claim 10, wherein the computing device is a hand-held device, and the user holds the device at the first and second distances to capture the at least one first image and the at least one second image.

358. In my opinion, Zhang, alone or in combination with Tanii, teaches claim 15's additional limitation for the reasons discussed in §X.D.6 (claim 6).

14. Claim 16: The method according to claim 10, wherein the first data and the second data comprise biometric data.

359. In my opinion, Zhang, alone or in combination with Tanii, teaches claim 16's additional limitation for the reasons discussed in §X.D.7 (claim 7).

15. Claim 17: The method according to claim 10, wherein the first data and the second data comprise a mapping of facial features.

360. In my opinion, Zhang, alone or in combination with Tanii, teaches claim 17's additional limitation for the reasons discussed in §X.D.8 (claim 8).

16. Claim 19: The method according to claim 10, wherein the user's face is held steady and the camera moves from the first location to the second location.

361. In my opinion, Zhang, alone or in combination with Tanii, teaches claim 19's additional limitation for the reasons discussed in §X.D.9 (claim 9).

17. Claim 20: The method according to claim 10, wherein the first data and the second data are maintained on the computing device.

362. In my opinion, Zhang discloses or suggests claim 20's additional limitation.

363. Zhang discloses an image capture component (102) and a live face detection module (104) that can both be implemented on the same computing device. *Zhang*, [0014]. In such instances, in my opinion, a POSITA would have understood that the biometric data would be maintained on the computing device so that all processes would be performed using a single device. If a single, local device is being used for all authentication procedures, there would be no need to store any data on a separate device. Moreover, a POSITA would have understood that biometric data is extremely sensitive, and transmitting that data over networks presents a security risk of being intercepted. A POSITA would have understood that there would be no need to take any security risks by transmitting such data if the single, local device is handling the entire biometric-authentication procedure.

E. Ground 2B: Zhang, Tanii, and Tahk (Claims 4, 13, 14)

1. Motivation to Combine

364. In my opinion, a POSITA would have been motivated to modify Zhang, with or without Tanii, in view of Tahk because Tahk provides a user-friendly way of ensuring that a face presented for facial authentication is properly framed. In my opinion, a POSITA would have understood that providing a real-time preview of what an image would look like prior to capturing the image, as well as providing express prompts—such as written instructions or oval shapes on the live-preview screen—as taught by Tahk, §VII.D (Tahk), would have been particularly useful for

authentication procedures that require capturing multiple images of a face, as taught by Zhang and Tanii.

365. For instance, providing a user a real-time preview would allow the user to actively adjust the position of the camera and/or their orientation of their face to properly frame their face for the image capture. Moreover, providing express prompts and oval shapes sized to guide the user to properly position their face would ensure images best suited for facial recognition can be captured, and that the faces would be captured from sufficiently different perspectives and/or distances to ensure the three-dimensional verification taught by Zhang, alone or in combination with Tanii, could be performed. This is particularly important for facial-recognition systems, which generally are known to have difficulty matching faces across different views. *See, e.g.*, Ex-1018, 32.

366. Moreover, a POSITA would have known that providing real-time image feedback, written instructions, and oval shapes to frame a face during a facial authentication process were all well-known and conventional techniques to provide user feedback during image capture as of the time of the invention. *See, e.g.*, Ex-1034, 7:16-8:7, Figs. 6B-7C (providing "an example of an interface used upon registering a facial image of a person to be authenticated" in which "the image of this person is displayed on a monitor" during registration, and oval-shaped prompts to indicate the size/distance of the user from the face authentication sensor); Ex-

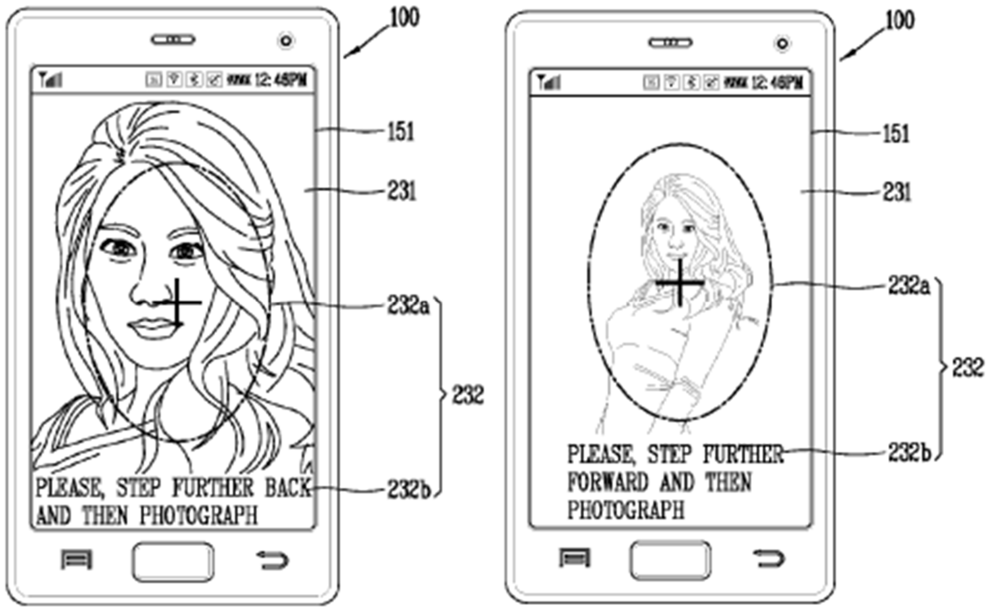
1035, 5:31-32 ("The computing device may present prompts that instruct the user to perform one or more liveness gestures"), 6:3-4 (same).

2. Claim 4: The system according to claim 1, wherein the computing device is configured to display one or more prompts on a screen of the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

367. In my opinion, Zhang combined with Tanii and Tahk teaches claim 4.

368. Zhang discloses taking a series of images sufficient to calculate a homography matrix. *See, e.g.*, *Zhang*, [0026], Figs. 1, 3. A POSITA would have understood that Zhang already discloses, or that Zhang combined with Tanii teach, taking a series of images at different distances between the face and the camera. *See* §§X.D.1 (motivation), X.D.2.i (1[d4]). However, Zhang and Tanii do not expressly teach providing a series of prompts to a user to guide them through different camera positions that would enhance calculations of the homography matrix.

369. Tahk, however, teaches that using one or more prompts on a screen ensures images of the face are captured at the correct distances. *See, e.g.*, Tahk, Figs. 8A-B ("Please step further back" and "Please step further forward," and presenting an oval to frame the face at the correct distance).



In my opinion, a POSITA would have been motivated by Tahk to modify Zhang, whether alone or in combination with Tanii, to expressly prompt a user to alter the distance of the camera in order to either capture sufficiently different images to perform a homography transformation (Zhang) or to capture an image with distance-induced distortions (Tanii) to ensure the images could be used to distinguish live from two-dimensional images of faces. *See* §X.E.1 (motivation).

3. Claim 13: The method according to claim 10, wherein the computing device is configured to display one or more prompts on a screen of the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

370. In my opinion, Zhang combined with Tanii and Tahk teaches claim 13 for the reasons discussed in §X.E.2 (claim 4).

4. Claim 14: The method according to claim 13, wherein the one or more prompts are ovals on the screen within which the face of the user is placed to capture the at least one first image and the at least one second image.

371. In my opinion, Zhang combined with Tanii and Tahk teaches claim 14.

372. Neither Zhang nor Tanii expressly teach using prompts to guide a user during the facial-authentication process. Tahk, however, teaches using oval prompts to frame a user's face. *See* §§X.E.2 (claim 4), X.E.3 (claim 13). A POSITA would have been motivated to modify Zhang, alone or in combination with Tanii, to provide oval-shaped prompts because they are a natural shape to appropriately size and frame a face at different distances. *See* §X.E.1 (motivation).

F. Ground 2C: Zhang, Tanii, and Hoyos (Claim 18)

1. Motivation to Combine

373. Zhang discloses implementing a process to verify the three-dimensionality of a user's face by capturing a series of images of a user using a mobile computing device, such as a phone or laptop. *See, e.g.*, *Zhang*, [0013]. In my opinion, a POSITA would have understood that mobile devices are often provided with user-facing cameras, particularly in mobile devices.

374. Although Zhang discloses using homography transformation to distinguish real, three-dimensional faces from pictures of a face, *see* §VII.C (Zhang), in my opinion, a POSITA would have appreciated that Zhang's process may be spoofed by presenting a non-planar picture of a face, because Zhang's homography

transformation process specifically looks for a *planar* structure in images that can be transformed nearly identically from one perspective to another. *See id.* Thus, Zhang's system could possibly be subverted by bending the picture in a way to trick the system, or applying the picture to a three-dimensional shape. For this reason, in my opinion, a POSITA would have therefore been motivated to look for secondary methods to ensure the user's face is from the user, and not a spoofer. In my experience, biometric systems often included multiple independent checks to ensure the liveness of the user (Derakhshani, for instance, is one example that provided separate spatial, behavioral, and reflectance metrics to each independently confirm liveness of the user).

375. A POSITA would have known that the use of reflectance of light off a face was a well-known liveness check. *See, e.g.*, *Derakhshani*, 18:5-19:11; *Hoyos*, [0018]-[0019], [0033]-[0036], Figs. 2-3. And Hoyos exemplifies this process by disclosing the use of patterned images intended to reflect off of the user's face. *Hoyos*, [0018]-[0019], [0033]-[0036], Figs. 2-3. In my opinion, a POSITA would have been motivated to incorporate a secondary liveness check based on reflectance, consistent with Hoyos, to ensure a user is not attempting to spoof Zhang's homography-transformation process.

376. In my opinion, a POSITA would have had a reasonable expectation of success in making this modification because Zhang's homography transformation

and Hoyos's reflectance measure operate on two distinct, modular principles that can be operated together; Zhang requires two images from two different perspectives, and Hoyos requires reflecting different light patterns during image capture.

2. Claim 18: The method according to claim 10, further comprising displaying an image on a screen of the computing device while capturing the at least one first image and/or the at least one second image, and processing the at least one first image and/or the at least one second image to detect a reflection of the displayed image off of the user's face.

377. In my opinion, Zhang combined with Tanii and Hoyos teaches claim 18's additional limitation.

378. Zhang does not expressly disclose displaying an image on the device's screen when capturing images, or detecting a reflection of the displayed image off of the user's face in the captured images. Hoyos, however, teaches that measuring reflectance of displayed images is a well-known method to verify the liveness of the user. *Hoyos*, [0018]-[0019]; [0033]-[0035]. In my opinion, a POSITA would have been motivated to modify Zhang to incorporate Hoyos's reflectance detection to provide an additional verification that the user is presenting a real, three-dimensional face. *See* §X.F.1 (motivation).

XI. ’606 PATENT: DETAILED EXPLANATION OF GROUNDS

A. Ground 1A: Derakhshani and Tanii (Claims 1-4, 6-7, 9-16, 18-20)

1. Motivation to Combine

379. In my opinion, a POSITA would have been motivated to combine Derakhshani and Tanii because both concern identifying and accounting for the three-dimensional nature of a face when capturing an image. They differ, however, in what principles are used to account for the face’s three-dimensionality. Derakhshani, for instance, uses changes in focus distance (e.g., image resolution for structures imperfectly in focus) and/or parallax effect to determine whether a face has depth. *See* §VII.A (Derakhshani). And although Tanii is not expressly directed to *evaluating* whether a face has depth, Tanii exemplifies the well-known distortions caused by the interaction between the camera’s lens and the three-dimensional nature of the face at different distances. *See* §VII.B (Tanii). A POSITA would have appreciated, therefore, that Tanii recognizes another alternative to evaluating the depth of a face, consistent with Derakhshani’s existing two approaches.

380. A POSITA would have recognized, for instance, that Derakhshani’s focus-distance approach and Tanii’s evaluation of distance-induced distortions are both attributable to classical optical effects such as refraction and diffraction caused by (among other factors) different distances between the camera and the object(s) being captured. *Derakhshani*, 16:57-60 (“Degree of focus is a measure of the extent to the image of the landmark is blurred by optical effects … (e.g., due to *diffraction*

and convolution with the aperture shape."); *Tanii*, [0048] (noting the "unnatural image" is caused by the angles of the face relative to the angle of the camera lens).

381. Derakhshani and Tanii differ, however, in the type of effect that is occurring. Specifically, Derakhshani takes advantage of the blurring of objects that are at distances *other than* the camera's focal plane (referred to by photographers as a "bokeh effect"), which makes those objects appear unfocused. *Derakhshani*, 16:54-57; §VII.A (Derakhshani). By adjusting the focus distance (or position of the focal plane by moving the camera) and evaluating when objects (or features of an object) in an image are clear versus when they are blurry, distance information can be derived. *Derakhshani*, 16:51-63; §VII.A (Derakhshani).

382. Tanii is more specifically concerned with a type of radial distortion that arises due to the interaction of certain (e.g., wide-angle) lenses and the three-dimensional nature of the face. §VII.B (Tanii). As Tanii explains, the convex shape of a three-dimensional face, when placed near the lens, exacerbates this type of distortion. *Tanii*, [0048]; §VII.B (Tanii). Thus, particularly when a camera incorporates a wide-angle lens, images of a face close to the camera will exhibit significant radial distortion in-part because of the distances between different facial features and the lens, and in-part because the face occupies both the center and the periphery of the camera's field of view so differences in radial distortion are more apparent. *Tanii*, [0047]; §VII.B (Tanii). But when the face is further from the camera

and occupies less of the image, the distortion will be less apparent because the face is more centered on the region of the lens where radial distortion is not as severe, and there is sufficient distance for the light rays from the face to strike this central portion of the lens. *Tanii*, [0047]; §VII.B (Tanii).

383. In my opinion, a POSITA would have appreciated that when evaluating multiple images taken at either different *focus* distances or *actual* distances, these different effects serve to provide information about an object's depth. In other words, a POSITA would have understood that Tanii teaches another obvious alternative to Derakhshani's existing two approaches to evaluate whether a face being captured is three-dimensional or not.

384. That said, a POSITA would have also had specific reasons to substitute Derakhshani's existing approaches with Tanii's distance-induced distortion analysis in certain circumstances. A POSITA would have understood, for instance, that implementing Derakhshani's focus-distance approach requires a camera with a sufficiently sized sensor and lens that could provide enough sensitivity to distinguish small differences in depth on the scale of a few centimeters when trying to evaluate the depth of a face. *See Derakhshani*, 16:48-51; Ex-1029, 3 (A 200mm lens focused at 12ft will have a smaller depth of field compared to a 20mm lens focused at 12ft).

385. But a POSITA would have also understood that the cameras typically found in mobile devices—especially around the 2014 timeframe—do not have this

ability; mobile devices typically incorporate wide-angle lenses to capture a wide field of view, with a fixed focal length and a large depth of field because of their small size. *Tanii*, [0007]; Ex-1030 ("Other features of a smartphone are obvious but worth stating, they almost always are fixed focal length, fixed aperture, with no shutter, sometimes with an ND filter (neutral density) and generally not very low F-number. In addition to keep modules thin, focal length is usually very short, which results in wide angle images with lots of distortion."). With such limited-capability cameras, it was known that distortions would therefore largely be a product of the lens shape and distance between the object and the lens. *See* Ex-1017, 177 ("The amount of spherical aberration, when the aperture and focal length are fixed, varies with both the ***object distance*** and the lens shape."). In other words, there is not enough room in mobile devices to incorporate large image sensors with small F-numbers (a measure of light-gathering ability of the camera) to allow these cameras to fine-tune the focus distance and induce blurring of out-of-plane objects. That is why, for instance, the iPhone introduced its "Portrait Mode" (in 2016, a few years after the earliest possible effective date) as a software-based *simulation* of the blurring effect that can only be achieved by much larger cameras. Ex-1031 (noting how blurring backgrounds was "previously only capable on DSLR cameras" prior to the iPhone's software-based "bokeh" effect).

386. For this reason, in my opinion, a POSITA would have been motivated to modify Derakhshani to capture at least two images at different *actual* distances and evaluate whether one exhibits more distance-induced distortion than the other, as suggested by Tanii. A POSITA would have been especially motivated to make this change when implementing biometric authentication in a mobile device as Derakhshani already envisions. *Derakhshani*, 5:23-26. A POSITA would have found such a modification obvious because both techniques merely involve the application of different well-known optics principles relating camera design and object's distance from the camera, and would have had a reasonable expectation of success in doing so because Tanii already taught a mechanism to identify such distance-induced distortions and thus indicate when the face being captured has depth. *See, e.g.*, *Tanii*, [0056].

387. Although Derakhshani separately discloses a process to verify the three-dimensionality of a face using parallax, in my opinion, a POSITA would have understood that evaluating for distance-induced distortion consistent with Tanii would be easier for users on a mobile device. Specifically, a POSITA would have naturally understood that mobile devices such as phones or laptops typically capture images of users at arm's length distances because that is how these devices are used (at arm's length). Moreover, a POSITA would have appreciated that facial features do not have *significant* differences in their depth (on the order of a few centimeters,

as opposed to meters between the face and a background). Thus, to evaluate for parallax at hand-held distances with suitable accuracy, a POSITA would expect that the user would need to move their device around their head, or could simulate a parallax effect by rotating their head around a stationary camera to create substantial differences in perspective and thus more parallax to more accurately verify the face as three-dimensional. But to do so would have involved moving the device out of the user's line of sight, meaning the user could not see exactly what they are capturing or know if what they were capturing is sufficient.

388. Evaluating for distance-induced distortions when the camera is held at different distances consistent with Tanii, however, could be accomplished while keeping the device directly in the user's direct line of sight, and would therefore be easier for users to verify that their face is, in fact, three dimensional. But, in my opinion, a POSITA would have also appreciated that biometric security is always subject to spoofing, and thus would have known that evaluating for distance-induced distortion consistent with Tanii could be *supplemented* by also evaluating for any parallax.

2. Independent Claim 1

a. 1[pre]: A method for verifying three-dimensionality of a user's face using images of the user's face captured using a camera equipped computing device, the method comprising:

389. If the preamble is limiting, in my opinion, Derakhshani discloses or suggests it.

390. Derakhshani discloses systems and methods for using a camera-equipped computing device for "biometric authentication." *See Derakhshani*, 1:11-25, 5:22-27, 6:3-5, 9:10-22, 18:1-3. Although Derakhshani uses the eye as the primary means of authentication, *see, e.g.*, *id.*, Abstract, as part of the ocular-authentication process, Derakhshani also verifies that the user's face is three-dimensional by capturing multiple images of a user's face at different focus distances or from different perspectives to calculate a "spatial metric" representing the face's three-dimensionality. *Id.*, 1:11-25, 3:14-15, 16:44-18:4.

b. 1[a]: capturing at least one first image of the user taken with the camera of the computing device at a first distance from the user;

391. In my opinion, Derakhshani discloses or suggests limitation 1[a].

392. Derakhshani discloses that, as part of the process to verify that the face is in fact three-dimensional, "two or more images of a subject" are captured using the camera of the computing device. *Derakhshani*, 1:44-46, 16:44-17:11, 17:45-18:4.

393. In my opinion, a POSITA would have understood that Derakhshani captures an image at a first distance. *Derakhshani*, 16:44-17:11. Specifically, a POSITA would have understood that there *must* be some distance between the camera and the face to capture the first image—enough so that the camera's field of view encompasses the face. If there were no distance between the camera and the face (e.g., if the camera were pressed up against the user's skin), then the field of view would be limited to just that patch of skin and the face would not be captured in the image, which would be useless for the three-dimensional verification process.

c. 1[b]: processing the at least one first image to obtain first biometric data from the at least one first image;

394. In my opinion, Derakhshani discloses or suggests limitation 1[b].

395. Derakhshani discloses that, as part of the process to verify that the face is in fact three-dimensional, "a landmark (e.g., an iris, an eye corner, a nose, an ear, or a background object) may be identified and located in the plurality of images." *Derakhshani*, 16:44-54 (focus-distance approach), 17:45-64 (parallax approach).

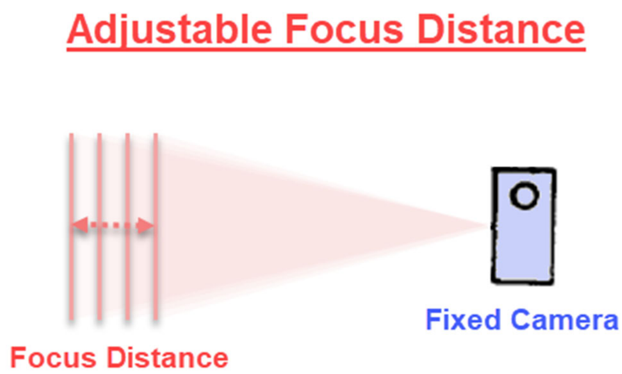
396. In my opinion, a POSITA would have understood that Derakhshani's identification of facial landmarks constitutes "biometric data" because "biometric data" generally refers to unique physical characteristics of an individual, which would include the positions of "landmarks" such as a user's eyes, nose, ears, and other such features. *See, e.g.*, Ex-1018, 2 ("biometric recognition can be defined as the science of establishing the identity of an individual based on the physical and/or

behavioral characteristics of the person."), 100-103 (describing the types of biometric data about a face used for facial-recognition systems, "such as the structure of the face components (e.g., eyes), [and] the relationship between facial components").

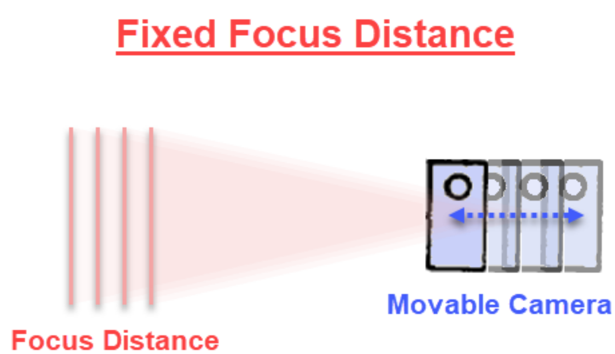
d. 1[c]: capturing at least one second image of the user taken with the camera of the computing device at a second distance from the user, the second distance being different than the first distance;

397. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 1[c].

398. Derakhshani discloses capturing "two or more images of a subject" using the camera. *Derakhshani*, 1:44-46, 16:44-17:11, 17:45-18:4; §XI.A.2.b (1[a]). When utilizing Derakhshani's focus-distance approach to evaluate depth, however, a POSITA would have understood that adjusting the *focus* distance of the camera does not require changing the *actual* distance between the camera if a stationary camera is capable of adjusting its lens position with respect to the image sensor. *See* §VII.A (Derakhshani explaining operation of the focus-distance approach).



But if the camera has a fixed focus distance (i.e. position of the lens with respect to the image sensor), as is found in many mobile devices (*see* §XI.A.1), a POSITA would have been motivated to instead implement Derakhshani's focus-distance approach by changing *actual* distance to capture multiple images, as shown below:

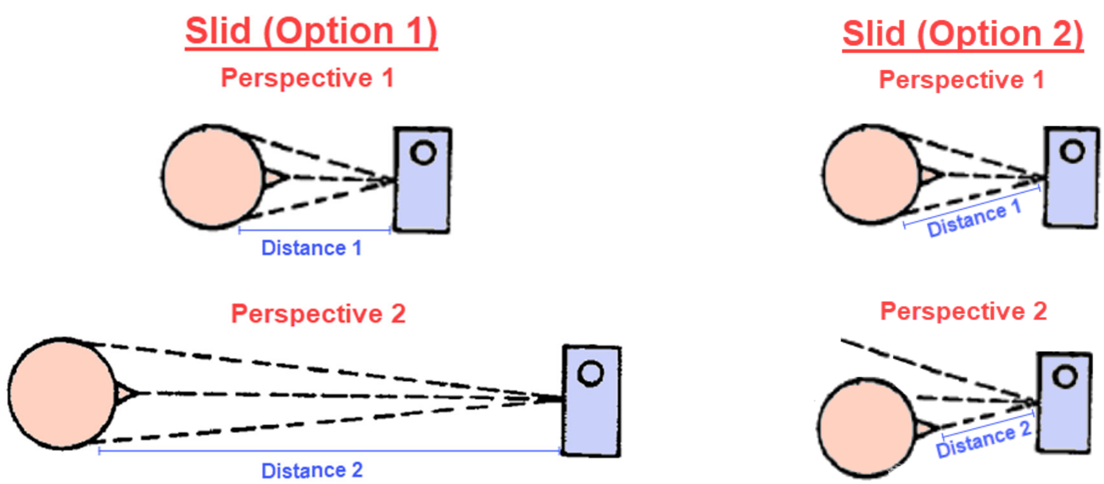


In other words, even if the focus distance of the camera cannot be changed, the "slices" of a face at different depths can be evaluated by moving the camera.

399. Regardless, in my opinion, a POSITA would have understood that Derakhshani's parallax approach captures multiple images from multiple distances, because Derakhshani discloses that "[a] plurality of images [are] taken from different perspectives on the subject," such as: (1) when "a single camera [is] rotated

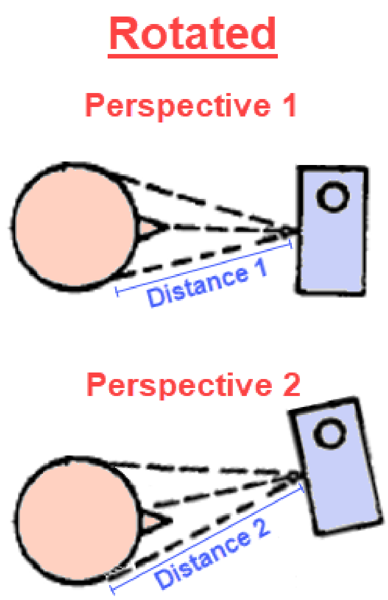
or slid slightly"; (2) "a user is prompted to move" between image captures; or (3) the sensor moves naturally, such as "where the sensor is a camera in a hand-held user device (e.g., a smartphone or tablet) [that] may naturally move relative to the users face due to involuntary haptic motion." *Derakhshani*, 17:45-18:4.

400. In my opinion, a POSITA would have understood that Derakhshani's use of the term "slid" means either of two things: (1) the camera is displaced front-to-back to increase or decrease the distance from the face; or (2) the camera is displaced side-to-side, both of which are depicted below:



In either case, a parallax effect would be evident if the face were three-dimensional because of the different perspectives of the face captured in each. For instance, a POSITA would have recognized that, with a front-to-back translation, more of the periphery of the face would be captured by the camera, and there may be other optical effects (e.g., distance-induced distortion) that are more apparent in the closer image than the further one. And with side-to-side translation, more features on the side of the face the camera favors would be captured, but features on the other side of the face may be obstructed due to the face's three-dimensionality.

401. Moreover, a POSITA would have understood that Derakhshani's use of the term "rotated" means the camera itself is rotated relative to the face. I have provided an example of rotation below that also includes some side-to-side translation to keep the face centered on the camera.



402. As these exemplary figures demonstrate, however, a POSITA would have understood that, regardless of whether the camera is "slid" or "rotated," distances between facial landmarks and the camera will change. In my opinion, a POSITA would have understood that any of these options results in "capturing at least one second image of the user … at a second distance from the user, the second distance being different than the first distance," as claimed, because there is no one

single "distance" between the camera and a three-dimensional user when changing the position/perspective of the camera; some distances will always change. However, even if the claims were limited to a front-to-back translation to change the *overall* distance between the camera and the user, a POSITA would have understood that Derakhshani discloses or suggests as much.

403. But even if Derakhshani does not expressly disclose taking two images at different distances, in my opinion, a POSITA would have been motivated to look to the differences in degree of distance-induced distortions exemplified by Tanii as an alternative or additional evaluation of the three-dimensionality of the face besides Derakhshani's focus-distance and parallax approaches. §XI.A.1 (motivation). When making this modification, a POSITA would have been motivated to modify Derakhshani in view of Tanii to expressly capture a second image at a second distance, and look for more distance-induced distortions in one image compared to the other to determine whether the face has depth. §XI.A.1 (motivation). Moreover, a POSITA would have had a reasonable expectation of success in making this modification because Tanii already taught a mechanism to identify such distance-induced distortions and thus indicate when the face being captured has depth. *See, e.g.*, *Tanii*, [0056].

e. 1[d]: processing the at least one second image to obtain second biometric data based on the at least one second image;

404. In my opinion, Derakhshani discloses or suggests limitation 1[d].

405. Derakhshani discloses processing the captured images to identify biometric "landmarks" in the face as part of the three-dimensional verification process. *Derakhshani*, 17:45-52; §XI.A.2.c (1[b]).

f. 1[e]: comparing the first biometric data with the second biometric data to determine whether the first biometric data matches the second biometric data;

406. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 1[e].

407. Derakhshani discloses that, regardless of whether the focus-distance approach or parallax approach is used, biometric features are identified and compared across each of the images to match them across multiple images. *Derakhshani*, 16:66-17:2 ("comparing the degree of focus for a landmark in images with different focus distances."); 17:45-64 (evaluating relative displacement of identified landmarks across images). In my opinion, a POSITA would have also appreciated that, when modifying Derakhshani to evaluate for distance-induced distortions exemplified by Tanii, biometric data would also be compared between images to then determine whether they exhibit distance-induced distortion relative to each other, consistent with Derakhshani. In other words, a POSITA would have

understood that each of the approaches to evaluate an object's depth taught by Derakhshani (focus distance or parallax), whether alone or combined with Tanii (distance-induced distortion), would require comparing biometric data points between multiple images. For this reason, a POSITA would have known, or at least been motivated to, match the biometric data between each of the images (such as matching the ears, eyes, and nose in one image to those same features in another) to evaluate the differences between them in different images.

g. 1[f]: comparing the first biometric data to second biometric data to determine whether differences between the at least one first image and the at least one second image match expected differences resulting from movement of the camera or the user which changed the distance between the user and camera from the first distance to the second distance;

408. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 1[f].

409. Derakhshani discloses that, when comparing the first biometric data to the second biometric data, a determination is made whether differences between the two exist. *Derakhshani*, 16:66-17:2 (for focus distance, "[b]y comparing the degree of focus for a landmark in images with different focus distances, the distance from the sensor to the landmark may be estimated."), 17:55-59 (for parallax, "[i]f all the landmarks in the images undergo the same apparent displacement due to the relative

motion of the sensor…then the subject viewed by the camera has a higher likelihood of being a two-dimensional spoof attack.”).

410. In my opinion, Derakhshani describes a comparison between images that looks for “expected” differences consistent with how the ’606 Patent uses the term because one would *expect* that following either the focus-distance or parallax approaches Derakhshani discloses would produce specific differences: the focus-distance approach would capture some images where certain facial features are blurred and others where those same features are clear, and the parallax approach would produce expected relative displacements of certain facial features depending on the change of perspective and distance between the specific features and the camera lens. Moreover, Derakhshani’s focus-distance approach looks for expected differences in the blurriness or clearness of facial landmarks by changing the *focus* distance (for adjustable-focus cameras) or *actual* distance (for fixed-focus cameras), and Derakhshani’s parallax approach looks for expected differences in the relative displacement of different facial landmarks by changing the *actual* distance alone. *See* §XI.A.2.d (1[c]).

411. Relatedly, a POSITA would have been particularly motivated to configure Derakhshani to capture images with specific, pre-defined configurations (e.g., a specific set of focus distances, or a specific position of the camera relative to the face) to minimize the variability between the images used for facial recognition

and specifically tailor the system to look for expected changes between images. For example, Derakhshani's focus-distance approach (with its loss of spatial frequency) would improve its performance if images were acquired with the face at different distances from the camera. Doing this with two or more distances would remove range ambiguity and decrease the variance in estimates of the distance from the camera to particular features. *See, e.g.*, Ex-1018, 32 (noting how facial-recognition systems often require controlling conditions such as a "fixed and simple background with controlled illumination" because "systems … have difficulty in matching face images captured from two different views, under different illumination conditions, and at different times."). In other words, rather than permitting users to change the focus distance or perspective of the camera any way they wish, which would require a system that could account for such variabilities, having the user follow a pre-determined protocol to capture images at set focus distances or perspectives would simplify the matching process.

412. In my opinion, however, a POSITA would have also understood that, when utilizing the distance-induced distortion approach exemplified by Tanii, the images captured from that process would also exhibit expected distortion based on the distance between the camera and the face. §XI.A.1 (motivation). In my opinion, a POSITA would have been motivated to look for and utilize these expected differences in distortion as an alternative or supplemental verification of three-

dimensionality of a face in Derakhshani, particularly in mobile devices that incorporate wide-angle lenses. §XI.A.1 (motivation). In doing so, a POSITA would have understood that verifying a three-dimensional face using distance-induced distortion would be accomplished by matching the positions of biometric features across the first and second images—as Derakhshani already discloses—but rather than look for blurriness/clearness or parallax of those biometric features, the images would instead be evaluated for expected differences in the distortion of those features caused by the distance-induced distortion. §XI.A.1 (motivation).

h. 1[g]: determining that the user's face is three-dimensional when:

413. In my opinion, Derakhshani discloses or suggests limitation 1[g].

414. Specifically, as I explain in further detail below, Derakhshani determines whether the user's face is three-dimensional or not depending on certain conditions. *Derakhshani*, 16:44-18:4.

i. 1[h]: the first biometric data does not match the second biometric data; and

415. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 1[h].

416. Derakhshani discloses that a face is determined to be three-dimensional when mismatches exist between the biometric landmarks (e.g., biometric data) using either the focus-distance or parallax approach. §XI.A.2.f (1[e]). Specifically, the

focus-distance approach evaluates whether facial landmarks are blurry in one image and clear in another, indicating depth. *Id*. And the parallax approach evaluates whether different facial landmarks are displaced by different amounts, also indicating depth. *Id*. In other words, a POSITA would have understood that, under either approach, Derakhshani is looking for a mismatch between the first and second biometric data to determine whether the face is three-dimensional or not.

417. Similarly, when modifying Derakhshani in view of Tanii, a POSITA would have understood that a mismatch between facial features (e.g., their size, position, and/or proportion) across two images (i.e., the first and second biometric data) caused by different degrees of distance-induced distortion would also indicate the face is three-dimensional. §§XI.A.1 (motivation); XI.A.2.d (1[c]).

418. In sum, in my opinion, Derakhshani alone or combined with Tanii looks for mismatches between first and second biometric data to indicate whether a face is three dimensional.

j. 1[i]: the second biometric data has the expected differences as compared to the first biometric data resulting from the change in distance between the user and the camera when capturing the at least one first image and the at least one second image.

419. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 1[i].

420. Derakhshani discloses a process that determines a face is three-dimensional when expected differences exist between the biometric landmarks (e.g., biometric data) using either the focus-distance or parallax approach. §§XI.A.2.g (1[f]), XI.A.2.i (1[h]). Moreover, Derakhshani's focus-distance approach looks for expected distortions in the blurriness or clearness of facial landmarks by changing *actual* distance (for fixed-focus cameras), and Derakhshani's parallax approach looks for expected differences in the relative displacement of different facial landmarks by changing the *actual* distance alone. *See* §XI.A.2.d (1[c]).

421. If Derakhshani's parallax approach for some reason cannot be considered to already disclose this limitation, however, in my opinion, Derakhshani combined with Tanii does. *See* §§XI.A.2.d (1[c]), XI.A.2.g (1[f]). Specifically, in my opinion, a POSITA would have appreciated that when modifying Derakhshani to specifically look for differences caused by distance-induced distortions (consistent with Tanii), a three-dimensional face would be indicated when one of the two sets of biometric data exhibits expected distance-induced distortions due to the change in distance of the camera. *See* §§XI.A.2.d (1[c]), XI.A.2.g (1[f]).

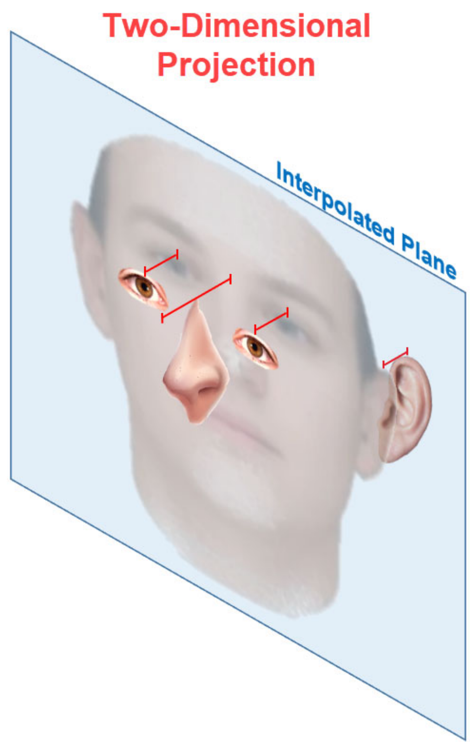
3. Claim 2

a. 2[a]: The method according to claim 1, further comprising: interpolating the first biometric data and the second biometric data to obtain estimated intermediate biometric data;

422. In my opinion, Derakhshani, alone or combined with Tanii, teaches 2[a]'s additional limitation.

423. Derakhshani discloses that one optional way the three-dimensionality of a face can be verified is by fitting "the locations of multiple landmarks…to the closest two dimensional plane and the average distance of the landmarks from this fit plane can be determined as the spatial metric." *Derakhshani*, 17:12-26.

424. In my opinion, a POSITA would have understood that Derakhshani identifies the position of this two-dimensional plane relative to the facial landmarks by matching up the landmarks that appear across different images, and then calculating an average distance between the various landmarks at their identified three-dimensional positions based on the series of images. If this average distance between the plane and landmarks is sufficiently large, the face is determined to be three dimensional. I have provided a graphic depiction of this process for demonstration purposes, with the plane being identified in blue, the position of the facial landmarks determined by analyzing the series of images, and the red line between the facial landmarks and the plane representing the distance between the two:



A POSITA would have therefore understood that the plane constitutes “interpolated” biometric data, because the term “interpolated” is generally understood to mean “to insert between other things” or “estimate values of (data or a function) between two known values.” Ex-1026, 654.

425. A POSITA would have further understood that a similar projection could be determined from Derakhshani’s parallax process. However, rather than use a two-dimensional plane projection based on images taken from the same perspective (like Derakhshani’s focus-distance approach), images taken from two different perspectives to evaluate for parallax would be better suited by using a three-dimensional model approach. For instance, it was well-known that three-

dimensional modeling of the perspective of a face could be estimated using images of a face. *See, e.g.*, Ex-1018, 117 (noting use of 2D and 3D modeling techniques to account for variations in perspective); Ex-1027, 8 (describing a method that involves estimating the position of a face to locate and match facial features); Ex-1015 (describing the generation of a three-dimensional model of a face based on two-dimensional images). In fact, Derakhshani expressly recognizes that the spatial metric can be determined by determining deviations between the images captured and a three-dimensional model of the face. *Derakhshani*, 17:27-44.

426. A POSITA would have understood that these three-dimensional models to which images are compared would be an "interpolation." In other words, a POSITA would have recognized, or at least been motivated to implement Derakhshani's parallax approach by constructing a three-dimensional, interpolated model based on the series of images captured to either: (1) compare it to an existing three-dimensional model generated during enrollment; or (2) determine whether the series of images can create a suitable three-dimensional model, which itself would indicate that the imaged face has three dimensions.

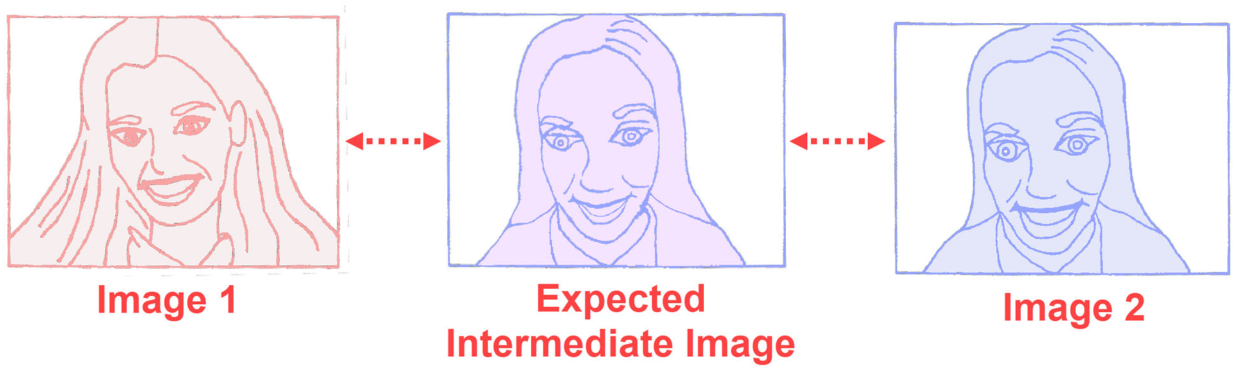
427. Based on a POSITA's understanding of Derakhshani, a POSITA would have further been motivated to derive interpolated biometric data based on the combination of Derakhshani and Tanii. Specifically, Tanii discloses that distance-induced distortions increase as distance between the face and camera decreases. *See*

Tanii, [0048]. A POSITA would have therefore understood that, all else being equal, distance-induced distortion depends on the distance between the user and the camera, and thus any set of images as the distance between the user and camera changes will have different degrees of facial warping, similar to the gradual changes in facial distortion that appears in the series of images below (although these images also altered the focal length of the camera to ensure the face remains a constant size in the frame, rather than just distance:

Ex-1022.

428. A POSITA reading Derakhshani—which discloses generating intermediate projections to evaluate depth—in view of Tanii therefore would have been motivated to interpolate intermediate biometric data with an intermediate, interpolated amount of distance-induced distortion based on the two images captured to create an array of potential distance-induced distortions that would indicate depth in a three-dimensional face, such a gradual projection from the lines of the facial

landmarks shown in blue (the distorted image) to the lines of the facial landmarks shown in red (the distorted image).



Performing this type of interpolation between the distorted and undistorted images would allow for further comparison with additional images, for instance, to ensure the distance-induced distortion matches expectations across a range of distances if the face were truly three-dimensional, as depicted below:

In my opinion, a POSITA would have understood that this would be akin to Derakhshani's modeling approach, but rather than build a model based on a two-dimensional projected plane or three-dimensional model of a head, the model would be of various degrees of expected distance induced distortion with which the captured images could be compared.

b. 2[b]: capturing at least one third image of the user taken with the camera of the computing device at a third distance from the user, the third distance being between the first distance and the second distance;

429. In my opinion, Derakhshani, alone or combined with Tanii, teaches 2[b]'s additional limitation.

430. Derakhshani discloses that, as part of the three-dimensional verification process, "a plurality" of images may be captured. *Derakhshani*, 16:44-46 (focus-distance embodiment), 17:45-47 (parallax embodiment).

431. In my opinion, a POSITA would have also understood generally that capturing more images would provide increased accuracy in verifying a three-dimensionality of the face because there would be more samples to evaluate, with the trade-off being an increase in processing demands. For instance, taking four images using the focus-distance approach would enable precise depth information of at least four facial landmarks that sit on different planes, such as the ears, eyes, mouth, and nose. For the parallax approach, fewer images would likely be necessary depending on how significant the change of perspective is—e.g., rotating the camera may reveal parallax in as little as two images, whereas sliding may benefit from an additional image—since that approach looks for displacement of facial landmarks

due to the change in perspective, which does not depend on taking images at different "slices" of depth like the focus-distance approach.

432. A POSITA would have also appreciated as a general matter that, in any set of images with more than two images captured at different distances, *see* §XI.A.2.d (1[c]), one image would have a minimum distance and one would have a maximum distance, with the rest existing somewhere in between. For instance, in a set of distances of 10cm, 50cm, and 1m, 10cm would be the minimum, and 1m would be the maximum, with 50cm existing in between the two.

433. Moreover, when modifying Derakhshani in view of Tanii to interpolate intermediate biometric data attributable to distance-induced distortions, a POSITA would have been further motivated to capture a third image at a distance that correlates to one of the interpolated data sets for further authentication of three-dimensional depth of the face in the captured images. *See* §§XI.A.2.d (1[c]), XI.A.3.a (2[a]).

c. 2[c]: processing the at least one third image to obtain third biometric data based on the at least one third image; and

434. In my opinion, Derakhshani, alone or combined with Tanii, teaches 2[c]'s additional limitation.

435. Derakhshani discloses processing the images to identify feature landmarks in each of the images. *See* §§XI.A.2.c (1[b]), XI.A.2.e (1[d]). Therefore,

in my opinion, a POSITA would have found it obvious to process any images captured by the camera to derive biometric data so that the biometric data could be compared between images, consistent with Derakhshani.

436. Furthermore, when modifying Derakhshani in view of Tanii, in my opinion, a POSITA would have found it obvious to acquire a third image and extract biometric data from the third image to compare it to the interpolated positions of the biometric data based on the first and second images. *See* §XI.A.3.b (2[b]).

d. 2[d]: comparing the estimated intermediate biometric data with the third biometric data to determine whether the third biometric data matches the estimated intermediate biometric data.

437. In my opinion, Derakhshani, alone or combined with Tanii, teaches 2[d]'s additional limitation.

438. Derakhshani discloses comparing multiple images to the two-dimensional projection or three-dimensional model interpolated from those images. *Derakhshani*, 17:12-44; §XI.A.3.a (2[a]). In my opinion, a POSITA would have therefore understood that once an interpolated projection or model of the face is generated consistent with Derakhshani, if a third image is captured, that too would be compared to the projection or model to estimate the distance or deviation of any facial landmarks in that image from the projection or model. §XI.A.3.a (2[a]).

439. Furthermore, in my opinion, a POSITA modifying Derakhshani in view of Tanii would have considered it obvious to acquire a third image and extract

biometric data from it to compare the biometric data to the interpolated, expected positions of the biometric data to determine if there is a match between the two. *See* §§XI.A.3.a (2[a]), XI.A.3.b (2[b]).

4. Claim 3: The method according to claim 1, further comprising verifying the presence of the user's ears in the at least one first image, and verifying the absence or reduced visibility of the user's ears in the at least one second image, wherein the first distance is larger than the second distance.

440. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 3's additional limitation.

441. Derakhshani discloses that "a landmark…an ear…may be identified and located." *Derakhshani*, 16:51-54; *see also* 17:14-19.

442. In my opinion, a POSITA would have appreciated that, when following Derakhshani's focus-distance approach, in some captured images, the ear would have reduced visibility (i.e., it is blurry) when it does not lie in the focal plane, and would be clear (e.g., a verified presence) when it does lie in the focal plane. *See* §XI.A.2.d (1[c]). A POSITA would have also appreciated that, in some circumstances, distances in which the ear would be clear would be greater than those with reduced visibility, such as when the focal plane is aligned behind the ears.

443. And when following Derakhshani's parallax approach, a POSITA would have also appreciated that some perspectives would obviously capture one or more ears (when both are exposed, such as a front-facing image from sufficient

distance), and other perspectives would only capture one (when the other is obstructed by the head), which would indicate that the user's face is three-dimensional. *See* §XI.A.2.d (1[c]) (providing an example figure in which a camera rotation would obfuscate one ear). If the object being captured were a two-dimensional picture of a face with ears, however, any perspective would capture both ears because the ears exist on a single plane of the picture.

444. Similarly, when modifying Derakhshani in view of Tanii to use distance-induced distortions to verify the three-dimensional nature of a face, Tanii teaches that the absence and presence of an ear is a natural result of the distance between the user and camera. Specifically, when a sufficient distance between the face and camera exists, the ears are captured because there is enough distance for the light rays from the ears to strike the camera's lens. *See Tanii*, [0048], Figs. 3A-3B.

Fig.3A

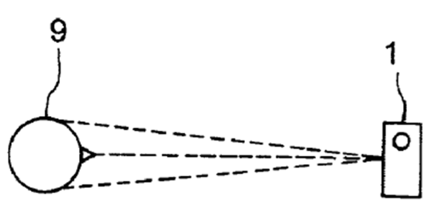
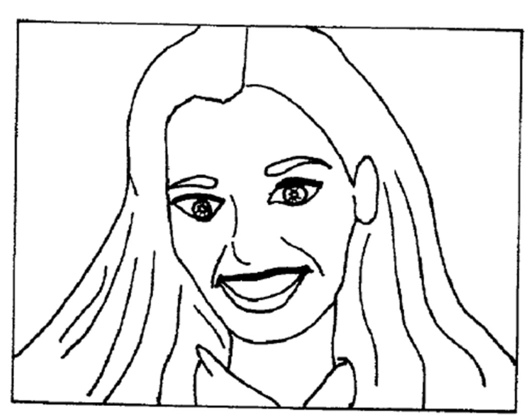


Fig.3B



But when a face is too close, the ears will not be captured because there is insufficient distance for the light rays from the ears to strike the camera's lens. *See Tanii*, [0048], Figs. 4A-4B.

Fig.4A

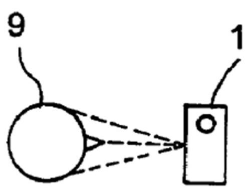
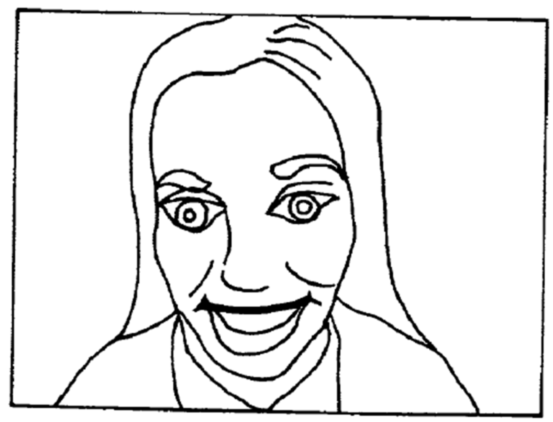
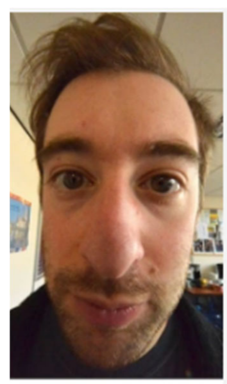


Fig.4B



This effect was well known and demonstrated in actual applications, as shown below.

85mm @ 200cm



8mm @ 20cm

445. In my opinion, therefore, a POSITA would have appreciated based on at least Tanii that the presence of a user's ears in one image at a sufficient distance, but absence of a user's ears in another image at a closer distance would be indicative of a three-dimensional face, and would have been motivated to modify Derakhshani

to verify the presence and absence of the ears between images as yet another indicator of the three-dimensional nature of the face.

5. Claim 4: The method according to claim 1, wherein the computing device is configured to display one or more prompts on a screen of the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

446. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 4's additional limitation.

447. Derakhshani discloses that the invention can be implemented in computing devices such as a "smart phone, a tablet computer, a television, a laptop computer, or a personal computer," *Derakhshani*, 5:22-27, which incorporate a camera, *id*., 5:23-27, 6:3-10, and a display. *Id*., 6:8-11, 9:22-24, 14:35-37, 22:33-38, 25:9-15. Derakhshani discloses displaying prompts to a user to guide the user to capture images of the user's face for authentication, *Derakhshani*, 5:23-32, 6:8-16, 9:22-26, including at more than once distance, *id.*, 17:64-66; §XI.A.2.d (1[c]).

448. But even if Derakhshani does not expressly disclose taking two images at different distances, doing so would have been obvious in view of Tanii to identify distance-induced distortions that indicate depth of a three-dimensional face. §XI.A.2.d (1[c]). When modifying Derakhshani to look for distance-induced distortions by capturing images at different distances consistent with Tanii, in my opinion, a POSITA would have been motivated to provide prompts to a user to

ensure the images are captured at the correct distances because Derakhshani already discloses providing prompts to correctly orient the user relative to the camera.

6. Claim 6: The method according to claim 4, wherein the computing device is a hand-held device, and the user holds the computing device at the first distance and the second distance to capture the at least one first image and the at least one second image.

449. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 6's additional limitation.

450. Derakhshani discloses that the biometric-authentication process can be implemented on a variety of different type of hand-held computing devices, such as "a laptop computer, a handheld computer…, a tablet computing device, a personal digital assistant (PDA), a cellular telephone…, a camera, a smart phone," and more. *See, e.g.*, *Derakhshani*, 8:11-28, 18:1-4. Derakhshani also recognizes that, to verify three-dimensionality of the face, "a single camera may be rotated or slide slightly," or that, when the device is hand-held, "the [camera] sensor may naturally move relative to the users face due to involuntary haptic motion" that may sufficiently capture a parallax effect. *Id.*, 17:59-18:4. Similarly, Tanii recognizes that distance-induced distortions often occur in mobile devices that have incorporated wide-angle lenses, and the amount of distortion is dictated by the distance between the user and the camera. *Tanii*, [0007], [0047]-[0048], Figs. 3A-B, 4A-B.

451. In my opinion, a POSITA would have also understood that, when performing a three-dimensional verification of the face on a mobile computing device, *see* §§XI.A.2.d (1[c]), XI.A.5 (claim 4), the user would adjust the distance by holding the mobile device and extending and retracting their arm (holding the mobile device at a first distance, then a second distance). A POSITA would have appreciated that mobile devices are routinely held to capture images, and holding the mobile device and adjusting distance would be a convenient and obvious way of changing the distance. In fact, Derakhshani already envisions evaluating depth based on the displacement of the user's arm while holding the device. §VII.A (Derakhshani); *Derakhshani*, 16:44-11, 17:45-18:4.

7. Claim 7: The method according to claim 6, wherein the computing device comprises a laptop or desktop computer and, with the computing device stationary, the user moves from the first distance to the second distance to capture the at least one first image and the at least one second image.

452. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 7's additional limitation.

453. Derakhshani discloses the computing device may be a "desktop computer [or] a laptop computer," (*Derakhshani*, 8:11-28), and that, in some instances, "a user is prompted to move in order to change the relative orientation of the [user] and [camera] sensor" during three-dimensional verification, (*Derakhshani*, 17:64-66).

454. When implementing a facial-authentication process on a more-stationary computing device (e.g., a desktop or laptop computer) consistent with Derakhshani, alone or in combination with Tanii, *see* §§XI.A.2.d (1[c]), XI.A.5 (claim 4), XI.A.6 (claim 6), in my opinion, a POSITA would have further understood that the user would physically move their face closer or further from the camera—as Derakhshani suggests—to capture images of the face at different distances, because that is a convenient and obvious way of changing the distance between a larger computing device and the user's face.

8. Claim 9: The method according to claim 1, wherein the first biometric data and the second biometric data are transmitted over a network to a server.

455. In my opinion, Derakhshani discloses or suggests claim 9's additional limitation.

456. Derakhshani discloses that the biometric-authentication process can be performed locally on the device, on a server, or split between the two. *See, e.g.*, *Derakhshani*, 9:27-58, 10:1-24. For instance, in some embodiments, "the authentication application 550 transmits captured image data to an authentication module (e.g., authentication modules 525 or 540) on a remote server (e.g., server systems 512 or 514) through the network 511." *Id.*, 9:27-31. Derakhshani also discloses that "processing of the image data for authentication purposes may be

performed by the authentication application…and the results…may be transmitted to an authentication module." *Id.*, 9:35-39.

457. In my opinion, a POSITA would have understood that what types of data to send to the server is a design choice that balances the processing capabilities of the image-capture device and the transmission bandwidths available, because Derakhshani discloses that "authentication functions may be distributed between the client and the server side processes in a manner suited [to] a particular application." *Id.*, 9:27-58, 10:1-24. If there is little bandwidth and ample processing power on the image capture device, a POSITA would be motivated to design the authentication on the device, so only a small amount of information needs to be transmitted to the server. If there is ample bandwidth available, but the image capture device is constrained in its processing power, a POSITA would be motivated to transmit raw image data to the server, so that the authentication could be done there. It would be obvious to a POSITA how to manage intermediate scenarios through standard engineering analysis.

458. In my opinion, a POSITA would have therefore found it obvious that, depending on the capabilities of the device used to capture the images, the device that captures the image may perform all, some, or none of the Derakhshani's processing steps. That means that, in some circumstances, a device will extract the biometric "landmarks" to limit the data transferred across a network (e.g., compared

to transmitting a full image), but offload the rest of the verification processing to a server, including transmitting the extracted biometric data over a network to the server for verification.

9. Independent Claims 10, 19

459. In my opinion, Derakhshani, alone or combined with Tanii, teaches the system claim 10 and method claim 19 for substantially the same reasons as method claim 1 because, other than limitation 10[pre]-[a] and 19[a]-[b], the claims are substantively identical as shown in the table below.

<table>
<tr><th>Claim 1</th><th>Claim 10</th><th>Claim 19</th><th>Reference</th></tr>
<tr><td>1[Pre]</td><td>10[Pre]</td><td>19[Pre]</td><td>§XI.A.2.a</td></tr>
<tr><td>-</td><td>10[a]</td><td>-</td><td></td></tr>
<tr><td>1[a]</td><td>10[b]</td><td rowspan="2">19[a]</td><td>§XI.A.2.b</td></tr>
<tr><td>1[b]</td><td>10[c]</td><td>§XI.A.2.c</td></tr>
<tr><td>1[c]</td><td>10[d]</td><td rowspan="2">19[b]</td><td>§XI.A.2.d</td></tr>
<tr><td>1[d]</td><td>10[e]</td><td>§XI.A.2.e</td></tr>
<tr><td>1[e]</td><td>10[f]</td><td>19[c]</td><td>§XI.A.2.f</td></tr>
<tr><td>1[f]</td><td>10[g]</td><td>19[d]</td><td>§XI.A.2.g</td></tr>
<tr><td>1[g]</td><td>10[h]</td><td>19[e]</td><td>§XI.A.2.h</td></tr>
<tr><td>1[h]</td><td>10[i]</td><td>19[f]</td><td>§XI.A.2.i</td></tr>
<tr><td>1[i]</td><td>10[j]</td><td>19[g]</td><td>§XI.A.2.j</td></tr>
</table>

a. 10[pre]: A system for verifying three-dimensionality of a user's face using images of the user's face captured using a camera equipped computing device, the system comprising:

460. If the preamble is limiting, in my opinion, Derakhshani discloses or suggests 10[pre]'s additional limitation.

461. Derakhshani discloses that the biometric-authentication process described is "embodied in a system," and provides a series of example systems where the process could be applied. *Derakhshani*, 1:20-21, 2:4-30, 5:22-27 (implementing invention in a "smart phone, a tablet computer, a television, a laptop computer, or a personal computer"); 5:65-8:28 (providing several example systems); 26:57-29:11 (system claims).

b. 10[a]: a computing device having a camera, screen, processor, and memory configured with non-transitory machine readable code that is executable by the processor, the machine readable code configured to:

462. If the preamble is limiting, in my opinion, Derakhshani discloses or suggests 10[pre]'s additional limitation.

463. Derakhshani discloses that the biometric-authentication process described is "embodied in a system," and provides a series of example systems where the process could be applied. *Derakhshani*, 1:20-21, 2:4-30, 5:22-27 (implementing invention in a "smart phone, a tablet computer, a television, a laptop computer, or a personal computer"); 5:65-8:28 (providing several example systems); 26:57-29:11 (system claims).

c. 19[a]: receiving first biometric data generated from at least one first image of the user taken with the camera of the computing device located at a first distance from the user;

464. In my opinion, Derakhshani discloses or suggests limitation 10[a].

465. Derakhshani discloses capturing and processing the first image taken at a first distance to obtain first biometric data. *See* §§XI.A.2.b (1[a]), XI.A.2.c (1[b]). In my opinion, a POSITA would have understood that limitation 19[a] presents a difference in perspective with respect to the system compared to claim 1. Specifically, where at least some of the limitations of claim 1 are from the perspective of the structure performing the method (e.g., a camera "captures," and a processor "processes"), claim 19 recites limitations directed to the transmission of certain data (e.g., a processor performing the processing *receives* the biometric data).

466. Because Derakhshani, alone or combined with Tanii, already describes the recited structures performing the recited functions, however, a POSITA would have known that Derakhshani also performs the necessary data transmissions as well for carrying those functions out. Specifically, a POSITA would have read Derakhshani as disclosing, or at least obviously suggesting that the processor performing the processing does, in fact, *receive* the biometric data, whether from its own processor (when all functions are carried out on a single device), or a different processor (when some functions are performed by a server), in order to carry out the disclosed functions.

d. 19[b]: receiving second biometric data generated from at least one second image of the user taken with the camera of the computing device located at a second distance from the user, the second distance being different than the first distance;

467. Derakhshani, alone or in combination with Tanii, teaches limitation 19[b].

468. Derakhshani, alone or in combination with Tanii, discloses capturing and processing the second image to obtain second biometric data at a distance different from the first distance. *See* §§XI.A.2.d, XI.A.2.e (1[c], 1[d]). In my opinion, a POSITA would have understood that limitation 19[b] presents a difference of perspective, but that Derakhshani, alone or combined with Tanii, teach these different perspectives as well. *See* §XI.A.9.c (19[a]).

10. Claim 11

a. 11[a]: The method according to claim 10, further comprising: interpolating the first biometric data and the second biometric data to obtain estimated intermediate biometric data;

469. In my opinion, Derakhshani discloses or suggests limitation 11[a] for the reasons discussed in §XI.A.3.a (2[a]).

b. 11[b]: capturing at least one third image of the user taken with the camera of the computing device at a third distance from the user, the third distance being between the first distance and the second distance;

470. In my opinion, Derakhshani discloses or suggests limitation 11[b] for the reasons discussed in §XI.A.3.b (2[b]).

c. 11[c]: processing the at least one third image to obtain third biometric data based on the at least one third image; and

471. In my opinion, Derakhshani discloses or suggests limitation 11[c] for the reasons discussed in §XI.A.3.c (2[c]).

d. 11[d]: comparing the estimated intermediate biometric data with the third biometric data to determine whether the third biometric data matches the estimated intermediate biometric data.

472. In my opinion, Derakhshani discloses or suggests limitation 11[d] for the reasons discussed in §XI.A.3.d (2[d]).

11. Claim 12: The method according to claim 10, further comprising verifying the presence of the user's ears in the at least one first image, and verifying the absence or reduced visibility of the user's ears in the at least one second image, wherein the first distance is larger than the second distance.

473. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 12's additional limitation for the reasons discussed in §XI.A.4 (claim 3).

12. Claim 13: The method according to claim 10, wherein the computing device is configured to display one or more prompts on a screen of the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

474. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 13's additional limitation for the reasons discussed in §XI.A.5 (claim 4).

13. Claim 14: The method according to claim 10, wherein comparing the first biometric data to the second biometric data and the determining that the user's face is three-dimensional occurs at a server that is remote from the camera equipped computing device.

475. In my opinion, Derakhshani discloses or suggests claim 14's additional limitation.

476. Derakhshani discloses that the biometric-authentication process can be performed locally on the device, on a server, or split between the two. *See, e.g.*, *Derakhshani*, 9:27-58, 10:1-24. In my opinion, a POSITA would have therefore understood that Derakhshani's three-dimensional verification—which is part of the biometric-authentication process—would obviously be configured in some circumstances to perform the verification on the server. *See* §XI.A.8 (claim 9) (discussing design choice of offloading certain processing).

14. Claim 15: The method according to claim 13, wherein the computing device is a hand-held device, and the user holds the computing device at the first distance and the second distance to capture the at least one first image and the at least one second image.

477. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 15's additional limitation for the reasons discussed in §XI.A.6 (claim 6).

15. Claim 16: The method according to claim 10, wherein the computing device comprises a laptop or desktop computer and, with the computing device stationary, the user moves from the first distance to the second distance.

478. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 16's additional limitation for the reasons discussed in §XI.A.7 (claim 7).

16. Claim 18: The method according to claim 10, wherein the first biometric data and the second biometric data are maintained on the computing device.

479. In my opinion, Derakhshani discloses or suggests claim 18's additional limitation.

480. Derakhshani discloses that the biometric-authentication process can be performed locally on the device, on a server, or split between the two. *See, e.g.*, *Derakhshani*, 9:27-58, 10:1-24. A POSITA would have understood that Derakhshani's three-dimensional verification—which is part of the biometric-authentication process—would be configured in some circumstances to perform locally on the device. In such instances, in my opinion, a POSITA would have understood that the biometric data would be maintained on the computing device so that all processes would be performed using a single device. If a single, local device is being used for all authentication procedures, there would be no need to store any data on a separate device. Moreover, a POSITA would have understood that biometric data is extremely sensitive, and transmitting that data over networks

presents a security risk of being intercepted. A POSITA would have understood that there would be no need to take any security risks by transmitting such data if the single, local device is handling the entire biometric-authentication procedure.

17. Claim 20: The method of claim 19, wherein the receiving of the first biometric data and the second biometric data occurs at a server and the first biometric data and the second biometric data are received over one or more of a LAN, WAN, or Internet type network.

481. In my opinion, Derakhshani discloses or suggests claim 20's additional limitation.

482. Derakhshani, alone or when combined with Tanii, teach receiving first and second biometric data. *See* §§XI.A.9.c (19[a]), XI.A.9.d (19[b]). Derakhshani further discloses that the biometric-authentication process can be performed locally on the device, on a server, or split between the two, *see, e.g.*, *Derakhshani*, 9:27-58, 10:1-24, and that when a server is used, devices used for biometric authentication can transmit authentication-related information over a network, including a LAN, WAN, or Internet-type network, *id*., 7:38-8:3.

483. In my opinion, a POSITA would have therefore understood that, when utilizing a server to perform the biometric authentication, first and second biometric data would be received by the server over one of those types of networks. In fact, the use of LAN, WAN, and Internet-type networks were some of the most well-known ways to transmit data between different computing devices.

B. Ground 1B: Derakhshani, Tanii, and Tahk (Claims 5, 8, 17)

1. Motivation to Combine

484. In my opinion, a POSITA would have been motivated to modify Derakhshani, with or without Tanii, in view of Tahk because Tahk provides a user-friendly way of ensuring that a face presented for facial authentication is properly framed. In my opinion, a POSITA would have understood that providing a real-time preview of what an image would look like prior to capturing the image, as well as providing express prompts—such as written instructions or oval shapes on the live-preview screen—as taught by Tahk, §VII.D (Tahk), would have been particularly useful for authentication procedures that require capturing multiple images of a face, as taught by Derakhshani and Tanii.

485. For instance, providing a user a real-time preview would allow the user to actively adjust the position of the camera and/or their orientation of their face to properly frame their face for the image capture. Moreover, providing express prompts and oval shapes sized to guide the user to properly position their face would ensure images best suited for facial recognition can be captured, and that the faces would be captured from sufficiently different perspectives and/or distances to ensure the three-dimensional verification taught by Derakhshani, alone or in combination with Tanii, could be performed. This is particularly important for facial-recognition

systems, which generally are known to have difficulty matching faces across different views. *See, e.g.*, Ex-1018, 32.

486. Moreover, a POSITA would have known that providing real-time image feedback, written instructions, and oval shapes to frame a face during a facial authentication process were all well-known and conventional techniques to provide user feedback during image capture as of the time of the invention. *See, e.g.*, Ex-1034, 7:16-8:7, Figs. 6B-7C (providing "an example of an interface used upon registering a facial image of a person to be authenticated" in which "the image of this person is displayed on a monitor" during registration, and oval-shaped prompts to indicate the size/distance of the user from the face authentication sensor); Ex-1035, 5:31-32 ("The computing device may present prompts that instruct the user to perform one or more liveness gestures"), 6:3-4 (same).

2. Claim 5: The method according to claim 4, wherein the one or more prompts are ovals sized on the screen within which the face of the user is placed to capture the at least one first image and the at least one second image at the first and second distances.

487. In my opinion, Derakhshani combined with Tanii and Tahk teaches claim 5.

488. As I have previously explained, Derakhshani, alone or in combination with Tanii, teaches providing prompts to user to properly frame themselves at different distances to capture images for biometric authentication. §XI.A.5 (claim

4). But Derakhshani and Tanii do not expressly describe using oval-shaped prompts to guide a user during the facial-authentication process.

489. In my opinion, however, a POSITA would have been motivated to provide such oval-shaped prompts (as well as express written instructions) in view of Tahk. *See, e.g.*, *Tahk*, Figs. 8A-B ("Please step further back" and "Please step further forward," and presenting an oval to frame the face at the correct distance). A POSITA would have been motivated to modify Derakhshani, alone or in combination with Tanii, to provide such oval-shaped prompts because they are a natural shape to appropriately size and frame a face at different distances. *See* §VII.D (Takh); XI.B.1 (motivation).

3. Claim 8: The method according to claim 1, further comprising displaying an image on a screen of the computing device while capturing the at least one first and/or the at least one second image.

490. In my opinion, Derakhshani combined with Tanii and Tahk teaches claim 8.

491. Derakhshani and Tanii do not expressly disclose displaying an image on a screen of the computing device while capturing the images of the user's face. Tahk, however, teaches displaying "a preview image for the face image." *See, e.g.*, *Tahk*, [0118], [0129], [0135], [0139], [0143], [0144], Fig. 8A-B.

492. In my opinion, a POSITA would have been motivated to modify the facial-authentication process taught by Derakhshani, alone or in combination with

Tanii, to provide a live "preview" of the image to be captured so that the user could appropriately frame their face prior to capturing the image. *See* §XI.B.1 (motivation).

493. In fact, providing a preview of the image the camera is intended to capture has been known for as long as cameras have existed. For instance, many camera designs have incorporated viewfinders to provide the user a preview of the image. Ex-1032. And when devices such as feature phones and smartphones began incorporating cameras, they also included previews of the image intended to be captured on their displays. Ex-1033. In other words, displaying the view of the camera as the image is intended to be captured would be extremely well-known to a POSITA.

4. Claim 17: The method according to claim 10, further comprising displaying an image on a screen of the computing device while capturing the at least one first image and the at least one second image.

494. In my opinion, Derakhshani combined with Tanii and Tahk teaches claim 17 for the reasons discussed in §XI.B.3 (claim 8).

C. Ground 2A: Zhang and Tanii (Claims 1-3, 9-12, 14, 16, 18-20)

1. Motivation to Combine

495. In my opinion, a POSITA would have been motivated to combine Zhang and Tanii because both concern identifying and accounting for the three-dimensional nature of a face when capturing an image. They differ, however, in what

principles are used to account for the face's three-dimensionality. Zhang, for instance, looks to dissimilarities in two images after one undergoes a mathematical homography. *See* §VII.C (Zhang). And although Tanii is not expressly directed to *evaluating* whether a face has depth like Zhang, Tanii exemplifies the well-known distortions caused by the interaction between the camera's lens and the three-dimensional nature of the face, *see* §VII.B (Tanii). A POSITA would have appreciated, therefore, that Tanii recognizes another clear alternative to evaluating the depth of a face, consistent with Zhang's existing homography transformation.

496. A POSITA would have recognized, as Tanii does, that distance-induced distortions occur because of the interactions between the shape of the camera lens and shape of the face, and the distortion in part depends on the distance between the face and the camera. §VII.B (Tanii); *Tanii*, [0048]. Accordingly, a POSITA would have understood from Tanii that, by taking two images from two different distances, a larger amount of distortion in the closer of the two images indicates whether a face is three-dimensional or not.

497. In my opinion, a POSITA would have therefore appreciated from Tanii that images captured by Zhang—without any modification—may exhibit distance-induced distortions based on the particular camera used to perform Zhang's process (e.g., particularly when a wide-angle lens with significant barrel distortion is used, as is common in computers and mobile devices). However, a POSITA would have

also appreciated that any distance-induced distortions would further enhance Zhang's homography-transformation process because a homography transformation cannot correct for these distortions.

498. For instance, if a homography transformation were applied to Tanii's Figure 4B (serving as Zhang's "first image") to compare to Figure 3B (serving as Zhang's "second image"), the transformation would not account for differences between the images caused by the distance-induced distortion.

Fig.4B

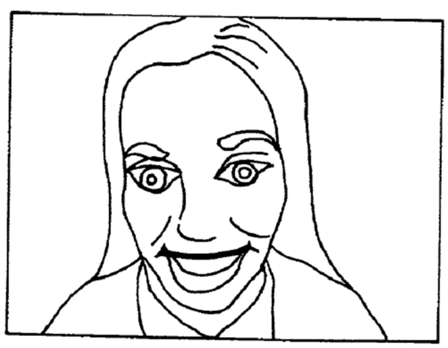
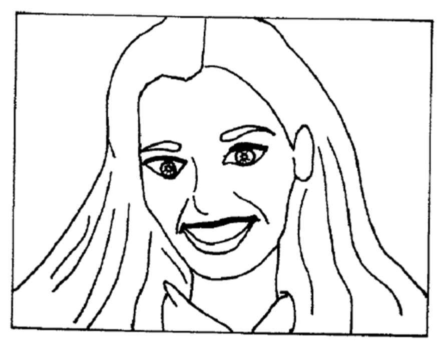


Fig.3B



Tanii, Figs. 3B, 4B. That is because Zhang relies on a mathematical principle that enables transforming the *perspective* of a planar object, such as a photograph being used to spoof the authentication procedure to a different *perspective*, §VII.C (Zhang), whereas the distortion identified by Tanii is *radial* and a byproduct of the lens' imperfections and the change in magnification with distance. A homography transform does not account for such radial distortions, but would instead transform the perspective of Tanii with its distortions intact. In other words, in a transformation of perspective with a three-dimensional object such as a real face, Tanii's distance-induced distortions would remain. Ultimately, however, when comparing the two images once one is transformed into the perspective of the other, there would remain differences attributable to the distance-induced distortion which, in my opinion, a POSITA would have understood would result in Zhang identifying the face as three-dimensional.

499. In my opinion, a POSITA would have therefore recognized that Zhang's existing process would be *enhanced* by prompting a user to capture two images and two distances—one of which would have increased distance-induced distortion—because if the face were three-dimensional, Zhang's existing procedure would identify the two images as different and indicate a three-dimensional face. The lack of a match between the two images would likely be enhanced by changes in radial distortion: it makes them even less like data from two planar objects which would produce a match.

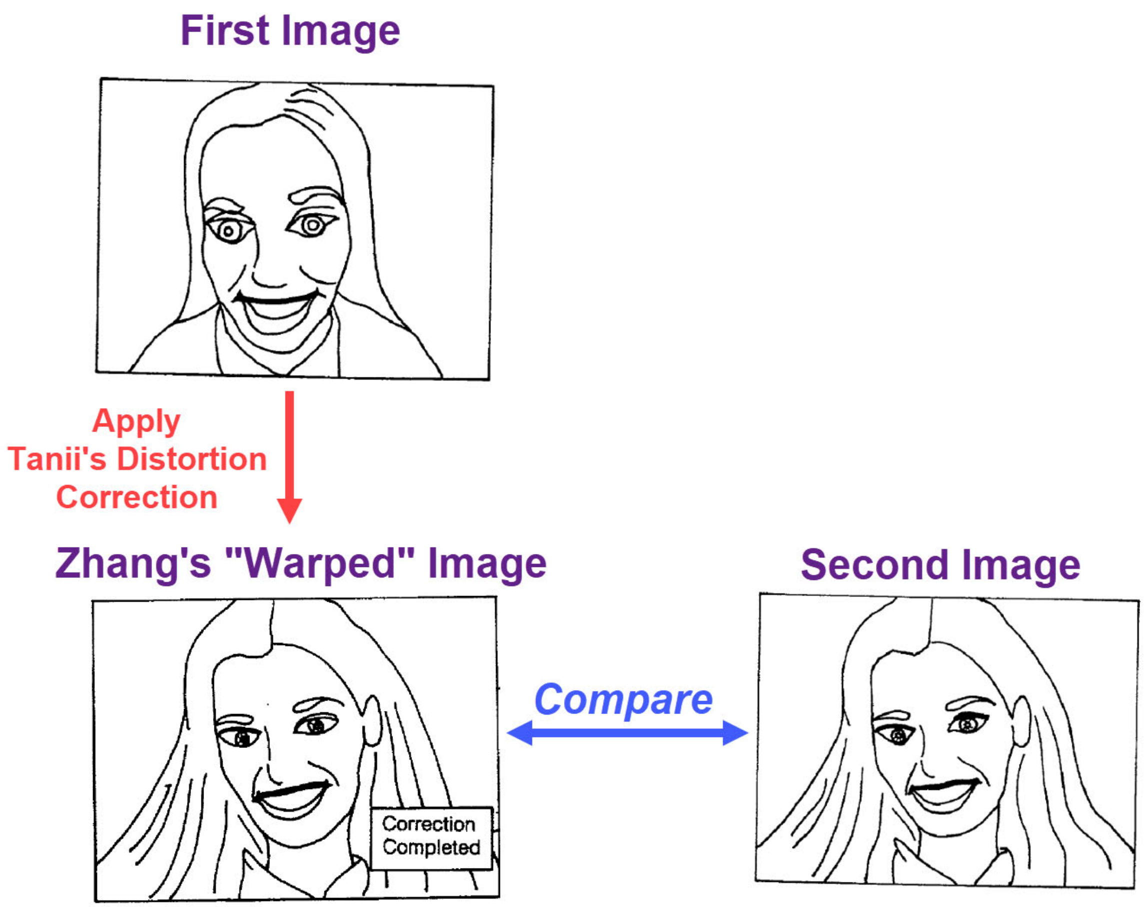
500. However, in my opinion, a POSITA would have also been motivated to modify Zhang's process in view of Tanii in either of two additional ways.

501. First, in my opinion, a POSITA would have been motivated to follow Zhang to verify the three dimensionality of a face during a facial authentication procedure by taking two or more images, applying mathematics to one of the images, and comparing the mathematically altered image to a second (unaltered) image. But

instead of the mathematics applied being a homography transformation, in my opinion, a POSITA would have been motivated to *substitute* Zhang's mathematics for those taught by Tanii to correct for distance-induced distortion. In other words, rather than change the perspective of one image to match the second image, a POSITA would correct the distortion of one image (to create what Zhang refers to as its "warped" image[5]) and compare the result to another image taken further away

[5] Zhang and Tanii both use the term "warped" to refer to different effects, but they are not inconsistent with one another. Specifically, Zhang uses the term "warped" to refer to the resulting image that has undergone homography transformation because the original relationship between the pixels in the image are modified. Tanii uses the term "warped" to refer to the distortions in an image of a face induced by the image-capture conditions (e.g., distance and lens geometry). When I refer to Zhang's "warping," I am referring to the result of a mathematical application to an image; and when I refer to Tanii's warping, I am referring to distance-induced distortion.

that does not exhibit the same degree of distance-induced distortions.

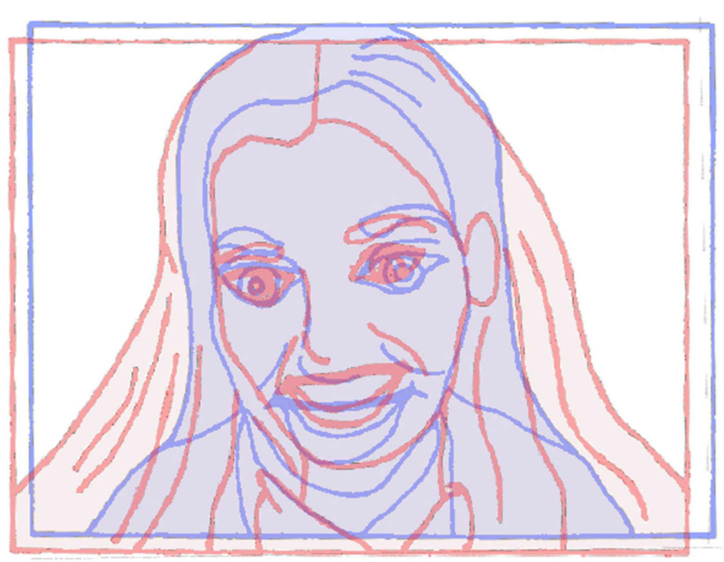


Tanii, Figs. 3B, 4B, 9.

502. A POSITA would have appreciated that if the "warped" (distortion-corrected) image and second image are sufficiently similar, that indicates a three-dimensional face because Tanii is correcting for distortions attributable the three-dimensionality of the user's face. By following this approach, a POSITA would have recognized that the only difference (besides the mathematics) is that the comparison

between the Zhang-Tanii "warped" (distortion-corrected) image would look for a match with the second image.

503. Alternatively, a POSITA would have appreciated that Zhang and Tanii could be modified to eliminate the mathematical transformation of a first image entirely. Once again, a POSITA would have been motivated to follow Zhang to verify the three dimensionality of a face during a facial authentication procedure by taking two or more images, but rather than apply mathematics to "warp" one of the images (e.g., using either a homography transform or distortion-correction procedure), the facial features would be mapped in each image, matched between the two images, and evaluated to determine whether differences attributable to distance-induced distortion appear (e.g., does the shape of the nose, size of the mouth or forehead, or do facial features shift by expected degrees relative to one another?). For instance, I have overlayed Tanii's two images to show how one (in blue) exhibits expected distortions while the other (in red) does not, resulting in various misalignments in facial features (assuming the faces are normalized in size):



In such circumstances, a POSITA would understand that two images would still be required, rather than just evaluating one image for distance-induced distortion. Otherwise, an imposter could provide a picture of a user with distance-induced distortion already applied to spoof the system; the need for a more-distance, undistorted image of the user for comparison would still be required.

504. In my opinion, a POSITA would have been motivated to make either of these two modifications for two reasons. First, a POSITA would have appreciated that Zhang's homography-transformation process may be spoofed by presenting a non-planar picture of a face, because Zhang's homography transformation process specifically looks for a *planar* structure in images that can be transformed nearly identically from one perspective to another. Thus, Zhang's system could possibly be subverted by bending the picture in a way to trick the system, or applying the picture to a three-dimensional shape. For this reason, in my opinion, a POSITA would have

therefore been motivated to look for other methods to ensure the user's face is from the user, and not a spoofer. A POSITA would have also appreciated that distance-induced distortion is more difficult to spoof, because it is induced by the interactions of geometries between the user's face and the camera's lens, and therefore could not be circumvented as easily. Second, a POSITA would have appreciated that either of the processes suggested by Tanii offers a potentially less computationally demanding than the homography mathematics proposed by Zhang, which may be more suitable for a low-power portable device.

505. In my opinion, a POSITA would have had a reasonable expectation of success in making this modification to Zhang because Tanii already taught a mechanism to identify (and correct) distance-induced distortions, *see, e.g.*, *Tanii*, [0056], and it was already well-known to use depth information about a face derived from a series of images to distinguish between live faces and two-dimensional images of faces. *See, e.g.*, Ex-1014, Abstract, [0031], [0036].

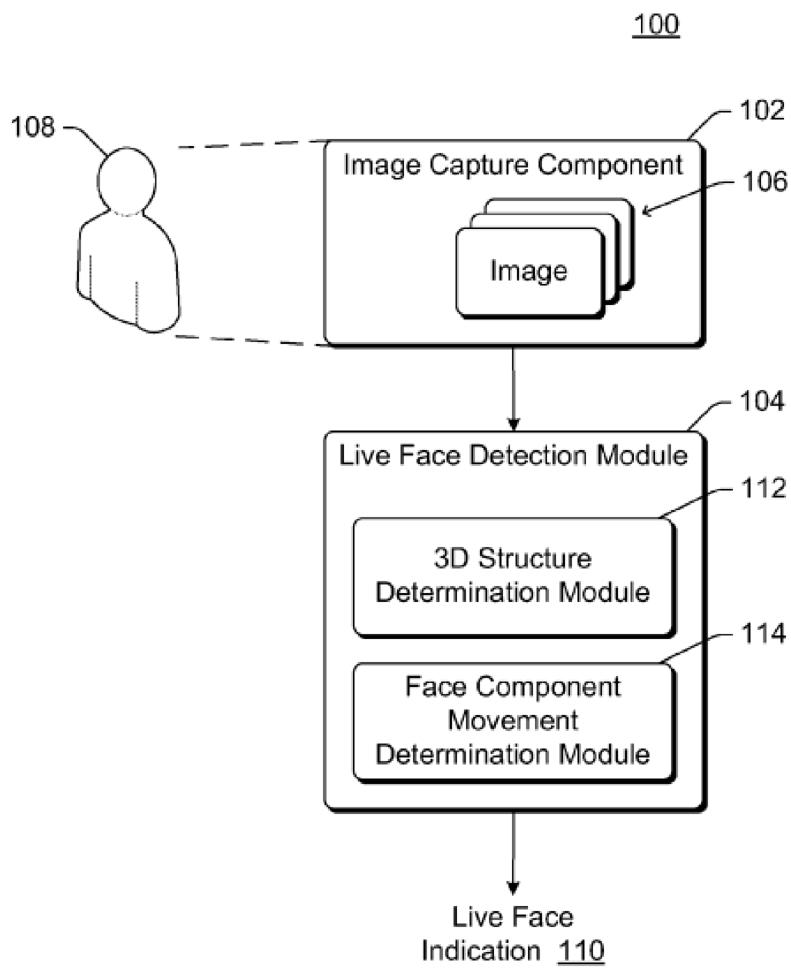
2. Independent Claim 1

a. 1[pre]: A method for verifying three-dimensionality of a user's face using images of the user's face captured using a camera equipped computing device, the method comprising:

506. If the preamble is limiting, in my opinion, Zhang discloses or suggests a method for verifying three-dimensionality of a user's face using images of the user's face captured using a camera equipped computing device.

507. Specifically, Zhang discloses a method "to determine whether a face in multiple images is a 3D structure or a flat surface," *Zhang*, [0026], Figs 2-3; *see also, e.g.*, *id.*, Abstract, [0003], to "authenticate a user for particular access," *id.* [0012]. To accomplish this, Zhang captures and analyzes multiple images of a user's face using the image capture component 102 implemented in a computing device (e.g., "a desktop computer, a laptop or notebook computer…[or] a cellular or other wireless phone"). *Zhang*, [0012]-[0013], [0016].

508. In my opinion, a POSITA would have understood that the "image capture component 102" would be a camera, because cameras are conventionally used to capture images, especially in computing devices. In fact, the "CCDs" and "CMOS" sensors Zhang references are the types of sensors commonly used in cameras. *Zhang*, [0016]; *see also, e.g.*, *Suzuki*, [0019] ("The camera unit includes solid-state image pickup elements such as CCD or CMOS"); Ex-1028, 3 ("Presently, there are two main technologies that can be used for the image sensor ***in a camera***, i.e., CCD (Charge-coupled Device) and CMOS (Complementary Metal-oxide Semiconductor)."



Zhang, Fig. 1.

b. 1[a]: capturing at least one first image of the user taken with the camera of the computing device at a first distance from the user;

509. In my opinion, Zhang discloses or suggests limitation 1[a].

510. Zhang discloses capturing a first image of a user as part of the authentication method. *Zhang*, [0016] ("user 108 presents himself or herself to image capture component 102, allowing component 102 to capture images 106 of user 108."), [0021].

511. In my opinion, a POSITA would have understood that Zhang's process captures an image at a first distance between the user and image capture component 102 in order to capture a picture of the user's face. *Zhang*, [0016]. Specifically, a POSITA would have understood that there *must* be some distance between the camera and the face to capture the first image so that the camera's field of view encompasses the face. If there were no distance between the camera and the face (e.g., if the camera were pressed up against the user's skin), then the field of view would be limited to just that patch of skin and the face would not be captured in the image, which would be useless for the three-dimensional verification process.

c. 1[b]: processing the at least one first image to obtain first biometric data from the at least one first image;

512. In my opinion, Zhang discloses or suggests limitation 1[b].

513. Zhang discloses processing the first image to extract "feature points" from the image. *Zhang*, [0027] ("[O]ne or more feature points are extracted from two images… A variety of different feature points can be extracted, such as a corner of an eye, a corner of a mouth, a tip of a nose, and so forth."), [0026] (disclosing "software, firmware, hardwire, or combin[ed]" implementations).

514. In my opinion, a POSITA would have understood that Zhang's extracted feature points constitute "biometric data" because "biometric data" generally refers to unique physical characteristics of an individual, which would include the positions of "feature points" such as a user's eyes, nose, mouth, and other

such features. *See, e.g.*, Ex-1018, 2 ("biometric recognition can be defined as the science of establishing the identity of an individual based on the physical and/or behavioral characteristics of the person."), 100-103 (describing the types of biometric data about a face used for facial-recognition systems, "such as the structure of the face components (e.g., eyes), [and] the relationship between facial components").

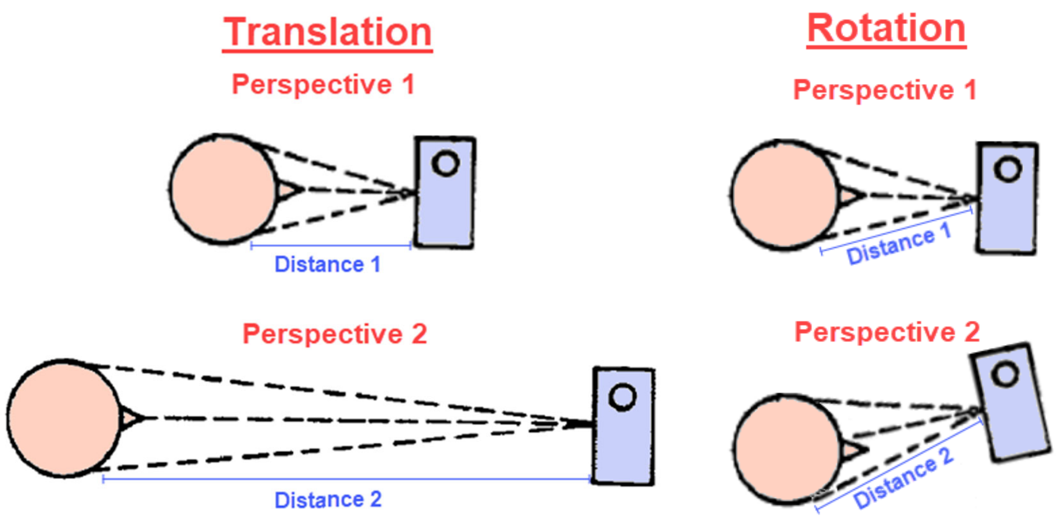
d. 1[c]: capturing at least one second image of the user taken with the camera of the computing device at a second distance from the user, the second distance being different than the first distance;

515. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 1[c].

516. Zhang discloses capturing a second image of a user as part of the authentication method. *Zhang*, [0016] ("Image capture component 102 captures multiple images").

517. Zhang does not expressly disclose that the second image is captured at a second distance different from the first distance of the first image. But, in my opinion, a POSITA would have understood that Zhang at least implicitly requires *some* change of distance. §XI.C.1 (motivation). Specifically, Zhang discloses a "3D structure determination module 112" that uses a "homography" technique to distinguish between a real face and a picture of a face by, *inter alia*, transforming a first image to the perspective of a second image and comparing the two. *Zhang*,

[0024], [0026]-[0035]; §VII.C (Zhang). In my opinion, a POSITA would have understood from Zhang that—like Derakhshani's parallax approach—the distances between the camera and at least some facial landmarks would change in order to obtain an image from a different perspective than the first and would obviously also encompass changing the overall distance between the camera and face as well. *See, e.g.*, §XI.A.2.d (in the context of Derakhshani, discussing changes of distance for parallax).



Moreover, a POSITA would have not only understood that providing images at different distances allows for a greater understanding of depth between objects in the scene, as exemplified in the paper Zhang references; Ex-1013, 22-25, but that taking pictures at different distances may induce distance-based distortion that would *enhance* the accuracy of Zhang's homography transformation to detect a three-dimensional face. §XI.C.1 (motivation).

518. Even if Zhang cannot be considered to disclose or suggest taking two images at different distances, however, a POSITA would have been motivated to do so in view of other prior art. For instance, a POSITA would have understood that distortions caused by camera lenses can indicate depth in the object being captured, as exemplified by Tanii. §XI.C.1 (motivation). Thus, even if Zhang does not already disclose this limitation, a POSITA would have been motivated to modify Zhang in view of Tanii to capture a second image at a second distance and evaluating the images for different degrees of distance-induced distortions to distinguish between live, three-dimensional faces and two-dimensional pictures of a face. §XI.C.1 (motivation).

e. 1[d]: processing the at least one second image to obtain second biometric data based on the at least one second image;

519. In my opinion, Zhang discloses or suggests limitation 1[d].

520. Zhang discloses processing the second image to obtain second feature-point biometric data from the image. *Zhang*, [0026]-[0027]; §XI.C.2.c (1[b]).

f. 1[e]: comparing the first biometric data with the second biometric data to determine whether the first biometric data matches the second biometric data;

521. In my opinion, Zhang discloses or suggests limitation 1[e].

522. Zhang discloses that "[t]he feature points extracted…are matched across the first and second images (act 304)" and, in my opinion, those feature points

constitute biometric data. *Zhang*, [0028]; §XI.C.2.c (1[b]). Zhang also discloses that the matching process may also "determine[] whether the first and second images include the same face," including "during the matching of feature points in act 304, if all (or at least a threshold number) of the feature points cannot be matched then it is determined that the first and second images are of different faces." *Zhang*, [0038].

g. 1[f]: comparing the first biometric data to second biometric data to determine whether differences between the at least one first image and the at least one second image match expected differences resulting from movement of the camera or the user which changed the distance between the user and camera from the first distance to the second distance;

523. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 1[f].

524. Zhang discloses that, after calculating a homography matrix between the first and second image, a "warped" version of the first image is created and then compared to the second image to determine whether expected differences exist. *Zhang*, [0025], [0031]. Zhang also discloses that, as part of the comparison, "any of a variety of conventional face detection algorithms or face recognition algorithms can be used to detect the face within each image, and the selected locations are the locations that are part of a face within at least one of the warped and second images." *Zhang*, [0032].

525. In my opinion, a POSITA would have understood that Zhang discloses comparing a first biometric data (e.g., the facial-feature locations in the first warped image) and second biometric data (e.g., the facial-feature locations in the second image) to determine whether differences between the two exist, in which it would be expected that a live face would have sufficient differences between the two images due to movement of the image capture component 102 (camera).

526. However, a POSITA would have also been aware that differences between two images—one with more distance-induced distortions and one with less—can also be used to distinguish between live, three-dimensional faces, and two-dimensional pictures of a face, as exemplified by Tanii. §XI.C.1 (motivation). And, in my opinion, a POSITA would have been motivated to look for these expected distortions as either a supplemental or alternative verification of three-dimensionality of a face. *Id*. A POSITA would have appreciated that verifying the three-dimensional nature of the face using distance-induced distortion would be accomplished by matching the positions of biometric features across the first and second images, consistent with Zhang. When modifying Zhang to look for such distortions as an indication of three-dimensionality rather than perform a homography transformation, *see* §XI.C.1, the comparison would evaluate whether one of the images exhibits the distance-induced distortion that would be expected when the user's face is captured at a close distance to the camera, and the other image

has less of the distance-induced distortion when captured further from the camera. *Id*.

h. 1[g]: determining that the user's face is three-dimensional when:

527. In my opinion, Zhang discloses or suggests limitation 1[g].

528. As explained in further detail below, Zhang determines whether the user's face is three-dimensional or not based on certain specified conditions. *Zhang*, [0025], [0034].

i. 1[h]: the first biometric data does not match the second biometric data; and

529. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 1[h].

530. Zhang discloses that "[t]he feature points extracted…are matched across the first and second images (act 304)" and, in my opinion, those feature points constitute data, and specifically biometric data. *Zhang*, [0028]; §XI.C.2.e (1[d]). Zhang also discloses that the matching process may also "determine[] whether the first and second images include the same face," including "during the matching of feature points in act 304, if all (or at least a threshold number) of the feature points cannot be matched then it is determined that the first and second images are of different faces." *Zhang*, [0038]. Zhang discloses that captured images are determined to be of a live, three-dimensional face when differences in the image data

exist after undergoing a homography transformation, *Zhang*, [0031], including when first biometric data (the position of facial features in the first "warped" image) does not match the second biometric data, *Zhang*, [0032]-[0034].

531. But even if Zhang does not disclose this limitation, in my opinion, Zhang combined with Tanii teaches it. *See* §§XI.C.2.d (1[c]), XI.C.2.g (1[f]). Specifically, a POSITA would have appreciated that, when modifying Zhang to evaluate differences arising from distance-induced distortions, a three-dimensional face would be indicated when one of the two sets of biometric data exhibits more distance-induced distortion (e.g., the first and second biometric data do not match). *See* §§XI.C.2.d (1[c]), XI.C.2.g (1[f]).

j. 1[i]: the second biometric data has the expected differences as compared to the first biometric data resulting from the change in distance between the user and the camera when capturing the at least one first image and the at least one second image.

532. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 1[i].

533. Zhang discloses that captured images are determined to be of a live, three-dimensional face when differences in the image data exist after undergoing a homography transformation, *Zhang*, [0031], including when first biometric data (the position of facial features in the first "warped" image) does not match the second biometric data, *Zhang*, [0032]-[0034]. These differences would be expected due to

a change in perspective (rotation and/or distance) of the camera between the two images. *See* §§XI.C.2.d (1[c]), XI.C.2.g (1[f]).

534. But even if Zhang does not disclose this limitation, Zhang combined with Tanii teaches it. *See* §§XI.C.2.d (1[c]), XI.C.2.g (1[f]). Specifically, a POSITA would have appreciated that, when modifying Zhang to evaluate differences arising from distance-induced distortions, a three-dimensional face would be indicated when one of the two sets of biometric data exhibits expected distance-induced distortions due to the change in distance of the camera. *See* §§XI.C.2.d (1[c]), XI.C.2.g (1[f]).

3. Claim 2

a. 2[a]: The method according to claim 1, further comprising: interpolating the first biometric data and the second biometric data to obtain estimated intermediate biometric data;

535. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 2[a].

536. Zhang discloses that, as part of the authentication process, the two images being compared may be "non-adjacent." *Zhang*, [0036]. Zhang explains that images are "non-adjacent" when additional images exist between the two images being compared for authentication. *Id.* In such instances, Zhang discloses performing some of the processes, such as "feature point extraction and feature point matching" using the intermediate images to "facilitate the feature matching process

when matching features across two images with one or more intervening images." *Id.* Zhang also discloses that the homography-transformation process can be applied to multiple pairs of images, whether the images are adjacent or non-adjacent. *Zhang*, [0037].

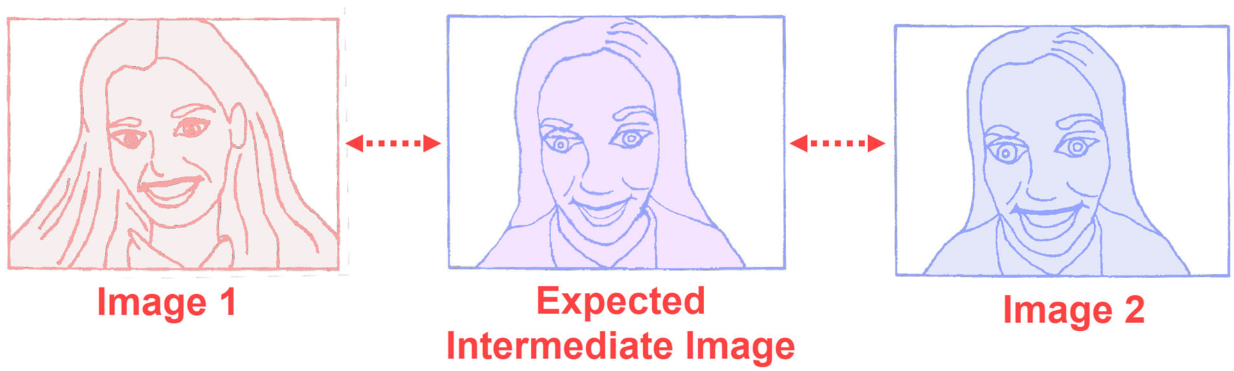
537. When a set of intermediate images exist *between* the first and second images, as Zhang discloses, in my opinion, a POSITA would have been motivated to generate predictions (i.e., interpolations) of what those intermediate images should look like based on Zhang's first and second images because using static images to build models or predictions of the face as a means of identifying a user was well-known in the art. Ex-2015, Abstract; *Derakhshani*, 17:27-44 (interpolating two-dimensional and three-dimensional models for comparison to acquired biometric data); Ex-1036, 8:19-27 (describing capturing one or more biometric features and calculating "change parameters" to evaluate whether the changes match expectations, or predictions of what the biometric features should look like). And a POSITA would have understood that building models or predictions of what Zhang's intermediate images *should* look like would further ensure against spoofing because a spoofer could not rely on artificial differences between the first and second images to have Zhang's system authenticate a face; the differences would also have to match what is expected *between* the two images.

538. Based on a POSITA's understanding of Zhang, a POSITA would have further been motivated to derive interpolated biometric data based on the combination of Zhang and Tanii. Specifically, Tanii discloses that distance-induced distortions increase as distance between the face and camera decreases. *See Tanii*, [0048]. A POSITA would have therefore understood that, all else being equal, distance-induced distortion depends on the distance between the user and the camera, and thus any set of images as the distance between the user and camera changes will have different degrees of facial warping, similar to the gradual changes in facial distortion that appears in the series of images below (although these images also altered the focal length of the camera to ensure the face remains a constant size in the frame, rather than just distance:

Ex-1022.

539. In my opinion, a POSITA reading Zhang—which discloses processing, interpolating, and evaluating intermediate images—in view of Tanii therefore would have been motivated to interpolate intermediate biometric data with an intermediate,

interpolated amount of distance-induced distortion based on the two non-adjacent images to create an array of intermediate distance-induced distortions that would indicate depth in a three-dimensional face, such a gradual projection from the lines of the facial landmarks shown in blue (the distorted image) to the lines of the facial landmarks shown in red (the distorted image).



Performing this type of interpolation between the distorted and undistorted images would allow for further comparison with additional images, for instance, to ensure the distance-induced distortion matches expectations if the face were truly three-dimensional, as depicted below:

b. 2[b]: capturing at least one third image of the user taken with the camera of the computing device at a third distance from the user, the third distance being between the first distance and the second distance;

540. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 2[b].

541. Zhang discloses capturing a series of intermediate images between two non-adjacent images. *Zhang*, [0035]-[0037]; *see* §XI.C.3.a (2[a]). A POSITA would have understood that these intermediate images would provide images at different positions (e.g., rotation or translation) of the camera relative to the first and second images. *See* §XI.C.3.a (2[a]).

542. When modifying Zhang in view of Tanii to interpolate intermediate biometric data attributable to distance-induced distortions, a POSITA would have been further motivated to capture a third image at a distance that correlates to one of the interpolated data sets for further authentication of three-dimensional depth of the face in the captured images. *See* §XI.C.3.a (2[a]).

c. 2[c]: processing the at least one third image to obtain third biometric data based on the at least one third image; and

543. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 2[c].

544. Zhang discloses processing a third (intermediate) image to obtain third biometric data. *Zhang*, [0036] ("the feature point extraction and feature point

matching in acts 302 and 304 can be generated for each adjacent pair of images in the sequence, which can facilitate the feature matching process when matching features across two images with one or more intervening images.").

545. Moreover, as discussed previously, when modifying Zhang in view of Tanii, a POSITA would have found it obvious to acquire a third image and extract biometric data from the third image to compare it to the interpolated positions of the biometric data based on the first and second images. *See* §§XI.C.3.a (2[a]), XI.C.3.b (2[b]).

d. 2[d]: comparing the estimated intermediate biometric data with the third biometric data to determine whether the third biometric data matches the estimated intermediate biometric data.

546. In my opinion, Zhang, alone or in combination with Tanii teaches limitation 2[d].

547. Zhang discloses tracking and comparing biometric features between the non-adjacent and intermediate images. *See Zhang*, [0036]-[0037]; §XI.C.3.a (2[a]).

548. In my opinion, a POSITA would have understood that, when interpolating what the intermediate images *should* look like based on the first and second images, a POSITA would have understood that the estimated, interpolated biometric data would be compared to the intermediate images to determine whether the intermediate images match what was predicted. *See* §XI.C.3.b (2[b]).

549. Furthermore, a POSITA would have been motivated to acquire a third image and extract biometric data from the third image to compare it to the interpolated positions of the biometric data based on the first and second images to determine if there is a match between the two. *See* §§XI.C.3.a (2[a]), XI.C.3.b (2[b]).

4. Claim 3: The method according to claim 1, further comprising verifying the presence of the user's ears in the at least one first image, and verifying the absence or reduced visibility of the user's ears in the at least one second image, wherein the first distance is larger than the second distance.

550. In my opinion, Zhang combined with Tanii teaches claim 3.

551. Zhang does not expressly disclose a process of verifying the presence in one image and absence in another of a user's ears.

552. However, when modifying Zhang in view of Tanii to use distance-induced distortions to verify the three-dimensional nature of a face, Tanii teaches that the absence and presence of an ear is a natural result of the distance between the user and camera. Specifically, when a sufficient distance between the face and camera exists, the ears are captured because there is enough distance for the light rays from the ears to strike the camera's lens. *See Tanii*, [0048], Figs. 3A-3B.

Fig.3A

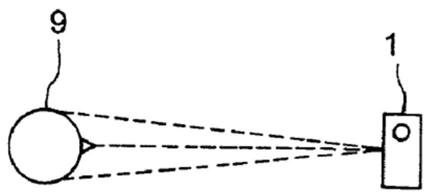
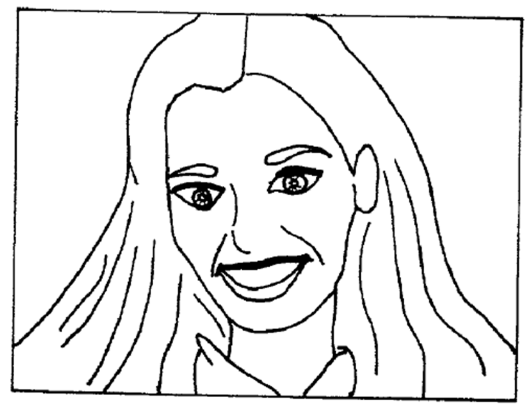


Fig.3B



But when a face is too close, the ears will not be captured because there is insufficient distance for the light rays from the ears to strike the camera's lens. *See Tanii*, [0048], Figs. 4A-4B.

Fig.4A

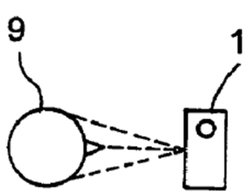
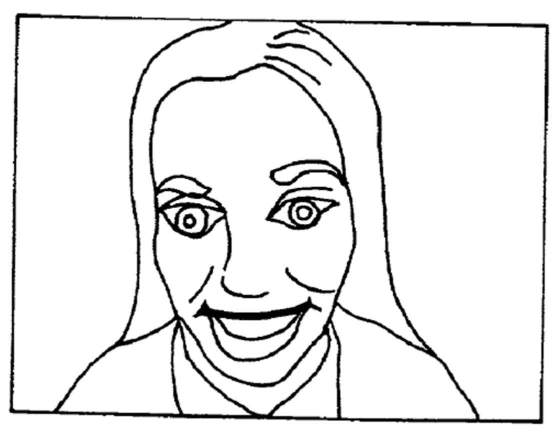
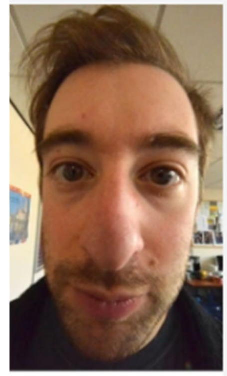


Fig.4B



This effect was well-known and demonstrated in actual applications, as shown below.

85mm @ 200cm



8mm @ 20cm

In my opinion, therefore, a POSITA would have appreciated based on at least Tanii that the presence of a user's ears in one image at a sufficient distance, but absence of a user's ears in another image at a closer distance would be indicative of a three-dimensional face, and would have been motivated to modify Zhang to verify the presence and absence of the ears between images as yet another indicator of the three-dimensional nature of the face.

5. Claim 9: The method according to claim 1, wherein the first biometric data and the second biometric data are transmitted over a network to a server.

553. In my opinion, Zhang discloses or suggests limitation claim 9.

554. Zhang discloses that the image capture component and live face detection module (104) with accompanying 3D structure determination module (112) can communicate and send data, including biometric facial feature data, over a variety of different networks, such as the Internet, a local area network (LAN), an intranet, etc. *Zhang*, [0014].

555. Although Zhang does not expressly disclose that the data is sent to a "server," in my opinion, a POSITA would have understood based on Zhang's disclosure that the use of a server would be implicit, or at least obvious. Specifically, servers are well-known networking infrastructure, and servers were known to be used for back-end processing of biometric data. *See, e.g.*, *Derakhshani*, 9:27-58, 10:1-24; Ex-1016, Abstract, [0040]-[0043]; Ex-1012, Fig. 1A, 5:24-50. Furthermore, in my opinion, a POSITA would have understood that what types of data to send to the server is a design choice that balances the processing capabilities of the image-capture device and the transmission bandwidths available. If there is little bandwidth and ample processing power on the image capture device, a POSITA would be motivated to design the authentication on the device, so only a small amount of information needs to be transmitted to the server. If there is ample bandwidth available, but the image capture device is constrained in its processing power, a POSITA would be motivated to transmit raw image data to the server, so that the authentication could be done there. It would be obvious to a POSITA how to manage intermediate scenarios through standard engineering analysis.

6. Independent Claims 10 and 19

556. In my opinion, Zhang, alone or combined with Tanii, teaches the system claim 10 and method claim 19 for substantially the same reasons as method

claim 1 because, other than limitations 10[pre]-[a] and 19[a]-[b], the claims are substantively identical as shown in the table below.

Claim 1	Claim 10	Claim 19	Reference
1[Pre]	10[Pre]	19[Pre]	§XI.C.2.a
-	10[a]	-	
1[a]	10[b]	19[a]	§XI.C.2.b
1[b]	10[c]		§XI.C.2.c
1[c]	10[d]	19[b]	§XI.C.2.d
1[d]	10[e]		§XI.C.2.e
1[e]	10[f]	19[c]	§XI.C.2.f
1[f]	10[g]	19[d]	§XI.C.2.g
1[g]	10[h]	19[e]	§XI.C.2.h
1[h]	10[i]	19[f]	§XI.C.2.i
1[i]	10[j]	19[g]	§XI.C.2.j

a. 10[pre]: A system for verifying three-dimensionality of a user's face using images of the user's face captured using a camera equipped computing device, the system comprising:

557. If the preamble is limiting, in my opinion Zhang discloses or suggests limitation 10[pre].

558. In addition to disclosing a facial authentication process to verify three-dimensionality of a user's face, *see, e.g.*, §XI.C.2 (claim 1), Zhang also discloses an associated system for performing the facial authentication process. *Zhang*, [0013] (noting process can be performed on a "desktop computer, a laptop or notebook computer, a notepad computer, a mobile station…, a cellular or other wireless phone, a digital camera or video camera…and so forth.").

b. 10[a]: a computing device having a camera, screen, processor, and memory configured with non-transitory machine readable code that is executable by the processor, the machine readable code configured to:

559. In my opinion, Zhang discloses or suggests limitation 10[a].

560. Zhang discloses a computing device that contains an image capture component, a display, processors, and computer-readable media (e.g., memory) storing software instructions. *Zhang*, [0063]-[0067]. As I explained previously, a POSITA would have understood that Zhang's "image capture component" is a camera. *See* §XI.C.2.b (1[a]).

c. 19[a]: receiving first biometric data generated from at least one first image of the user taken with the camera of the computing device located at a first distance from the user;

561. In my opinion, Zhang discloses or suggests limitation 19[a].

562. Zhang discloses capturing a first image and processing the first image taken at a first distance to obtain first biometric data. *See* §§XI.C.2.b, XI.C.2.c (1[a], 1[b]).

563. In my opinion, a POSITA would have understood that limitation 19[a] presents a difference in perspective with respect to the system compared to claim 1. Specifically, where at least some of the limitations of claim 1 are from the perspective of the structure performing the method (e.g., a camera "captures," and a

processor "processes"), claim 19 recites limitations directed to the transmission of certain data (e.g., a processor performing the processing *receives* the biometric data).

564. Because Zhang, alone or combined with Tanii, already describes the recited structures performing the recited functions, however, a POSITA would have known that Zhang also performs the necessary data transmissions as well for carrying those functions out. Specifically, a POSITA would have read Zhang as disclosing, or at least obviously suggesting that the processor performing the processing does, in fact, *receive* the biometric data, whether from its own processor (when all functions are carried out on a single device), or a different processor (when some functions are performed by a server), in order to carry out the disclosed functions.

d. 19[b]: receiving second biometric data generated from at least one second image of the user taken with the camera of the computing device located at a second distance from the user, the second distance being different than the first distance;

565. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 19[b].

566. Zhang, alone or in combination with Tanii, discloses capturing and processing the second image to obtain second biometric data at a distance different than the first distance. *See* §§XI.C.2.d, XI.C.2.e (1[c], 1[d]). In my opinion, a POSITA would have understood that limitation 19[b] presents a difference of

perspective, but that Zhang, alone or combined with Tanii, teach these different perspectives as well. *See* §XI.C.6.c (19[a]).

7. Claim 11

a. 11[a]: The method according to claim 10, further comprising: interpolating the first biometric data and the second biometric data to obtain estimated intermediate biometric data;

567. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 11[a] for the reasons discussed in §XI.C.3.a (2[a]).

b. 11[b]: capturing at least one third image of the user taken with the camera of the computing device at a third distance from the user, the third distance being between the first distance and the second distance;

568. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 11[b] for the reasons discussed in §XI.C.3.b (2[b]).

c. 11[c]: processing the at least one third image to obtain third biometric data based on the at least one third image; and

569. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 11[c] for the reasons discussed in §XI.C.3.c (2[c]).

d. 11[d]: comparing the estimated intermediate biometric data with the third biometric data to determine whether the third biometric data matches the estimated intermediate biometric data.

570. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 11[d] for the reasons discussed in §XI.C.3.d (2[d]).

8. Claim 12: The method according to claim 10, further comprising verifying the presence of the user's ears in the at least one first image, and verifying the absence or reduced visibility of the user's ears in the at least one second image, wherein the first distance is larger than the second distance.

571. In my opinion, Zhang combined with Tanii teaches claim 12's additional limitation for the reasons discussed in §XI.C.4 (claim 3).

9. Claim 14: The method according to claim 10, wherein comparing the first biometric data to the second biometric data and the determining that the user's face is three-dimensional occurs at a server that is remote from the camera equipped computing device.

572. In my opinion, Zhang discloses or suggests claim 14's additional limitation.

573. Zhang discloses an image capture component (102) can be physically separate from live face detection module (104), and can communicate over a network (such as the internet or local area network). *Zhang*, [0014].

574. Although Zhang does not expressly disclose hosting the live face detection module (104) on a server to compare the first and second biometric data to determine whether the user's face is three-dimensional, in my opinion, a POSITA would have understood that sending data to a more-powerful server for processing is a design choice that balances the processing capabilities of the image-capture device and the transmission bandwidths available. *See Derakhshani*, 9:27-58, 10:1-

24; Ex-1016, Abstract, [0040]-[0043]; Ex-1012, Fig. 1A, 5:24-50; *see* §XI.C.5 (claim 9) (discussing the obvious use of conventional servers in Zhang's system).

575. For this reason, in my opinion, a POSITA would have been motivated to implement Zhang's live face detection module (104) on a server when providing facial authentication features on a computing device with fewer computational resources available.

10. Claim 16: The method according to claim 10, wherein the computing device comprises a laptop or desktop computer and, with the computing device stationary, the user moves from the first distance to the second distance.

576. In my opinion, Zhang combined with Tanii and Tahk teaches claim 16.

577. Zhang discloses that the face authentication process can be implemented on a variety of different type of computing devices, such as a desktop computer or laptop or notebook computer. *Zhang*, [0013].

578. When implementing a facial-authentication process on a more-stationary computing device (e.g., a desktop or laptop computer) consistent with Zhang, alone or in combination with Tanii, *see* §XI.C.2.d (1[c]), in my opinion, a POSITA would have further understood that the user would physically move their face closer or further from the camera to capture images of the face at different distances, because that is a convenient and obvious way of changing the distance between a larger computing device and the user's face.

11. Claim 18: The method according to claim 10, wherein the first biometric data and the second biometric data are maintained on the computing device.

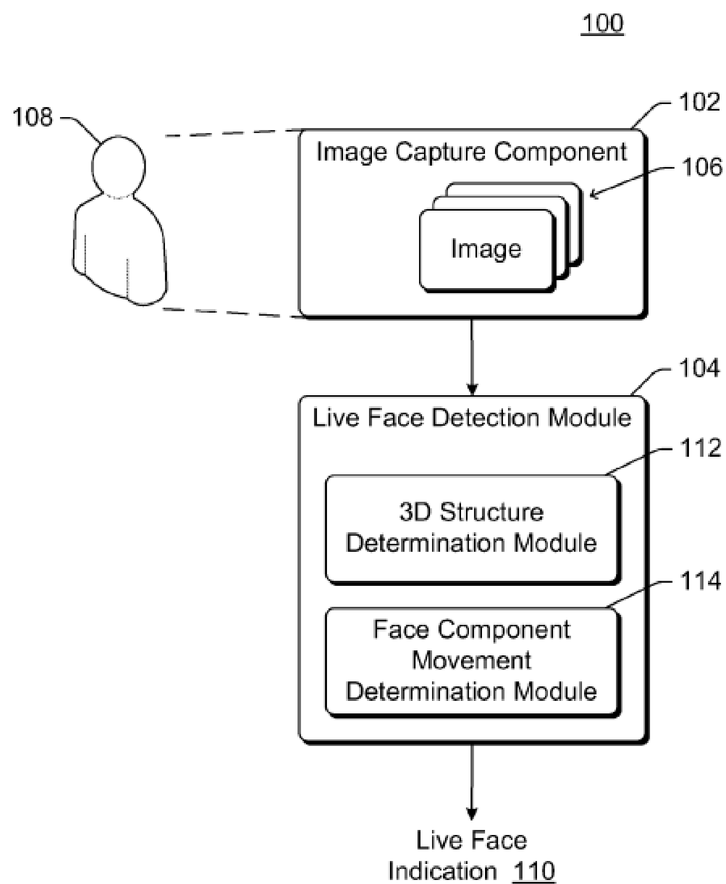
579. In my opinion, Zhang discloses or suggests claim 18.

580. Zhang discloses an image capture component (102) and a live face detection module (104) that can both be implemented on the same computing device. *Zhang*, [0014]. In such instances, in my opinion, a POSITA would have understood that the biometric data would be maintained on the computing device so that all processes would be performed using a single device. If a single, local device is being used for all authentication procedures, there would be no need to store any data on a separate device. Moreover, a POSITA would have understood that biometric data is extremely sensitive, and transmitting that data over networks presents a security risk of being intercepted. A POSITA would have understood that there would be no need to take any security risks by transmitting such data if the single, local device is handling the entire biometric-authentication procedure.

12. Claim 20: The method of claim 19, wherein the receiving of the first biometric data and the second biometric data occurs at a server and the first biometric data and the second biometric data are received over one or more of a LAN, WAN, or Internet type network.

581. In my opinion, Zhang discloses or suggest claim 20's additional limitation.

582. Zhang discloses an image capture component 102 and the live face detection module 104 "can communicate with one another via any of a variety of different networks, such as the Internet, a local area network (LAN)." *Zhang*, Fig. 1, [0014].



583. In my opinion, a POSITA would have also understood that offloading an authentication procedure from a local device to a server was well-known, and modifying Zhang to offload the authentication process to a server would have been straightforward. *See* §§VII.C.7 (claim 9), VII.C.11 (claim 14). In doing so, a

POSITA would have appreciated that the local device and server would obviously communicate over a network (such as WAN or LAN), to perform the authentication procedure—including transmitting first and second data from the capture device to the server for processing—particularly for devices with limited processor capabilities. *See* §§VII.C.7 (claim 9), VII.C.11 (claim 14).

D. Ground 2B: Zhang, Tanii, and Tahk (Claims 4-8, 13, 15-17)

1. Motivation to Combine

584. In my opinion, a POSITA would have been motivated to modify Zhang, with or without Tanii, in view of Tahk because Tahk provides a user-friendly way of ensuring that a face presented for facial authentication is properly framed. In my opinion, a POSITA would have understood that providing a real-time preview of what an image would look like prior to capturing the image, as well as providing express prompts—such as written instructions or oval shapes on the live-preview screen—as taught by Tahk, §VII.D (Tahk), would have been particularly useful for authentication procedures that require capturing multiple images of a face, as taught by Zhang and Tanii.

585. For instance, providing a user a real-time preview would allow the user to actively adjust the position of the camera and/or their orientation of their face to properly frame their face for the image capture. Moreover, providing express prompts and oval shapes sized to guide the user to properly position their face would

ensure images best suited for facial recognition can be captured, and that the faces would be captured from sufficiently different perspectives and/or distances to ensure the three-dimensional verification taught by Zhang, alone or in combination with Tanii, could be performed. This is particularly important for facial-recognition systems, which generally are known to have difficulty matching faces across different views. *See, e.g.*, Ex-1018, 32.

586. Moreover, a POSITA would have known that providing real-time image feedback, written instructions, and oval shapes to frame a face during a facial authentication process were all well-known and conventional techniques to provide user feedback during image capture as of the time of the invention. *See, e.g.*, Ex-1034, 7:16-8:7, Figs. 6B-7C (providing "an example of an interface used upon registering a facial image of a person to be authenticated" in which "the image of this person is displayed on a monitor" during registration, and oval-shaped prompts to indicate the size/distance of the user from the face authentication sensor); Ex-1035, 5:31-32 ("The computing device may present prompts that instruct the user to perform one or more liveness gestures"), 6:3-4 (same).

2. Claim 4: The method according to claim 1, wherein the computing device is configured to display one or more prompts on a screen of the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

587. In my opinion, Zhang combined with Tanii and Tahk teaches claim 4.

588. Zhang discloses taking a series of images sufficient to calculate a homography matrix. *See, e.g.*, *Zhang*, [0026], Figs. 1, 3. In my opinion, a POSITA would have understood that Zhang already discloses, or that Zhang combined with Tanii teaches, taking a series of images at different distances between the face and the camera. *See* §§XI.C.1 (motivation), XI.C.2.d (1[c]). However, Zhang and Tanii do not expressly teach providing a series of prompts to a user to guide them through different camera positions that would enhance calculations of the homography matrix.

589. Tahk, however, teaches that using one or more prompts on a screen ensures images of the face are captured at the correct distances. *See, e.g.*, Tahk, Figs. 8A-B ("Please step further back" and "Please step further forward," and presenting an oval to frame the face at the correct distance). In my opinion, a POSITA would have been motivated by Tahk to modify Zhang, whether alone or in combination with Tanii, to expressly prompt a user to alter the distance of the camera in order to either capture sufficiently different images to perform a homography transformation (Zhang) or to capture an image with distance-induced distortions (Tanii) to ensure the images could be used to distinguish live from two-dimensional images of faces. *See also* §XI.D.1 (motivation).

3. Claim 5: The method according to claim 4, wherein the one or more prompts are ovals sized on the screen within which the face of the user is placed to capture the at least one first image and the at least one second image at the first and second distances.

590. In my opinion, Zhang combined with Tanii and Tahk teaches claim 5.

591. Neither Zhang nor Tanii expressly teach using prompts to guide a user during the facial-authentication process. Tahk, however, teaches using oval prompts to frame a user's face. *See* §XI.D.2 (claim 4). And for the same reasons, a POSITA would have been motivated to modify Zhang, alone or in combination with Tanii, to provide such prompts because ovals are a natural shape to appropriately size and frame a face at different distances. *See* §§XI.D.1 (motivation), XI.D.2 (claim 4).

4. Claim 6: The method according to claim 4, wherein the computing device is a hand-held device, and the user holds the computing device at the first distance and the second distance to capture the at least one first image and the at least one second image.

592. In my opinion, Zhang combined with Tanii and Tahk teaches claim 6.

593. Zhang discloses that the face authentication process can be implemented on a variety of different types of hand-held computing devices, such as a cellular or other wireless phone, a digital camera or video camera. *Zhang*, [0013]. Moreover, Tanii notes that distance-induced distortions often occur in mobile devices that have incorporated wide-angle lenses, and the amount of distortion is dictated by the distance between the user and the camera. *Tanii*, [0007],

[0047]-[0048], Figs. 3A-B, 4A-B. And Takh also teaches the use of a camera on a mobile phone. *See, e.g.*, *Takh*, [0040], Figs. 2A-B, 5, 8A-B. And Takh also teaches the use of a camera on a mobile phone. *See, e.g.*, *Takh*, [0040], Figs. 2A-B, 5, 8A-B.

594. In my opinion, a POSITA would have also understood that, when performing a three-dimensional verification of the face on a mobile computing device, *see* §§XI.C.2.d (1[c]), XI.D.2 (claim 4), the user would adjust the distance by holding the mobile device and extending and retracting their arm (holding the mobile device at a first distance, then a second distance). A POSITA would have appreciated that mobile devices are routinely held to capture images, and holding the mobile device and adjusting distance would be a convenient and obvious way of changing the distance.

5. Claim 7: The method according to claim 6, wherein the computing device comprises a laptop or desktop computer and, with the computing device stationary, the user moves from the first distance to the second distance to capture the at least one first image and the at least one second image.

595. In my opinion, Zhang, alone or in combination with Tanii, teaches claim 7's additional limitation for the reasons discussed in §XI.C.10 (claim 16).

6. Claim 8: The method according to claim 1, further comprising displaying an image on a screen of the computing device while capturing the at least one first and/or the at least one second image.

596. In my opinion, Zhang combined with Tanii and Tahk teaches claim 8.

597. Zhang and Tanii do not expressly disclose displaying an image on a screen of the computing device while capturing the images of the user's face. Tahk, however, teaches displaying "a preview image for the face image." *See, e.g.*, *Tahk*, [0118], [0129], [0135], [0139], [0143], [0144], Fig. 8A-B.

598. In my opinion, a POSITA would have been motivated to modify the facial-authentication process taught by Zhang, alone or in combination with Tanii, to provide a live "preview" of the image to be captured so that the user could appropriately frame their face prior to capturing the image. *See* §XI.D.1 (motivation).

599. In fact, providing a preview of the image the camera is intended to capture has been known for as long as cameras have existed. For instance, many camera designs have incorporated viewfinders to provide the user a preview of the image. Ex-1032. And when devices such as feature phones and smartphones began incorporating cameras, they also included previews of the image intended to be captured on their displays. Ex-1033. In other words, displaying the view of the camera as the image is intended to be captured would be extremely well-known to a POSITA.

7. Claim 13: The method according to claim 10, wherein the computing device is configured to display one or more prompts on a screen of the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

600. In my opinion, Zhang combined with Tanii and Tahk teaches claim 13 for the reasons discussed in §XI.D.2 (claim 4).

8. Claim 15: The method according to claim 13, wherein the computing device is a hand-held device, and the user holds the computing device at the first distance and the second distance to capture the at least one first image and the at least one second image.

601. In my opinion, Zhang, alone or in combination with Tanii, teaches claim 15's additional limitation for the reasons discussed in §XI.D.4 (claim 6).

9. Claim 17: The method according to claim 10, further comprising displaying an image on a screen of the computing device while capturing the at least one first image and the at least one second image.

602. In my opinion, Zhang combined with Tanii and Tahk teaches claim 8 for the reasons discussed in §XI.D.6 (claim 8).

XII. '938 PATENT: DETAILED EXPLANATION OF GROUNDS

A. Ground 1A: Derakhshani and Tanii (Claims 1-10, 12-24)

1. Motivation to Combine

603. In my opinion, a POSITA would have been motivated to combine Derakhshani and Tanii because both concern identifying and accounting for the three-dimensional nature of a face when capturing an image. They differ, however,

in what principles are used to account for the face's three-dimensionality. Derakhshani, for instance, uses changes in focus distance (e.g., image resolution for structures imperfectly in focus) and/or parallax effect to determine whether a face has depth. *See* §VII.A (Derakhshani). And although Tanii is not expressly directed to *evaluating* whether a face has depth, Tanii exemplifies the well-known distortions caused by the interaction between the camera's lens and the three-dimensional nature of the face at different distances, *see* §VII.B (Tanii). A POSITA would have appreciated, therefore, that Tanii recognizes another alternative to evaluating the depth of a face, consistent with Derakhshani's existing two approaches.

604. A POSITA would have recognized, for instance, that Derakhshani's focus-distance approach and Tanii's evaluation of distance-induced distortions are both attributable to classical optical effects such as refraction and diffraction caused by (among other factors) different distances between the camera and the object(s) being captured. *Derakhshani*, 16:57-60 ("Degree of focus is a measure of the extent to the image of the landmark is blurred by optical effects … (e.g., due to *diffraction* and convolution with the aperture shape."); *Tanii*, [0048] (noting the "unnatural image" is caused by the angles of the face relative to the angle of the camera lens).

605. Derakhshani and Tanii differ, however, in the type of effect that is occurring. Specifically, Derakhshani takes advantage of the blurring of objects that are at distances *other than* the camera's focal plane (referred to by photographers as

a "bokeh effect"), which makes those objects appear unfocused. *Derakhshani*, 16:54-57; §VII.A (Derakhshani). By adjusting the focus distance (or position of the focal plane by moving the camera) and evaluating when objects (or features of an object) in an image are clear versus when they are blurry, distance information can be derived. *Derakhshani*, 16:51-63; §VII.A (Derakhshani).

606. Tanii is more specifically concerned with a type of radial distortion that arises due to the interaction of certain (e.g., wide-angle) lenses and the three-dimensional nature of the face. §VII.B (Tanii). As Tanii explains, the convex shape of a three-dimensional face, when placed near the lens, exacerbates this type of distortion. *Tanii*, [0048]; §VII.B (Tanii). Thus, particularly when a camera incorporates a wide-angle lens, images of a face close to the camera will exhibit significant radial distortion in-part because of the distances between different facial features and the lens, and in-part because the face occupies both the center and the periphery of the camera's field of view so differences in radial distortion are more apparent. *Tanii*, [0047]; §VII.B (Tanii). But when the face is further from the camera and occupies less of the image, the distortion will be less apparent because the face is more centered on the region of the lens where radial distortion is not as severe, and there is sufficient distance for the light rays from the face to strike this central portion of the lens. *Tanii*, [0047]; §VII.B (Tanii).

607. In my opinion, a POSITA would have appreciated that when evaluating multiple images taken at either different *focus* distances or *actual* distances, these different effects serve to provide information about an object's depth. In other words, a POSITA would have understood that Tanii teaches another obvious alternative to Derakhshani's existing two approaches to evaluate whether a face being captured is three-dimensional or not.

608. That said, a POSITA would have also had specific reasons to substitute Derakhshani's existing approaches with Tanii's distance-induced distortion analysis in certain circumstances. A POSITA would have understood, for instance, that implementing Derakhshani's focus-distance approach requires a camera with a sufficiently sized sensor and lens that could provide enough sensitivity to distinguish small differences in depth on the scale of a few centimeters when trying to evaluate the depth of a face. *See Derakhshani*, 16:48-51; Ex-1029, 3 (A 200mm lens focused at 12ft will have a smaller depth of field compared to a 20mm lens focused at 12ft).

609. But a POSITA would have also understood that the cameras typically found in mobile devices—especially around the 2014 timeframe—do not have this ability; mobile devices typically incorporate wide-angle lenses to capture a wide field of view, with a fixed focal length and a large depth of field because of their small size. *Tanii*, [0007]; Ex-1030 ("Other features of a smartphone are obvious but worth stating, they almost always are fixed focal length, fixed aperture, with no

shutter, sometimes with an ND filter (neutral density) and generally not very low F-number. In addition to keep modules thin, focal length is usually very short, which results in wide angle images with lots of distortion.”). With such limited-capability cameras, it was known that distortions would therefore largely be a product of the lens shape and distance between the object and the lens. *See* Ex-1017, 177 (“The amount of spherical aberration, when the aperture and focal length are fixed, varies with both the ***object distance*** and the lens shape.”). In other words, there is not enough room in mobile devices to incorporate large image sensors with small F-numbers (a measure of light-gathering ability of the camera) to allow these cameras to fine-tune the focus distance and induce blurring of out-of-plane objects. That is why, for instance, the iPhone introduced its “Portrait Mode” (in 2016, a few years after the earliest possible effective date) as a software-based *simulation* of the blurring effect that can only be achieved by much larger cameras. Ex-1031 (noting how blurring backgrounds was “previously only capable on DSLR cameras” prior to the iPhone’s software-based “bokeh” effect).

610. For this reason, in my opinion, a POSITA would have been motivated to modify Derakhshani to capture at least two images at different *actual* distances and evaluate whether one exhibits more distance-induced distortion than the other, as suggested by Tanii. A POSITA would have been especially motivated to make this change when implementing biometric authentication in a mobile device as

Derakhshani already envisions. *Derakhshani*, 5:23-26. A POSITA would have found such a modification obvious because both techniques merely involve the application of different well-known optics principles relating camera design and object's distance from the camera, and would have had a reasonable expectation of success in doing so because Tanii already taught a mechanism to identify such distance-induced distortions and thus indicate when the face being captured has depth. *See, e.g.*, *Tanii*, [0056].

611. Although Derakhshani separately discloses a process to verify the three-dimensionality of a face using parallax, in my opinion, a POSITA would have understood that evaluating for distance-induced distortion consistent with Tanii would be easier for users on a mobile device. Specifically, a POSITA would have naturally understood that mobile devices such as phones or laptops typically capture images of users at arm's length distances because that is how these devices are used (at arm's length). Moreover, a POSITA would have appreciated that facial features do not have *significant* differences in their depth (on the order of a few centimeters, as opposed to meters between the face and a background). Thus, to evaluate for parallax at hand-held distances with suitable accuracy, a POSITA would expect that the user would need to move their device around their head, or could simulate a parallax effect by rotating their head around a stationary camera to create substantial differences in perspective and thus more parallax to more accurately verify the face

as three-dimensional. But to do so would have involved moving the device out of the user's line of sight, meaning the user could not see exactly what they are capturing or know if what they were capturing is sufficient. Evaluating for distance-induced distortions when the camera is held at different distances consistent with Tanii, however, could be accomplished while keeping the device directly in the user's direct line of sight, and would therefore be easier for users to verify that their face is, in fact, three dimensional. But, in my opinion, a POSITA would have also appreciated that biometric security is always subject to spoofing, and thus would have known that evaluating for distance-induced distortion consistent with Tanii could be *supplemented* by also evaluating for any parallax.

2. Independent Claim 1

a. 1[pre]: A non-transient computer readable medium containing non-transitory machine executable code configured to determine if the three-dimensional shape is consistent with that of a human face, the non-transitory machine executable code configured to:

612. If the preamble is limiting, in my opinion, Derakhshani discloses or suggests it.

613. Derakhshani discloses a "computing device" with "a machine-readable repository," *Derakhshani*, 7:15-23, that can run a "computer program" with "instructions that, when executed, perform one or more methods, such as those described," *id.*, 22:51-64; *see also id.*, 24:61-25:8. By disclosing a machine-readable

repository and a program that runs on stored instructions, in my opinion, a POSITA would have understood that Derakhshani is referring to conventional "non-transient" (e.g., non-temporary) computer memory—which is a computer-readable medium—with "non-transient" (e.g., non-temporary, stored) executable code to perform Derakhshani's process.

614. Derakhshani also discloses an authentication process to be run using a computer program that verifies the user's face is three-dimensional by capturing multiple images of a user's face at different focus distances or from different perspectives to calculate a "spatial metric" representing the face's three-dimensionality. *Id.,* 1:11-25, 3:14-15, 16:44-18:4.

b. 1[a]: receive or derive first biometric data from at least one first image of a user taken with a computing device camera located at a first distance from the user;

615. In my opinion, Derakhshani discloses or suggests limitation 1[a].

616. Derakhshani discloses that, as part of the verification process, "two or more images of a subject" are captured using the camera of the computing device. *Derakhshani*, 1:44-46, 16:44-17:11, 17:45-18:4.

617. In my opinion, a POSITA would have understood that Derakhshani captures an image at a first distance. *Derakhshani*, 16:44-17:11. Specifically, a POSITA would have understood that there *must* be some distance between the camera and the face to capture the first image—enough so that the camera's field of

view encompasses the face. If there were no distance between the camera and the face (e.g., if the camera were pressed up against the user's skin), then the field of view would be limited to just that patch of skin and the face would not be captured in the image, which would be useless for the three-dimensional verification process.

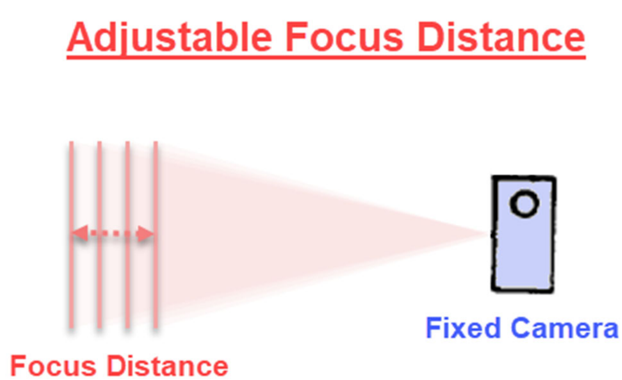
618. Derakhshani also discloses that, as part of the process to verify that the face is in fact three-dimensional, "a landmark (e.g., an iris, an eye corner, a nose, an ear, or a background object) may be identified and located in the plurality of images." *Derakhshani*, 16:44-54 (focus-distance approach), 17:45-64 (parallax approach).

619. In my opinion, a POSITA would have understood that Derakhshani's identification of facial landmarks constitutes deriving "biometric data" because "biometric data" generally refers to unique physical characteristics of an individual, which would include the positions of "landmarks" such as a user's eyes, nose, ears, and other such features. *See, e.g.*, Ex-1018, 2 ("biometric recognition can be defined as the science of establishing the identity of an individual based on the physical and/or behavioral characteristics of the person."), 100-103 (describing the types of biometric data about a face used for facial-recognition systems, "such as the structure of the face components (e.g., eyes), [and] the relationship between facial components").

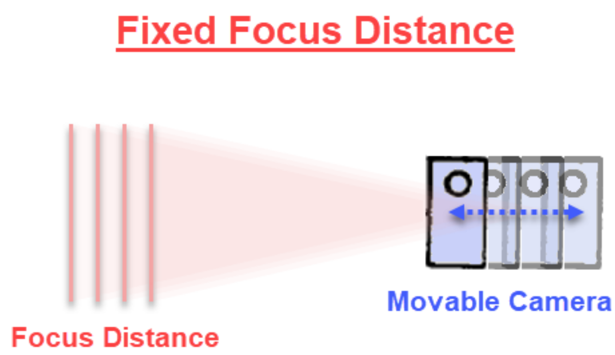
c. **1[b]: receive or derive second biometric data from at least one second image of the user taken with the computing device camera located at a second distance from the user, the second distance being different than the first distance;**

620. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 1[b].

621. Derakhshani discloses capturing "two or more images of a subject" using the camera. *Derakhshani*, 1:44-46, 16:44-17:11, 17:45-18:4; §XII.A.2.b (1[a]). When utilizing Derakhshani's focus-distance approach to evaluate depth, however, a POSITA would have understood that adjusting the *focus* distance of the camera does not require changing the *actual* distance between the camera if a stationary camera is capable of adjusting its focus distance. *See* §VII.A (Derakhshani explaining operation of the focus-distance approach).



But if the camera has a fixed focus distance (i.e. position of the lens with respect to the image sensor), as is found in many mobile devices (*see* §XII.A.1), a POSITA would have been motivated to instead implement Derakhshani's focus-distance approach by changing *actual* distance to capture multiple images, as shown below:

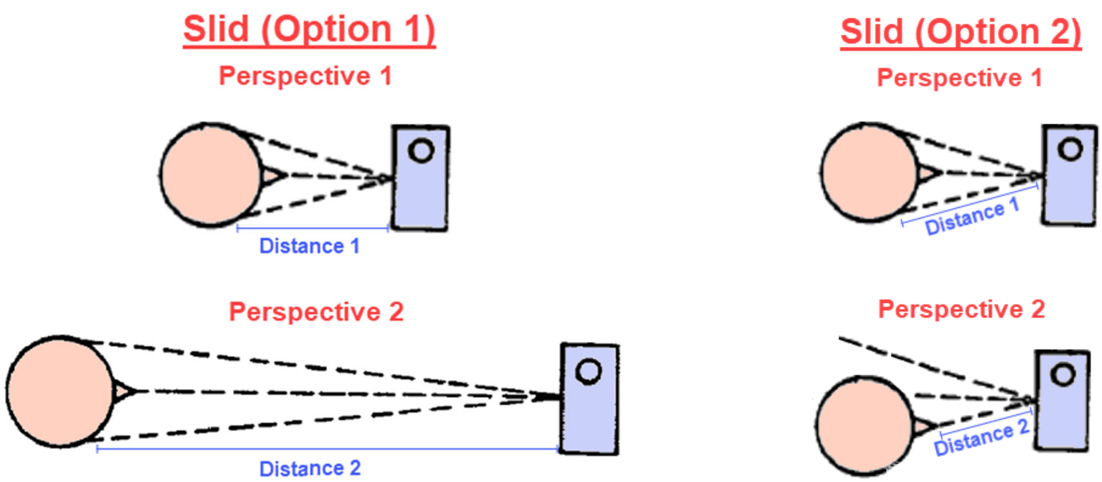


In other words, even if the focus distance of the camera cannot be changed, the "slices" of a face at different depths can be evaluated by moving the camera.

622. Regardless, in my opinion, a POSITA would have understood that Derakhshani's parallax approach captures multiple images from multiple distances, because Derakhshani discloses that "[a] plurality of images [are] taken from different perspectives on the subject," such as: (1) when "a single camera [is] rotated *or slid* slightly"; (2) "a user is prompted to move" between image captures; or (3) the sensor moves naturally, such as "where the sensor is a camera in a hand-held user device (e.g., a smartphone or tablet) [that] may naturally move relative to the users face due to involuntary haptic motion." *Derakhshani*, 17:45-18:4.

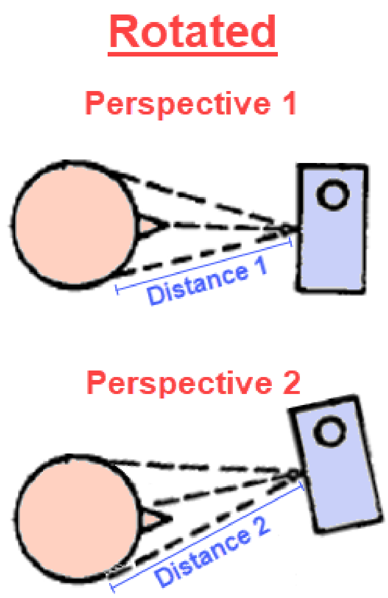
623. In my opinion, a POSITA would have understood that Derakhshani's use of the term "slid" means either of two things: (1) the camera is displaced front-

to-back to increase or decrease the distance from the face; or (2) the camera is displaced side-to-side, both of which are depicted below:



In either case, a parallax effect would be evident if the face were three-dimensional because of the different perspectives of the face captured in each. For instance, a POSITA would have recognized that, with a front-to-back translation, more of the periphery of the face would be captured by the camera, and there may be other optical effects (e.g., distance-induced distortion) that are more apparent in the closer image than the further one. And with side-to-side translation, more features on the side of the face the camera favors would be captured, but features on the other side of the face may be obstructed due to the face's three-dimensionality.

624. Moreover, a POSITA would have understood that Derakhshani's use of the term "rotated" means the camera itself is rotated relative to the face. I have provided an example of rotation below that also includes some side-to-side translation to keep the face centered on the camera.



625. As these exemplary figures demonstrate, however, a POSITA would have understood that, regardless of whether the camera is "slid" or "rotated," distances between facial landmarks and the camera will change. In my opinion, a POSITA would have understood that any of these options results in a "second image of the user taken with the computing device camera located at a second distance from the user, the second distance being different than the first distance" as claimed, because there is no one single "distance" between the camera and a three-dimensional user when changing the position/perspective of the camera; some distances will always change. However, even if the claims were limited to a front-to-back translation to change the *overall* distance between the camera and the user, a POSITA would have understood that Derakhshani discloses or suggests as much.

626. But even if Derakhshani does not expressly disclose taking two images at different distances, in my opinion, a POSITA would have been motivated to look to differences in the degree of distance-induced distortions exemplified by Tanii as an alternative or additional evaluation of the three-dimensionality of the face besides Derakhshani's focus-distance and parallax approaches. §XII.A.1 (motivation). When making this modification, a POSITA would have been motivated to modify Derakhshani in view of Tanii to expressly capture a second image at a second distance, and look for more distance-induced distortions in one image compared to the other to determine whether the face has depth. §XII.A.1 (motivation). Moreover, a POSITA would have had a reasonable expectation of success in making this modification because Tanii already taught a mechanism to identify such distance-induced distortions and thus indicate when the face being captured has depth. *See, e.g.*, *Tanii*, [0056].

627. Finally, Derakhshani discloses processing the captured images to identify biometric "landmarks" in the face as part of the three-dimensional verification process, (*Derakhshani*, 17:45-52), which a POSITA would have understood to constitute deriving "biometric data," (§XII.A.2.b (1[b])).

d. 1[c]: compare the first biometric data with second biometric data for expected differences that result from characteristics of a human face and the at least one first image and the at least one second image being captured at different distances from the user;

628. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 1[c].

629. Derakhshani discloses comparing the first biometric data to the second biometric data to determine whether differences exist between the two. *Derakhshani*, 16:66-17:2 (for focus distance, "[b]y comparing the degree of focus for a landmark in images with different focus distances, the distance from the sensor to the landmark may be estimated."), 17:55-59 (for parallax, "[i]f all the landmarks in the images undergo the same apparent displacement due to the relative motion of the sensor…then the subject viewed by the camera has a likelihood of being a two-dimensional spoof attack."). Moreover, Derakhshani's focus-distance approach looks for expected differences in the blurriness or clearness of facial landmarks by changing the *actual* distance (for fixed-focus cameras), and Derakhshani's parallax approach looks for expected differences in the relative displacement of different facial landmarks by changing the *actual* distance alone. *See* §XII.A.2.c (1[b]).

630. In my opinion, Derakhshani describes a comparison between images that looks for "expected" differences consistent with how the '938 Patent uses the term because one would *expect* that following either the focus-distance or parallax

approaches Derakhshani discloses would produce specific differences: the focus-distance approach would capture some images where certain facial features are blurred and others where those same features are clear, and the parallax approach would produce images with facial features displacing by different amounts relative to one another depending on the change of perspective and distance between the specific features and the camera lens. And a POSITA would have appreciated that those expected differences between the two images would be attributable to "characteristics" of a human face—specifically, its three-dimensionality—and the fact that distance (whether focal distance or actual distance) changed between the two images, because it is this change of distance that imparts the differences in focus of a three-dimensional face or changes in the relative displacement of facial features.

631. Relatedly, a POSITA would have been particularly motivated to configure Derakhshani to capture images with specific, pre-defined configurations (e.g., a specific set of focus distances, or a specific position of the camera relative to the face) to minimize the variability between the images used for facial recognition and specifically tailor the system to look for expected changes between images. For example, Derakhshani's focus-distance approach (with its loss of spatial frequency) would improve its performance if images were acquired with the face at different distances from the camera. Doing this with two or more distances would remove range ambiguity and decrease the variance in estimates of the distance from the

camera to particular features. *See, e.g.*, Ex-1018, 32 (noting how facial-recognition systems often require controlling conditions such as a "fixed and simple background with controlled illumination" because "systems … have difficulty in matching face images captured from two different views, under different illumination conditions, and at different times."). In other words, rather than permitting users to change the focus distance or perspective of the camera any way they wish, which would require a system that could account for such variabilities, having the user follow a pre-determined protocol to capture images at set focus distances or perspectives would simplify the matching process.

632. In my opinion, however, a POSITA would have also understood that, when utilizing the distance-induced distortion approach exemplified by Tanii, the images captured from that process would also exhibit expected distortion based on the distance between the camera and the face. §XII.A.1 (motivation). In my opinion, a POSITA would have been motivated to look for and utilize these expected differences in distortion as an alternative or supplemental verification of three-dimensionality of a face in Derakhshani, particularly in mobile devices that incorporate wide-angle lenses. §XII.A.1 (motivation). In doing so, a POSITA would have understood that verifying a three-dimensional face using distance-induced distortion would be accomplished by matching the positions of biometric features across the first and second images—as Derakhshani already discloses—but rather

than look for blurriness/clearness or parallax of those biometric features, the images would instead be evaluated for expected differences in the distortion of those features caused by the distance-induced distortion. §XII.A.1 (motivation).

e. 1[d]: determine that the three-dimensional shape is not exhibited when the second biometric data does not have expected differences compared to the first biometric data, the expected differences comprising at least differences due to the change in the relative distance between the user's facial features and the camera when the at least one first image was captured at the first distance and the at least one second image was captured at the second distance, wherein the expected differences result from fish-eye type distortion in at least one of the at least one first image and the at least one second image and due to the three-dimensional nature of the human face and the change in distance between the camera and the user.

633. In my opinion, Derakhshani alone or in combination with Tanii teaches limitation 1[d].

634. Derakhshani discloses that a determination is made whether or not a face is three-dimensional depending on whether mismatches exist between the biometric landmarks (e.g., biometric data) using either the focus-distance or parallax approach. §XII.A.2.d (1[c]). Specifically, a face would not be considered three-dimensional if, using the focus-distance approach, facial landmarks are not blurry in some images and clear in others—but are instead either all blurry or all clear, together—because that suggests all facial landmarks exist on the same (two-dimensional) plane. *See, e.g.*, *Derakhshani* 16:44-17:44. And a face would not be

considered three-dimensional if, using the parallax approach, all facial landmarks are displaced by the exact same amount because that also suggests all facial landmarks exist on the same plane. *Id*., 17:12-18:4. In other words, a POSITA would have understood that, under either approach, Derakhshani looks for a mismatch between the first and second biometric data to determine whether expected differences between the images exist, and if they do not, the image is considered to be of a two-dimensional face.

635. Although Derakhshani does not expressly look for an expected "fish-eye" distortion in at least one of the two images, evaluating images for this type of distortion would have been obvious in view of Tanii. Specifically, a POSITA would have been motivated to modify Derakhshani's three-dimensional verification method to look for different degrees of "fish-eye" type distortion depending on the distance between the user's face and the camera (and more specifically the camera's lens and image sensor), because this type of distortion was also known to be attributable to the three-dimensional geometry of the imaged objects, similar to Derakhshani's focus-distance and parallax approaches. §XII.A.1 (motivation); XII.A.2.d (1[c]). A POSITA would have therefore understood that, when modifying Derakhshani in view of Tanii, the combination would evaluate whether facial features exhibit different expected degrees of "fish-eye" distortion when images are captured at different distances, and if the images do ***not*** exhibit different degrees of

"fish-eye" distortion, a POSITA would have understood that it would be determined that the face is ***not*** three dimensional because only a three-dimensional face would exhibit these types of distance-induced distortions. §XII.A.1 (motivation).

3. Claim 2: The non-transient computer readable medium of claim 1 wherein the expected differences appear as changes in the relative size and shape of facial features of the user.

636. In my opinion, Derakhshani alone or combined with Tanii teaches claim 2.

637. Derakhshani uses facial landmarks, such as "an iris, an eye corner, a nose, an ear," to determining three-dimensionality using at least two images *Derakhshani*, 16:51-52.

638. In my opinion, a POSITA would have understood generally that, keeping all else equal (such as the camera lenses and settings), changing the distance between the face and camera will inherently change the size of facial landmarks because changing the distance will alter the lateral magnification of the camera lens. For instance, other prior art (Tahk) depicts this principle in practice:

FIG. 8A

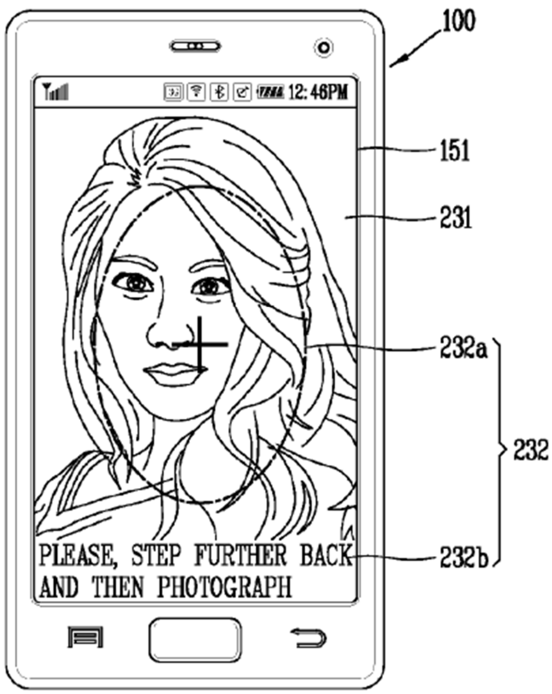
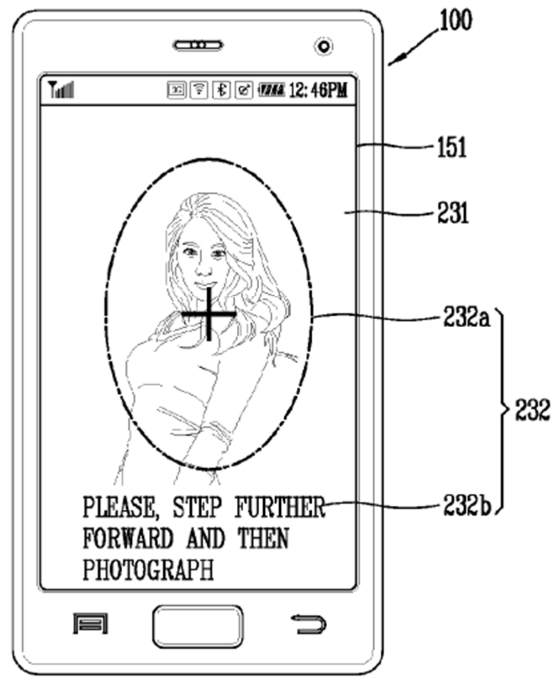


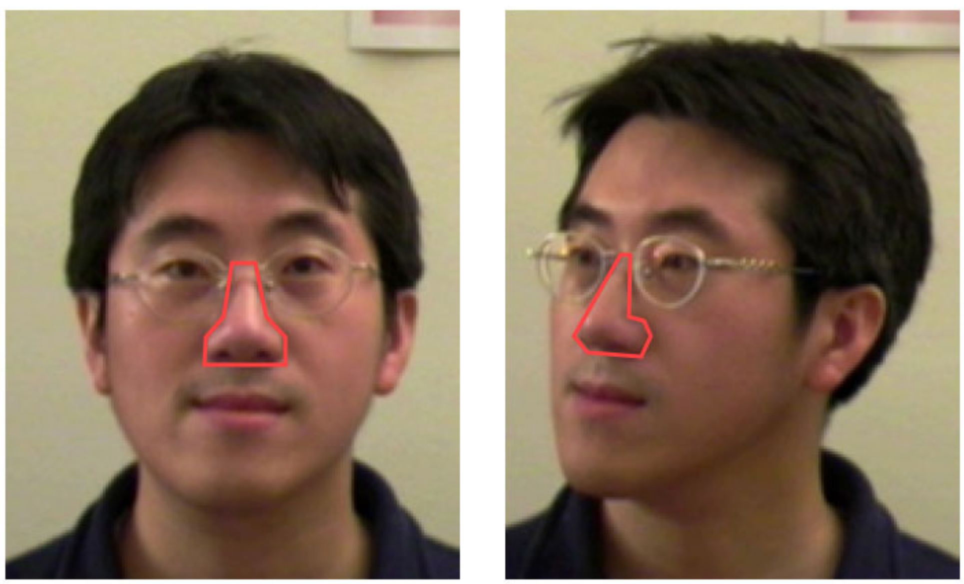
FIG. 8B



Tahk, Figs. 8A-B. Specifically, Tahk's Figure 8A shows a face that is closer to the camera than Figure 8B, making the facial features appear bigger than when the image is captured further away.

639. Relatedly, a POSITA would have further appreciated that Derakhshani's focus-distance and parallax approaches evaluate changes in the *shape* of facial features. Specifically, because Derakhshani's focus-distance approach looks for changes in the blurriness of facial features between two images, a POSITA would have appreciated that a blurry facial landmark will have a different (less-defined) shape compared to a clear facial landmark in another image:

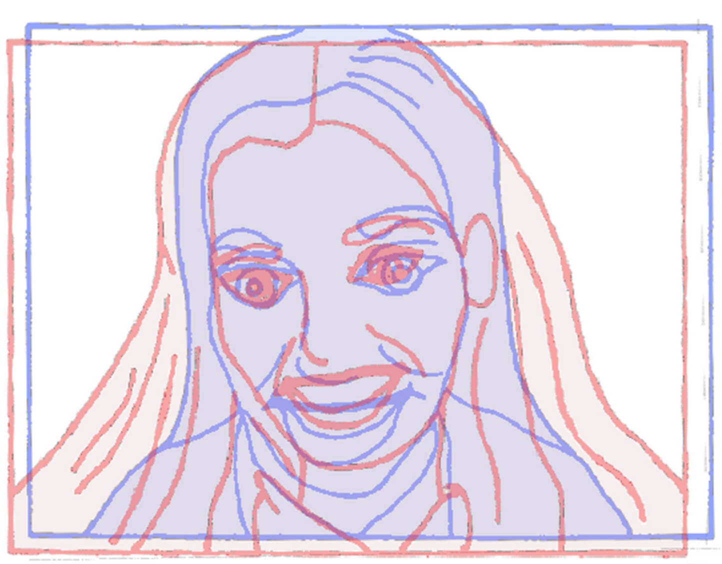
And Derakhshani's parallax approach envisions changing perspectives, which a POSITA would have understood meant different facial landmarks may be captured from different perspectives, giving the appearance of changed shape, as demonstrated below:



(a) (b)

Ex-1018, 99 (annotated). Although this example depicts a significant change in perspective, a POSITA would have appreciated that smaller changes of perspective would also change the apparent shape of facial features, but to a smaller degree.

640. When modifying Derakhshani in view of Tanii to look for different degrees of distance-induced distortion, however, a POSITA would have appreciated that the expected differences in a face caused by distance-induced distortion is the relative size and shape of facial features of the user. *Tanii*, [0047] ("where the main object 9 and the cellular phone 1 are close together…an unnatural image results in which the perspective is exaggerated."), [0056] (describing "a warp in which the peripheral areas of the main object appear reduced in size relative to the center area.").



In my opinion, a POSITA would have therefore known to look for these expected differences in size and shape of facial features to determine whether the face is three-dimensional.

4. Claim 3: The non-transient computer readable medium of claim 1 wherein determining that three-dimensionality is not exhibited happens during an authentication session.

641. In my opinion, Derakhshani discloses or suggests claim 3's additional limitation.

642. Derakhshani discloses that a "liveness score" is calculated during an authentication session as part of Derakhshani's anti-spoofing countermeasures. *Derakhshani*, 4:53-63. Derakhshani also discloses that the "liveness score" is based on one or more "liveness metrics" including the "spatial metric" to verify a user's face is three-dimensional. *Derakhshani*, 14:59-63, 15:26-31, Fig. 7

643. In my opinion, a POSITA would have therefore understood that calculating Derakhshani's "spatial metric" occurs during an authentication session to determine whether a live face is being presented for authentication. *See Derakhshani*, 9:39-48, 11:5-16. Moreover, in my opinion, a POSITA would have been motivated to modify the calculation of Derakhshani's spatial metric to calculate differences in distance-induced distortion—instead of or in addition to Derakhshani's existing focus-distance or parallax approaches—to verify the three-dimensionality of a face during an authentication session, in view of Tanii. *See* §XII.A.1 (motivation).

5. Claim 4: The non-transient computer readable medium of claim 1 wherein the computing device camera is part of a computing device and the machine executable code is configured to display an interface on the computing device's screen to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

644. In my opinion, Derakhshani, alone or combined with Tanii, teaches claim 4's additional limitation.

645. Derakhshani discloses that the invention can be implemented in computing devices such as a "smart phone, a tablet computer, a television, a laptop computer, or a personal computer," *Derakhshani*, 5:22-27, which incorporate a camera, *id.*, 5:23-27, 6:3-10, and a display. *Id.*, 6:8-11, 9:22-24, 14:35-37, 22:33-38, 25:9-15. Derakhshani also discloses displaying prompts to a user to guide the user to capture images of the user's face for authentication, *id.*, 5:23-32, 6:8-16, 9:22-26, including at more than once distance, *id.*, 17:64-66; §XII.A.2.c (1[b]).

646. And even if Derakhshani does not expressly disclose taking two images at different distances, doing so would have been obvious in view of Tanii to identify distance-induced distortions that indicate depth of a three-dimensional face. §§XII.A.1 (motivation), XII.A.2.c (1[b]). When modifying Derakhshani to look for distance-induced distortions by capturing images at different distances consistent with Tanii, in my opinion, a POSITA would have been motivated to provide prompts to a user to ensure the images are captured at the correct distances because

Derakhshani already discloses providing prompts to correctly orient the user relative to the camera.

6. Claim 5

a. 5[a]: The non-transient computer readable medium of claim 1 wherein the machine executable code is further configured to compare at least portions of the first data, second data, or both to enrollment data derived from an enrollment image, the enrollment image captured and stored prior to an authentication session; and

647. In my opinion, Derakhshani discloses or suggests 5[a]'s additional limitation.

648. Derakhshani discloses capturing and analyzing multiple images of a user and comparing the user's features to a previously stored "reference record" to authenticate the user. *Derakhshani*, 4:19-24; 7:20-34; 8:60-64; 9:31-34. In my opinion, a POSITA would have understood the "reference record" to be "enrollment data" because the process Derakhshani describes to generate and then use the "reference record" for authentication is consistent with a typical biometric-authentication enrollment procedure. *See* §V.A (biometric security overview). Specifically, Derakhshani discloses that the system captures one or more initial reference images of the user during a registration process, extracts features from the reference images, stores the extracted features as the reference record, and then subsequently compares later-captured images to the reference record. *Derakhshani*,

7:19-34 ("To create a reference record *for a new user* and enrollment or registration process may be carried out."); 9:31-34 ("The collection of image data from user may also facilitate authentication against a reference record for a user identity."); 13:62-14:9 (describing authentication matching against a reference record). Then, during the authentication process, Derakhshani compares the extracted features from the captured images (i.e., portions of the first data, second data, or both) to the user's enrollment reference record to determine a match score. *Id*. 9:59-67; 13:62-14:9; 17:32-36. This is consistent with a conventional biometric enrollment and authentication process. *See, e.g.*, Ex-1018, 4-11 (providing overview of biometric authentication and verification).

b. 5[b]: determining the user is not authenticated when the first data, the second data, or both do not sufficiently correspond to the enrollment data.

649. In my opinion, Derakhshani discloses or suggests 5[b]'s additional limitation.

650. Derakhshani discloses calculating a match score during the authentication process based on the comparison of features extracted from the first and second image to the corresponding features in an enrollment reference record. *Derakhshani*, 13:62-14:9. Derakhshani also discloses that, when the match score is low—because the first or second data, or both, do not sufficiently correspond to the enrollment data—it is determined the user is not authenticated. *Id*., 14:25-35. This

is consistent with conventional biometric-authentication processes. *See, e.g.*, Ex-1018, 17 ("[B]iometric systems mostly decide on a person's identity based on a *close* match between the template and the query, where the strength of the match (or the degree of similarity) is represented by the match score.").

7. Claim 6: The non-transient computer readable medium of claim 1 wherein the computing device camera is part of a computing device and the computing device is a hand-held device, and the user holds the device at the first distance to capture the at least one first image and then holds the computing device at the second distance to capture the at least one second image.

651. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 6's additional limitation.

652. Derakhshani discloses that the biometric-authentication process can be implemented on a variety of different type of hand-held computing devices, such as "a laptop computer, a handheld computer…, a tablet computing device, a personal digital assistant (PDA), a cellular telephone…, a camera, a smart phone," and more, *see, e.g.*, *Derakhshani*, 8:11-28, 18:1-4, which incorporate a camera, *id*., 5:23-27, 6:3-10. Derakhshani also recognizes that, to verify three-dimensionality of the face, "a single camera may be rotated or slide slightly," or that, when the device is hand-held, "the [camera] sensor may naturally move relative to the users face due to involuntary haptic motion" that may sufficiently capture a parallax effect. *Id.*, 17:59-18:4. Similarly, Tanii recognizes that distance-induced distortions often occur in

mobile devices that have incorporated wide-angle lenses, and the amount of distortion is dictated by the distance between the user and the camera. *Tanii*, [0007], [0047]-[0048], Figs. 3A-B, 4A-B.

653. In my opinion, a POSITA would have also understood that, when performing a three-dimensional verification of the face on a mobile computing device, *see* §§XII.A.2.c (1[b]), XII.A.2.d (1[c]), XII.A.2.e (1[d]), the user would adjust the distance by holding the mobile device and extending and retracting their arm (holding the mobile device at a first distance, then a second distance). A POSITA would have appreciated that mobile devices are routinely held to capture images, and holding the mobile device and adjusting distance would be a convenient and obvious way of changing the distance. In fact, Derakhshani already envisions evaluating depth based on the displacement of the user's arm while holding the device. §VII.A (Derakhshani); *Derakhshani*, 16:44-11, 17:45-18:4.

8. Claim 7: The non-transient computer readable medium of claim 1 wherein the first biometric data and the second biometric data comprise image data of facial features.

654. In my opinion, Derakhshani discloses or suggests claim 7's additional limitation.

655. Derakhshani discloses processing the captured images to identify and locate facial biometric "landmarks" (e.g., an iris, an eye corner, a nose, a mouth, an ear). *Derakhshani*, 16:44-54. In my opinion, a POSITA would have understood that

Derakhshani's identification of facial landmarks from the captured images constitutes image data of facial features. In fact, many different conventional methods were known to use image data to perform facial authentication. *See, e.g.*, Ex-1018, 109-27.

9. Independent Claim 8

a. 8[pre]: A method for determining when a user, based on images of the user's face, does not exhibit three-dimensionality, the method comprising:

656. If the preamble is limiting, in my opinion, Derakhshani discloses or suggests it.

657. Specifically, Derakhshani discloses a method (carried out by computer-readable code) to determine images of a user's face does not exhibit three-dimensionality. *See* §XII.A.2.a (1[pre]).

b. 8[a]: capturing at least one first image of the user's face taken with a camera located at a first distance from the user's face, the camera associated with a computing device;

658. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 8[a] for the reasons discussed in §XII.A.2.b (1[a]).

c. 8[b]: processing the at least one first image or a portion thereof to create first data;

659. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 8[b] for the reasons discussed in §XII.A.2.b (1[a]).

d. 8[c]: moving the camera to a second distance from the user's face, where the second distance is different from the first distance;

660. In my opinion, Derakhshani alone or combined with Tanii teaches limitation 8[c].

661. Derakhshani (whether alone or combined with Tanii) teaches capturing a series of images as part of the three-dimensional verification of a face. *See* §XII.A.2.c (1[b]). Derakhshani also envisions a camera moving relative to a user's face. *Derakhshani*, 17:59-18:4; §XII.A.7 (claim 6). In my opinion, a POSITA would have understood that, to capture multiple images at multiple distances, either the camera would need to be moved from the first to the second distance, or the subject would need to be moved in relation to the camera from a first distance to a second distance because these are the only two ways to change the relative distance between the two. *See* §XII.A.2.c (1[b]).

662. Furthermore, Derakhshani discloses implementing the method using mobile computing devices that incorporate a camera such as "smart phone[s]." *Derakhshani*, 5:22-27. Based on these teachings, in my opinion, a POSITA would have understood that Derakhshani at least suggests moving a hand-held device with a camera in relation to the user's face because that is the more user-convenient option than moving the user's face relative to the camera, and because there are only two

possible options for changing the distance between the user's face and the camera (either moving the camera or the user). *See* §XII.A.7 (claim 6).

e. 8[d]: capturing, at the second distance, at least one second image of the user's face taken with the camera associated with the computing device;

663. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 8[d] for the reasons discussed in §XII.A.2.c (1[b]).

f. 8[e]: processing the at least one second image or a portion thereof to create second data;

664. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 8[e] for the reasons discussed in §XII.A.2.c (1[b]).

g. 8[f]: examining the first data and the second data to determine whether differences between the first data and the second data indicate an expected type of distorting change in at least one image that is consistent with a real person being imaged and which is indicative of three-dimensionality;

665. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 8[f]. *See* §XII.A.2.d (1[c], describing expected differences), XII.A.2.e (1[d], describing expected differences as distorting changes from Derakhshani's focus-distance approach and Derakhshani-Tanii's distance-induced distortion approach).

h. 8[g]: determining the user's face is not three-dimensional when the first data and the second data do not have expected differences indicating the user exhibits three-dimensionality.

666. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 8[g] for the reasons discussed in §XII.A.2.e (1[d]).

10. Claim 9

a. 9[a]: The method of claim 8 further comprising: capturing one or more additional images at distances from the user's face that are between the first distance and the second distance;

667. In my opinion, Derakhshani, alone or combined with Tanii, teaches 9[a]'s additional limitation.

668. Derakhshani discloses that, as part of the three-dimensional verification process, "a plurality" of images may be captured. *Derakhshani*, 16:44-46 (focal-distance embodiment), 17:45-47 (parallax embodiment).

669. In my opinion, a POSITA would have understood generally that capturing more images would be preferred, as it would provide the biometric-authentication system more samples to evaluate to authenticate the identity of the user. For instance, with the focus-distance approach, capturing only two images would be able to evaluate depth of facial landmarks on only two planes (or planar regions). Capturing additional images at additional focus distances would therefore provide additional information about the depth of facial landmarks on the face on additional focal planes. And for the parallax approach, capturing more than two

images would further ensure sufficient differences in perspective to evaluate whether parallax exists. For this reason, in my opinion, a POSITA would have known to take *at least* two images to prevent spoofing attacks, but more would have been preferred. The upper limit, in my opinion, would be dictated by the user resistance to the time of the authentication procedure.

670. Additionally, Tanii discloses that distance-induced distortions increase as distances between the face and camera decreases. *See Tanii*, [0048]. In my opinion, a POSITA would have understood that a relationship exists between the extent of distance-induced distortion and distance based on Tanii. Thus, for any set of captured images at different distances, the closest will display the most distance-induced distortion, the furthest will display the least, and any intermediate images captured at intermediate distances will display intermediate levels of distortion. For the same reasons discussed with respect to Derakhshani, in my opinion, a POSITA would have therefore been motivated to capture additional images at one or more distances between the first and second distance to determine whether the series of images at different distances exhibit an expected range of distance-induced distortion to further confirm the three-dimensionality of the user's face. Once again, a POSITA would have appreciated that capturing additional images increases accuracy, but comes at a cost of processing power and acquisition time.

b. 9[b]: for at least one of the one or more additional images, generating additional data;

671. In my opinion, Derakhshani, alone or combined with Tanii, teaches 9[b]'s additional limitation.

672. Derakhshani discloses processing the images to identify feature landmarks in each of the images. *See* §§XII.A.2.b (1[a]), XII.A.2.c (1[b]). In my opinion, a POSITA would have understood that each image would be processed to generate biometric data for that image under either the Derakhshani or Derakhshani-Tanii approaches so that the image could be used in a biometric-authentication process. *See* §XII.A.10.a (9[a]). If an image were captured and *not* processed, then it would not be useful for biometric authentication.

c. 9[c]: examining the additional data, the first data, and the second data, or portions thereof, to determine whether expected differences therebetween indicate the user's face exhibits three-dimensionality.

673. Derakhshani, alone or combined with Tanii, teaches 9[c]'s additional limitation.

674. Derakhshani discloses verifying the three-dimensionality of a face by evaluating expected differences between the facial landmarks (e.g., data) using either the focus-distance or parallax approach. §§XII.A.2.d-XII.A.2.e (1[c]-[d]). And Derakhshani and Tanii together teach verifying the three-dimensionality of a

face by evaluating for expected differences caused by distance-induced distortion. *See id.*

675. In my opinion, a POSITA would have understood that, when capturing additional images under any approach, facial landmarks would be identified and compared to the first and second biometric data as well to determine whether expected differences exist. *See* §§XII.A.10.a (9[a]; describing tradeoffs of accuracy versus processing power and time); XII.A.10.b (9[b]; describing processing of all captured images for use in biometric authentication).

11. Claim 10: The method of claim 8 further comprising displaying one or more prompts on a screen associated with the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

676. Derakhshani, alone or in combination with Tanii, teaches claim 10's additional limitation.

677. Derakhshani discloses displaying prompts to guide the user to capture images for authentication, *Derakhshani*, 5:23-32, 6:8-16, 9:22-26, including at more than once distance, *id.*, 17:64-66; §§XII.A.2.c (1[b]), XII.A.5 (cl.4).

678. But even if Derakhshani does not expressly disclose taking two images at different distances, in my opinion, a POSITA would have been motivated to do so in view of Tanii to identify distance-induced distortions that indicate depth of a three-dimensional face. §§XII.A.1 (motivation); XII.A.2.c (1[b]). Moreover, when

modifying Derakhshani to look for distance-induced distortions by capturing images at different distances consistent with Tanii, in my opinion, a POSITA would have been motivated to provide prompts to a user to ensure the images are captured at the correct distances because Derakhshani already discloses providing prompts to correctly orient the user relative to the camera.

12. Claim 12: The method of claim 8 wherein the computing device is a hand-held device, and the user holds the computing device at the first distance from the user's face when capturing at least one first image and holds the computing device at the second distance from the user's face when capturing the at least one second image.

679. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 12's additional limitation. §XII.A.7 (cl.6).

680. In my opinion, a POSITA also would have understood that, when adjusting the distance of a hand-held computing device, the device would move between a first distance from the user's face and a second distance from the user's face because Derakhshani and Tanii are both concerned with differences in images of a user's face captured in different conditions, such as distances. *See* §XII.A.2.c (1[b]).

13. Claim 13: The method of claim 8 wherein the first data and the second data comprise at least in part biometric data.

681. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 13's additional limitation. *See* §§XII.A.2.b-XII.A.2.c (1[a]-1[b]).

682. In my opinion, a POSITA would have also understood that data created from an image (or portion of an image) of a user's face would comprise biometric data. *Id.*

14. Claim 14: The method of claim 8 wherein moving the camera comprises moving the camera linearly toward or away from the user's face.

683. In my opinion, Derakhshani, alone or combined with Tanii teaches claim 14's additional limitation.

684. Derakhshani discloses that "[a] plurality of images [are] taken from different perspectives on the subject," *Derakhshani*, 17:45-18:4, including a camera that is "slid" or "rotated." *See* §XII.A.2.c (1[b]). In my opinion, a POSITA would have understood that, when Derakhshani discloses the camera being "slid," that includes a linear translation towards or away from the user's face to capture images on different focal planes when following the focus-distance approach using a fixed-focus-distance camera, or to change perspectives for the parallax effect. §XII.A.2.c (1[b]).

685. Relatedly, Tanii teaches that images captured at different distances exhibit different degrees of distortion. *Tanii*, [0047]-[0048], [0056], Figs. 3A-B, 4A-B. In my opinion, a POSITA would have understood that Tanii depicts a *linear* movement of the camera towards or away from the user's face because the face shown in Figures 3A-B and 4A-B are from the same face-on perspective, but with

different degrees of distortion. In my opinion, therefore, a POSITA would have understood that, when modifying Derakhshani in view of Tanii to look for distance-induced distortions, the camera would be moved linearly towards or away from the user's face between image captures. §XII.A.2.c (1[b]).

15. Claim 15: The method of claim 8 further comprising illuminate a screen of the computing device while capturing the at least one first image and/or the at least one second image to improve quality of an image being captured.

686. In my opinion, Derakhshani teaches claim 15's additional limitation.

687. Derakhshani discloses, in addition to a "spatial metric," calculating a separate "reflectance metric" that measures changes in surface glare on the eye due to changes in a light source such as the illumination of the screen to further verify the "liveness" of the user's face. *Derakhshani*, 18:8-19. In my opinion, a POSITA would have therefore understood that illuminating the screen while capturing the first and second images would improve the quality of the image—at least for authentication purposes—by inducing a glare in the eyes of the user.

688. However, a POSITA would have also understood more generally that using a "flash" of light during image capture was a well-known and conventional way to improve image quality generally, and particularly in dim-lit environments. In fact, the use of flash is nearly as old as the camera itself, with the use of flash powders and lamps in the late 1800s. In my opinion, therefore, a POSITA would have found it further obvious to use a "flash" of light by illuminating the screen of the device to

improve the overall image quality, which was also well-known. *See, e.g.*, §VII.E (Suzuki).

16. Claim 16: The method of claim 8 wherein a face of the user is held steady when capturing the at least one first image and the at least one second image and the camera moves from the first location to the second location.

689. Derakhshani, alone or combined with Tanii, teaches claim 16's additional limitation.

690. Derakhshani and Tanii both teach or suggest moving the camera to capture images at two different distances. *See* §§XII.A.2.c (1[b]), XII.A.14 (cl.14).

691. In my opinion, a POSITA would have understood that, when moving the camera to capture images from different distances, the user's face would be stationary (e.g., steady) both during each image capture (to ensure each image is not blurry due to the camera's exposure time) and during movement of the camera (to isolate any differences between images to those attributable to the change of distance). Moreover, a POSITA would have appreciated that holding the user's face steady and moving the camera closer and further away would be more user friendly than requiring the user to move their head closer and further from the camera while holding the camera steady.

17. Independent Claim 17

a. 17[pre]: A method, performed using a computing device, for providing authentication of a person during an authentication session, the method comprising:

692. If the preamble is limiting, Derakhshani discloses or suggests it.

693. First, it is worth noting that although the ’938 Patent uses the term “authentication” in other claims (such as claim 5, which recites an “authentication session”) to refer to comparisons to enrollment data to authenticate a user’s identity, claim 17 uses the term “authentication” to refer to authenticating the three-dimensionality of the face, which is made clear in limitation 17[g] and claim 19. Accordingly, in my opinion, the ’938 Patent does not use the term “authentication” consistently to refer to comparisons to enrollment data.

694. Regardless, Derakhshani discloses a computer-implemented authentication method for authenticating the three-dimensionality of a person’s face during an authentication session. *See, e.g.*, *Derakhshani*, Abstract, 1:11-2:3; §§XII.A.2.a (1[pre]), XII.A.6 (cl.5).

b. 17[a]: capturing a first image of a head of the person with a camera at a first distance from the person, the camera associated with the computing device;

695. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 17[a]. *See* §XII.A.2.b (1[a]).

696. Furthermore, in my opinion, a POSITA would have understood that, when capturing an image of a user's face, the image would be of the user's head because the face is part of the head.

c. 17[b]: changing a distance between the person and the camera to a second distance, which is different from the first distance;

697. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 17[b] for the reasons discussed in §§XII.A.2.c (1[b]), XII.A.9.d (8[c]).

d. 17[c]: capturing a second image of the head of the person with the camera at the second distance from the person;

698. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 17[c] for the reasons discussed in §§XII.A.2.c (1[b]), XII.A.17.b (17[a]).

e. 17[d]: comparing one or more aspects of the head from the first image or first biometric data derived from the first image to one or more aspects of the head from the second image or second biometric data derived from the second image to determine whether expected differences are not present, wherein the expected differences:

699. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 17[d]. *See* §XII.A.2.d (1[c]).

700. In my opinion, a POSITA would have also understood that, when comparing facial landmarks consistent with Derakhshani, that would be a

comparison of "aspects of the head," because facial landmarks are present on the user's head.

f. 17[e]: would be present when the first image and second images of the head of the person being captured at different distances has three-dimensional characteristics but not if the head did not have three-dimensional characteristics; and

701. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 17[e] for the reasons discussed in §§XII.A.1 (motivation), XII.A.2.e (1[d]), XII.A.17.e (17[d]).

g. 17[f]: the expected differences result from differences in relative dimensions of a person's face appearing different when capturing images is done close to the person's face and far from the persons face; and

702. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 17[f] for the reasons discussed in §§XII.A.1 (motivation), XII.A.2.e (1[d]), XII.A.17.e (17[d]).

703. Furthermore, in my opinion, a POSITA also would have understood that each of the approaches disclosed by Derakhshani (focus distance or parallax), or taught by Derakhshani combined with Tanii (distance-induced distortion), include an evaluation for "expected differences" in the relative dimensions of a person's face when one image is captured close to the person's face and another image is captured far from the person's face. *See* §§XII.A.1 (motivation), XII.A.2.c-XII.A.2.e (1[b]-1[d]). Specifically, Derakhshani's focus-distance approach (using a fixed-focus-

distance camera) looks for features that are blurry and those that are clear in both a close and far image to understand the relative depth of those features. *See* §§XII.A.1 (motivation), XII.A.2.e (1[d]). Derakhshani's parallax approach looks for changes in relative displacement of facial features between a close and far image. *Id.* And finally, the Derakhshani-Tanii focus-distance approach looks for different degrees of distance-induced distortion—which alters the relative dimensions of facial features—between a close and far image. *Id.*

h. 17[g]: if the expected differences are not present, denying authentication of the person and providing notice thereof to one or more of the person, a third party, or a software application, wherein the authentication is authentication of liveness, three-dimensionality, or both.

704. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 17[d] for the reasons discussed in §XII.A.2.e (1[d]).

705. Derakhshani also discloses that when an authentication attempt is rejected as a spoof attempt, the system provides notice of the rejection to the user or a third party, and that the authentication is one of at least liveliness and/or three-dimensionality. *Id.*, 4:53-63; 8:67-9:4; 11:17-26.

18. Claim 18: The method of claim 17 wherein the steps of comparing, denying authentication, and providing notice are performed by a server that is remote from the computing device.

706. In my opinion, Derakhshani discloses or suggests claim 18's additional limitation.

707. Derakhshani discloses the biometric-authentication process can be performed by a server system that is remote from the computing device, including the comparing, denying authentication, and providing notice steps. *Derakhshani*, 9:27-30; 10:1-24, 11:22-26.

19. Claim 19: The method of claim 17 wherein the authentication is authentication of three-dimensionality.

708. Derakhshani discloses or suggests claim 19's additional limitation.

709. Derakhshani discloses an authentication process that includes verifying the three-dimensionality of a user. *Derakhshani*, 9:59-67; 16:44-18:4.

20. Independent Claim 20

a. 20[pre]: A method for determining whether a user exhibits three-dimensionality, the method comprising:

710. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 20[pre] for the reasons discussed in §§XII.A.2.a (1[pre]), XII.A.9.a (8[pre]).

b. 20[a]: capturing at least one first image of a user's face taken with a camera located a first distance from the user, the camera associated with a computing device;

711. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 20[a] for the reasons discussed in §XII.A.2.b (1[a]).

c. 20[b]: processing the at least one first image or a portion thereof to create first data, the first data derived from the user's face;

712. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 20[b] for the reasons discussed in §XII.A.2.b (1[a]).

d. 20[c]: intentionally moving the camera from the first location to a second location, the second location being a second distance from the user, or the user moving to change a distance between the user and the camera from the first distance to the second distance;

713. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 20[c]. *See* §§XII.A.2.c (1[b]), XII.A.9.d (8[c]).

714. In my opinion, a POSITA would have also understood that any movement of the camera to evaluate focus distance or parallax effects (Derakhshani alone) or distance-induced distortion effects (Derakhshani and Tanii) would include intentional movement of the camera to induce these effects. Derakhshani, for instance, expressly notes that the "sensor is moved about the subject to collect image data from different orientations relative to the subject," including by "rotat[ing] or slid[ing] the camera." *Derakhshani*, 17:59-64. Tanii likewise explains that the extent

of distance-induced distortion depends on the distance between the user and the camera (*Tanii*, [0047]), and thus modifying Derakhshani to evaluate distance-induced distortion would require intentional movement of the camera between image captures. Although Derakhshani separately discloses that the parallax effect may also be measured using minor "involuntary" movements, (*Derakhshani*, 18:1-4), in my opinion, Derakhshani is merely providing one example of the *minimal* movement needed to capture a parallax effect. A POSITA would have known, however, that the smaller the displacement of the camera from one perspective to another, the error rate based on the parallax effect will increase. Thus, a POSITA would have been motivated to capture images from different perspectives using *more* than involuntary haptic movement (e.g., by instead intentionally moving the camera) to increase the amount of parallax induced and thereby decrease the error rate.

e. 20[d]: capturing at least one second image of the user's face taken with the camera located a second distance from the user, the second distance being different than the first distance;

715. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 20[d] for the reasons discussed in §XII.A.2.c (1[b]).

f. 20[e]: processing the at least one second image or a portion thereof to create second data, the second data derived from the user's face;

716. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 20[e] for the reasons discussed in §XII.A.2.c (1[b]).

g. 20[f]: analyzing the first data to determine at least if the first data exhibits first characteristics that indicate the first data was derived from an image of the user captured at the first distance;

717. In my opinion, Derakhshani discloses or suggests limitation 20[f].

718. Derakhshani discloses that, as part of the verification process, "a landmark (e.g., an iris, an eye corner, a nose, an ear, or a background object) may be identified and located in the plurality of images." *Derakhshani*, 16:44-54 (focus-distance approach), 17:45-64 (parallax approach).

719. In my opinion, a POSITA would have understood that, when analyzing the images for facial landmarks, the landmarks would exhibit characteristics (e.g., size, distance from other landmarks, or other features) that indicates the image was taken at some distance, because—when all else is held equal, such as the camera system, lenses, and focus settings used—those types of characteristics of the landmarks depend on the distance between the user and the camera. *See, e.g.*, Ex-1018, 107 (Fig. 3.9 providing an example of how distance from the camera affects resolution of facial features).

720. Moreover, a POSITA would have understood that, when evaluating images for specific types of distortion (like focus-distance blurring disclosed by Derakhshani or distance-induced distortion taught by Zhang combined with Tanii), it is especially important for the three-dimensional verification process to be able to orient which image is taken at which distance for the focus-distance and distance-

induced distortion approaches in order to assess the three-dimensionality of the face. For instance, in the focus-distance approach, if one image shows a clear nose but all other facial features are blurry, it is important to know that the image is associated with the shortest focus distance used (for adjustable focus-distance cameras) or was taken furthest from the user (for fixed-focus distance cameras) to indicate the nose is the most forwardly protruding feature from the face. And for the distance-induced distortion approach, it is important to characterize which image is the "close" image (e.g., Tanii's Fig. 4B) to assess whether it specifically exhibits expected degrees of distance-induced distortion.

h. 20[g]: analyzing the second data to determine at least if the second data exhibits second characteristics that indicate the second data was derived from an image the user captured at the second distance, wherein the first characteristics or the second characteristics include at least distortion within the at least one first image or the at least one second image;

721. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 20[g] for the reasons discussed in §XII.A.20.g (20[f]).

722. Moreover, when applying Derakhshani's focus-distance approach or the Derakhshani-Tanii distance-induced distortion approach, a POSITA would have understood that the images are analyzed to identify specific distortion characteristics (e.g., blurriness/clearness for focus-distance; barrel or fish-eye distortions for distance-induced distortion). §§XII.A.1 (motivation), XII.A.2.e (1[d]).

i. 20[h]: determining the user does not exhibit the expected degree of three-dimensionality when either or both of the following occur:

723. In my opinion, Derakhshani discloses or suggests limitation 20[h].

724. As explained in further detail below, Derakhshani determines that the user's face is not three-dimensional when there is a match between the biometric landmarks and no expected differences exist. *Derakhshani*, 16:44-18:4.

j. 20[h1]: the step of analyzing the first data determines the first data does not exhibit first characteristics that indicate the first data was derived from an image of the user captured at the first distance; or

725. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 20[h1].

726. Derakhshani discloses that a face is determined to not be three-dimensional when expected differences do not exist in the facial landmarks (e.g., biometric data) of the first and second images. *See* §XII.A.2.e (1[d]).

727. In my opinion, a POSITA would have understood that, if neither the first nor second images display any expected differences—whether looking for Derakhshani's focus-distance or the distance-induced distortion taught by Derakhshani in view of Tanii—then a determination is made that either the first or second image was not taken at a closer (e.g., first or second) distance from the face, because such distortions would be expected to be induced in at least one image. In other words, a POSITA would have appreciated that if the "first" image is intended

to be the “closer” image, then a lack of distortions (or characteristics, as claimed) in the “first” data would indicate the first data was not derived from an image of the user captured at the first (closer) distance.

k. 20[h2]: the step of analyzing the second data determines the second data does not exhibit second characteristics that indicate the second data was derived from an image of the user captured at the second distance.

728. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 20[h2]. *See* XII.A.20.j (20[h1]).

729. Similarly, a POSITA would have appreciated that if the “second” image is intended to be the “closer” image, then a lack of distortions (or characteristics, as claimed) in the “second” data would indicate the first data was not derived from an image of the user captured at the first (closer) distance.

21. Claim 21: The method of claim 20 further comprising displaying one or more prompts on a screen associated with the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

730. In my opinion, Derakhshani discloses or suggests claim 21’s additional limitation for the reasons discussed in §XII.A.11 (cl.10).

22. Claim 22: The method of claim 20 wherein the at least one first image and the at least one second image are captured with a hand-held computing device, and the user holds the computing device at the first distance when capturing at least one first image and at the second distances when capturing the at least one second image.

731. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 22's additional limitation for the reasons discussed in §XII.A.12 (cl.12).

23. Claim 23: The method of claim 20 wherein the first data and the second data comprise at least in part biometric data.

732. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 23's additional limitation for the reasons discussed in §XII.A.13 (cl.13).

24. Claim 24: The method of claim 20 wherein the first data and the second data comprise at least in part image data of facial features.

733. In my opinion, Derakhshani discloses or suggests claim 24's additional limitation for the reasons discussed in §XII.A.8 (cl.7).

B. Ground 1B: Derakhshani, Tanii, and Takh (Claim 11)

1. Motivation to Combine

734. In my opinion, a POSITA would have been motivated to modify Derakhshani, with or without Tanii, in view of Tahk because Tahk provides a user-friendly way of ensuring that a face presented for facial authentication is properly

framed. In my opinion, a POSITA would have understood that providing a real-time preview of what an image would look like prior to capturing the image, as well as providing express prompts—such as written instructions or oval shapes on the live-preview screen—as taught by Tahk, §VII.D (Tahk), would have been particularly useful for authentication procedures that require capturing multiple images of a face, as taught by Derakhshani and Tanii.

735. For instance, providing a user a real-time preview would allow the user to actively adjust the position of the camera and/or their orientation of their face to properly frame their face for the image capture. Moreover, providing express prompts and oval shapes sized to guide the user to properly position their face would ensure images best suited for facial recognition can be captured, and that the faces would be captured from sufficiently different perspectives and/or distances to ensure the three-dimensional verification taught by Derakhshani, alone or in combination with Tanii, could be performed. This is particularly important for facial-recognition systems, which generally are known to have difficulty matching faces across different views. *See, e.g.*, Ex-1018, 32.

736. Moreover, a POSITA would have known that providing real-time image feedback, written instructions, and oval shapes to frame a face during a facial authentication process were all well-known and conventional techniques to provide user feedback during image capture as of the time of the invention. *See, e.g.*, Ex-

1034, 7:16-8:7, Figs. 6B-7C (providing "an example of an interface used upon registering a facial image of a person to be authenticated" in which "the image of this person is displayed on a monitor" during registration, and oval-shaped prompts to indicate the size/distance of the user from the face authentication sensor); Ex-1035, 5:31-32 ("The computing device may present prompts that instruct the user to perform one or more liveness gestures"), 6:3-4 (same).

2. Claim 11: The method of claim 10 wherein the one or more prompts are an on the screen shape within which an image of a face of the user is aligned during capture the at least one first image and the at least one second image.

737. In my opinion, Derakhshani combined with Tanii and Tahk teaches claim 11.

738. As I have previously explained, Derakhshani, alone or in combination with Tanii, teaches providing prompts to user to properly frame themselves at different distances to capture images for biometric authentication. §XII.A.11 (claim 10). But Derakhshani and Tanii do not expressly describe using shaped prompts to guide a user during the facial-authentication process.

739. In my opinion, however, a POSITA would have been motivated to provide such shaped prompts (and specifically oval-shaped prompts) in view of Tahk. *See, e.g.*, *Tahk*, Figs. 8A-B (presenting an oval-shaped prompt to frame the face at the correct distance). A POSITA would have been motivated to modify Derakhshani, alone or in combination with Tanii, to provide such oval-shaped

prompts because they are a natural shape to appropriately size and frame a face at different distances. *See* §VII.D (Takh); XII.B.1 (motivation).

C. Ground 2A: Zhang and Tanii (Claims 1-3, 5-9, 12-14, 16-20, 22-24)

1. Motivation to Combine

740. In my opinion, a POSITA would have been motivated to combine Zhang and Tanii because both concern identifying and accounting for the three-dimensional nature of a face when capturing an image. They differ, however, in what principles are used to account for the face's three-dimensionality. Zhang, for instance, looks to dissimilarities in two images after one undergoes a mathematical homography. *See* §VII.C (Zhang). And although Tanii is not expressly directed to *evaluating* whether a face has depth like Zhang, Tanii exemplifies the well-known distortions caused by the interaction between the camera's lens and the three-dimensional nature of the face, *see* §VII.B (Tanii). A POSITA would have appreciated, therefore, that Tanii recognizes another clear alternative to evaluating the depth of a face, consistent with Zhang's existing homography transformation.

741. A POSITA would have recognized, as Tanii does, that distance-induced distortions occur because of the interactions between the shape of the camera lens and shape of the face, and the distortion in part depends on the distance between the face and the camera. §VII.B (Tanii); *Tanii*, [0048]. Accordingly, a POSITA would have understood from Tanii that, by taking two images from two different distances,

a larger amount of distortion in the closer of the two images indicates whether a face is three-dimensional or not.

742. In my opinion, a POSITA would have therefore appreciated from Tanii that images captured by Zhang—without any modification—may exhibit distance-induced distortions based on the particular camera used to perform Zhang's process (e.g., particularly when a wide-angle lens with significant barrel distortion is used, as is common in computers and mobile devices). However, a POSITA would have also appreciated that any distance-induced distortions would further enhance Zhang's homography-transformation process because a homography transformation cannot correct for these distortions.

743. For instance, if a homography transformation were applied to Tanii's Figure 4B (serving as Zhang's "first image") to compare to Figure 3B (serving as Zhang's "second image"), the transformation would not account for differences between the images caused by the distance-induced distortion.

Fig.4B

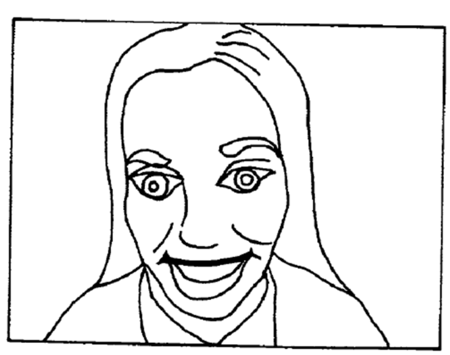
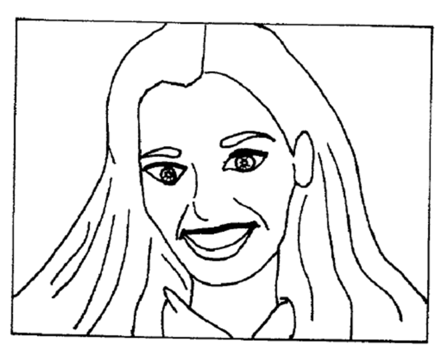


Fig.3B



Tanii, Figs. 3B, 4B. That is because Zhang relies on a mathematical principle that enables transforming the *perspective* of a planar object, such as a photograph being used to spoof the authentication procedure to a different *perspective*, §VII.C (Zhang), whereas the distortion identified by Tanii is *radial* and a byproduct of the lens' imperfections and the change in magnification with distance. A homography transform does not account for such radial distortions, but would instead transform the perspective of Tanii with its distortions intact. In other words, in a transformation of perspective with a three-dimensional object such as a real face, Tanii's distance-induced distortions would remain. Ultimately, however, when comparing the two images once one is transformed into the perspective of the other, there would remain differences attributable to the distance-induced distortion which, in my opinion, a POSITA would have understood would result in Zhang identifying the face as three-dimensional.

744. In my opinion, a POSITA would have therefore recognized that Zhang's existing process would be *enhanced* by prompting a user to capture two images and two distances—one of which would have increased distance-induced distortion—because if the face were three-dimensional, Zhang's existing procedure would identify the two images as different and indicate a three-dimensional face. The lack of a match between the two images would likely be enhanced by changes

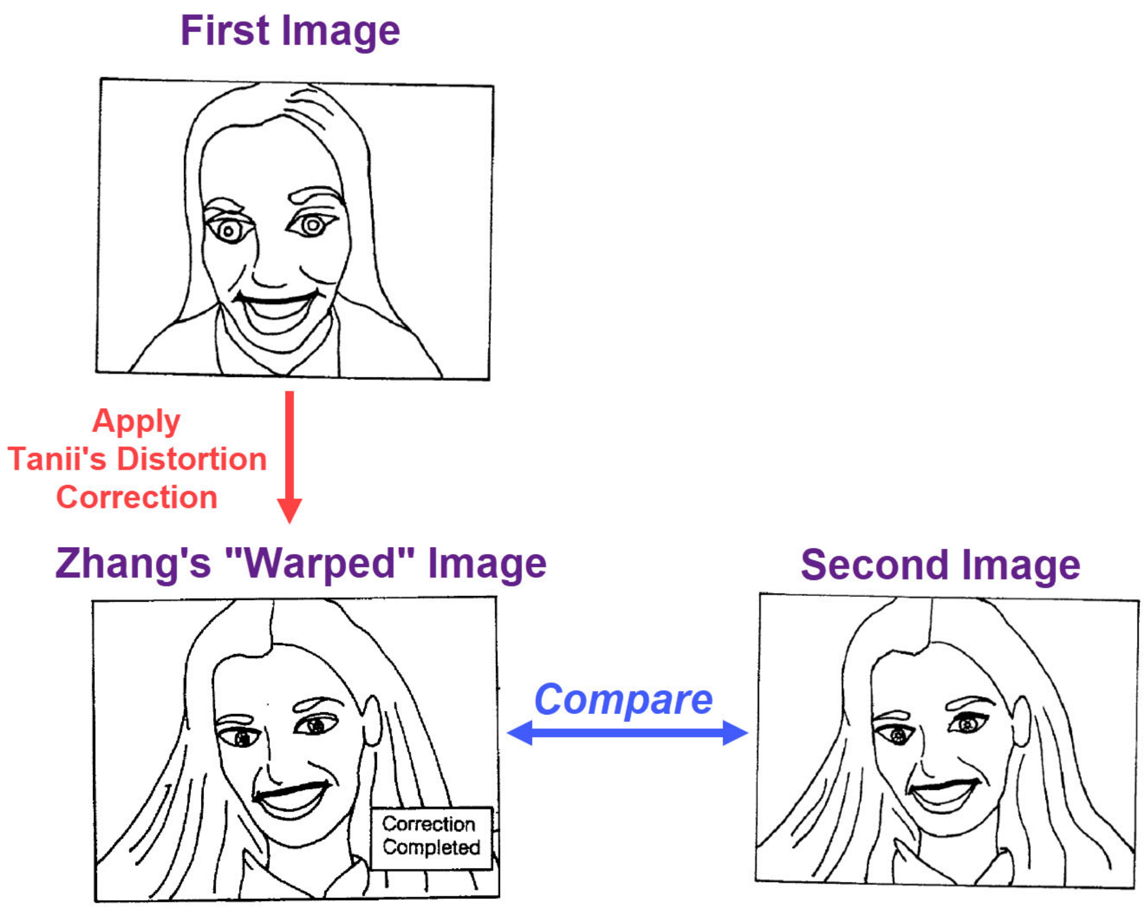
in radial distortion: it makes them even less like data from two planar objects which would produce a match.

745. However, in my opinion, a POSITA would have also been motivated to modify Zhang's process in view of Tanii in either of two additional ways.

746. First, in my opinion, a POSITA would have been motivated to follow Zhang to verify the three dimensionality of a face during a facial authentication procedure by taking two or more images, applying mathematics to one of the images, and comparing the mathematically altered image to a second (unaltered) image. But instead of the mathematics applied being a homography transformation, in my opinion, a POSITA would have been motivated to *substitute* Zhang's mathematics for those taught by Tanii to correct for distance-induced distortion. In other words, rather than change the perspective of one image to match the second image, a POSITA would correct the distortion of one image (to create what Zhang refers to as its "warped" image[6]) and compare the result to another image taken further away

[6] Zhang and Tanii both use the term "warped" to refer to different effects, but they are not inconsistent with one another. Specifically, Zhang uses the term "warped" to refer to the resulting image that has undergone homography transformation because the original relationship between the pixels in the image are modified. Tanii uses the term "warped" to refer to the distortions in an image of a face induced by the image-

that does not exhibit the same degree of distance-induced distortions.



Tanii, Figs. 3B, 4B, 9.

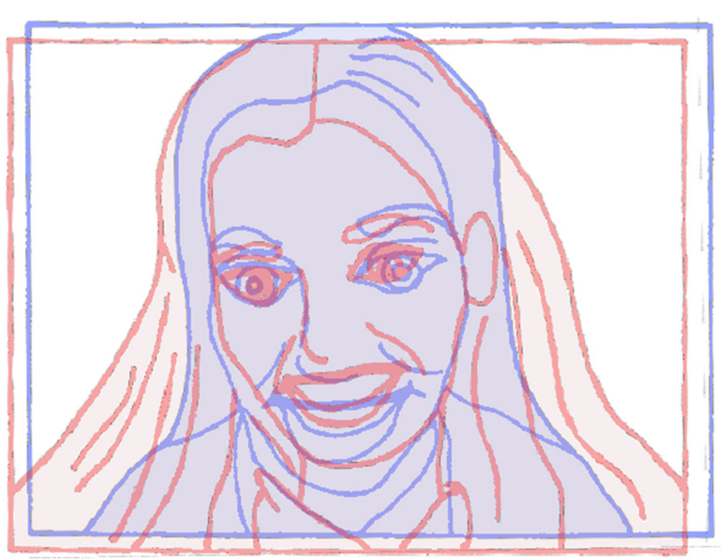
747. A POSITA would have appreciated that if the "warped" (distortion-corrected) image and second image are sufficiently similar, that indicates a three-

---

capture conditions (e.g., distance and lens geometry). When I refer to Zhang's "warping," I am referring to the result of a mathematical application to an image; and when I refer to Tanii's warping, I am referring to distance-induced distortion.

dimensional face because Tanii is correcting for distortions attributable the three-dimensionality of the user's face. By following this approach, a POSITA would have recognized that the only difference (besides the mathematics) is that the comparison between the Zhang-Tanii "warped" (distortion-corrected) image would look for a match with the second image.

748. Alternatively, a POSITA would have appreciated that Zhang and Tanii could be modified to eliminate the mathematical transformation of a first image entirely. Once again, a POSITA would have been motivated to follow Zhang to verify the three dimensionality of a face during a facial authentication procedure by taking two or more images, but rather than apply mathematics to "warp" one of the images (e.g., using either a homography transform or distortion-correction procedure), the facial features would be mapped in each image, matched between the two images, and evaluated to determine whether differences attributable to distance-induced distortion appear (e.g., does the shape of the nose, size of the mouth or forehead, or do facial features shift by expected degrees relative to one another?). For instance, I have overlayed Tanii's two images to show how one (in blue) exhibits expected distortions while the other (in red) does not, resulting in various misalignments in facial features (assuming the faces are normalized in size):



In such circumstances, a POSITA would understand that two images would still be required, rather than just evaluating one image for distance-induced distortion. Otherwise, an imposter could provide a picture of a user with distance-induced distortion already applied to spoof the system; the need for a more-distance, undistorted image of the user for comparison would still be required.

749. In my opinion, a POSITA would have been motivated to make either of these two modifications for two reasons. First, a POSITA would have appreciated that Zhang's homography-transformation process may be spoofed by presenting a non-planar picture of a face, because Zhang's homography transformation process specifically looks for a *planar* structure in images that can be transformed nearly identically from one perspective to another. Thus, Zhang's system could possibly be subverted by bending the picture in a way to trick the system, or applying the picture to a three-dimensional shape. For this reason, in my opinion, a POSITA would have

therefore been motivated to look for other methods to ensure the user's face is from the user, and not a spoofer. A POSITA would have also appreciated that distance-induced distortion is more difficult to spoof, because it is induced by the interactions of geometries between the user's face and the camera's lens, and therefore could not be circumvented as easily. Second, a POSITA would have appreciated that either of the processes suggested by Tanii offers a potentially less computationally demanding than the homography mathematics proposed by Zhang, which may be more suitable for a low-power portable device.

750. In my opinion, a POSITA would have had a reasonable expectation of success in making this modification to Zhang because Tanii already taught a mechanism to identify (and correct) distance-induced distortions, *see, e.g.*, *Tanii*, [0056], and it was already well-known to use depth information about a face derived from a series of images to distinguish between live faces and two-dimensional images of faces. *See, e.g.*, Ex-1014, Abstract, [0031], [0036].

2. Independent Claim 1

a. 1[pre]: A non-transient computer readable medium containing non-transitory machine executable code configured to determine if the three-dimensional shape is consistent with that of a human face, the non-transitory machine executable code configured to:

751. If the preamble is limiting, in my opinion, Zhang discloses or suggests it.

752. Zhang discloses a computing device, *Zhang*, [0013], that can run "software, with instructions being executed" that can be stored in "computer readable media," *id.*, [0066]-[0071], Fig. 6. By disclosing a computer-readable media with instructions, in my opinion, a POSITA would have understood that Zhang is referring to conventional "non-transient" (e.g., non-temporary) computer-readable medium with "non-transient" (e.g., non-temporary, stored) executable code to perform Zhang's process.

753. Zhang also discloses computer-based methods "to determine whether a face in multiple images is a 3D structure or a flat surface," *Zhang*, [0013], [0026], Figs 2-3; *see also, e.g.*, *id.*, Abstract, [0003], to "authenticate a user for particular access," *id.*, [0012].

b. 1[a]: receive or derive first biometric data from at least one first image of a user taken with a computing device camera located at a first distance from the user;

754. In my opinion, Zhang discloses or suggests limitation 1[a].

755. Zhang discloses capturing a first image of a user as part of the authentication method. *Zhang*, [0016] ("user 108 presents himself or herself to image capture component 102, allowing component 102 to capture images 106 of user 108."), [0021].

756. In my opinion, a POSITA would have understood that Zhang's process captures an image at a first distance between the user and image capture component

102 in order to capture a picture of the user's face. *Zhang*, [0016]. Specifically, a POSITA would have understood that there *must* be some distance between the camera and the face to capture the first image so that the camera's field of view encompasses the face. If there were no distance between the camera and the face (e.g., if the camera were pressed up against the user's skin), then the field of view would be limited to just that patch of skin and the face would not be captured in the image, which would be useless for the three-dimensional verification process. In my opinion, a POSITA would have also understood that the "image capture component 102" would be a camera, because cameras are conventionally used to capture images, especially in computing devices. In fact, the "CCDs" and "CMOS" sensors Zhang references are the types of sensors commonly used in cameras. *Zhang*, [0016]; *see also, e.g.*, *Suzuki*, [0019] ("The camera unit includes solid-state image pickup elements such as CCD or CMOS"); Ex-1028, 3 ("Presently, there are two main technologies that can be used for the image sensor ***in a camera***, i.e., CCD (Charge-coupled Device) and CMOS (Complementary Metal-oxide Semiconductor).")

757. Zhang also discloses processing the first image to extract "feature points" from the image. *Zhang*, [0027] ("[O]ne or more feature points are extracted from two images… A variety of different feature points can be extracted, such as a corner of an eye, a corner of a mouth, a tip of a nose, and so forth."), [0026] (disclosing "software, firmware, hardwire, or combin[ed]" implementations).

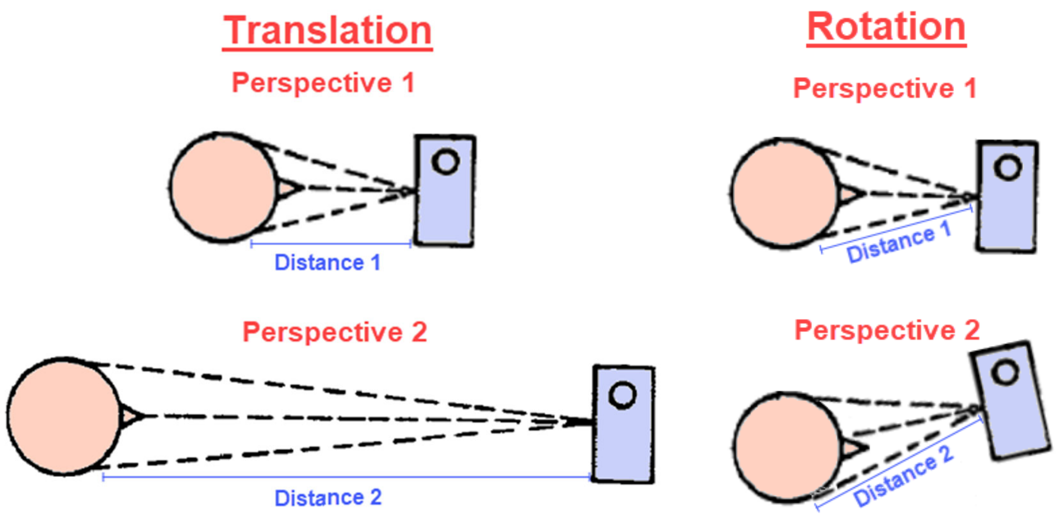
758. In my opinion, a POSITA would have understood that Zhang's extracted feature points constitute deriving "biometric data" because "biometric data" generally refers to unique physical characteristics of an individual, which would include the positions of "feature points" such as a user's eyes, nose, mouth, and other such features. *See, e.g.*, Ex-1018, 2 ("biometric recognition can be defined as the science of establishing the identity of an individual based on the physical and/or behavioral characteristics of the person."), 100-103 (describing the types of biometric data about a face used for facial-recognition systems, "such as the structure of the face components (e.g., eyes), [and] the relationship between facial components").

c. 1[b]: receive or derive second biometric data from at least one second image of the user taken with the computing device camera located at a second distance from the user, the second distance being different than the first distance;

759. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 1[b].

760. Zhang discloses capturing a second image of a user as part of the authentication method. *Zhang*, [0016] ("Image capture component 102 captures multiple images"). Zhang also discloses processing the second image to obtain second feature-point biometric data from the image. *Zhang*, [0026]-[0027]; §XII.C.2.b (1[a]).

761. Zhang does not expressly disclose that the second image is captured at a second distance different from the first distance of the first image. But, in my opinion, a POSITA would have understood that Zhang at least implicitly requires *some* change of distance. §XII.C.1 (motivation). Specifically, Zhang discloses a "3D structure determination module 112" that uses a "homography" technique to distinguish between a real face and a picture of a face by, *inter alia*, transforming a first image to the perspective of a second image and comparing the two. *Zhang*, [0024], [0026]-[0035]; §VII.C (Zhang). In my opinion, a POSITA would have understood from Zhang that—like Derakhshani's parallax approach—the distances between the camera and at least some facial landmarks would change in order to obtain an image from a different perspective than the first and would obviously also encompass changing the overall distance between the camera and face as well. *See, e.g.*, §XII.A.2.c (in the context of Derakhshani, discussing changes of distance for parallax).



Moreover, a POSITA would have not only understood that providing images at different distances allows for a greater understanding of depth between objects in the scene, as exemplified in the paper Zhang references; Ex-1013, 22-25, but that taking pictures at different distances may induce distance-based distortion that would *enhance* the accuracy of Zhang's homography transformation to detect a three-dimensional face. §XII.C.1 (motivation).

762. Even if Zhang cannot be considered to disclose or suggest taking two images at different distances, however, a POSITA would have been motivated to do so in view of other prior art. For instance, a POSITA would have understood that distortions caused by camera lenses can indicate depth in the object being captured, as exemplified by Tanii. §XII.C.1 (motivation). Thus, even if Zhang does not already disclose this limitation, a POSITA would have been motivated to modify Zhang in view of Tanii to capture a second image at a second distance and evaluating the

images for different degrees of distance-induced distortions to distinguish between live, three-dimensional faces and two-dimensional pictures of a face. §XII.C.1 (motivation).

d. 1[c]: compare the first biometric data with second biometric data for expected differences that result from characteristics of a human face and the at least one first image and the at least one second image being captured at different distances from the user;

763. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 1[c].

764. Zhang discloses that, after calculating a homography matrix between the first and second image, a "warped" version of the first image is created and then compared to the second image to determine whether expected differences exist. *Zhang*, [0025], [0031]. Zhang also discloses that, as part of the comparison, "any of a variety of conventional face detection algorithms or face recognition algorithms can be used to detect the face within each image, and the selected locations are the locations that are part of a face within at least one of the warped and second images." *Zhang*, [0032].

765. In my opinion, a POSITA would have understood that Zhang discloses comparing a first biometric data (e.g., the facial-feature locations in the first warped image) and second biometric data (e.g., the facial-feature locations in the second image) to determine whether differences between the two exist, in which it would

be expected that a live face would have sufficient differences between the two images due to movement of the image capture component 102 (camera). And a POSITA would have appreciated that those expected differences between the two images would be attributable to "characteristics" of a human face—specifically, its three-dimensionality—and the fact that perspective (which includes changes in distance) changed between the two images, because it is this change of perspective that imparts the differences in focus of a three-dimensional face or changes in the relative displacement of facial features when Zhang is looking for a two-dimensional planar face.

766. However, a POSITA would have also been aware that differences between two images—one with more distance-induced distortions and one with less—can also be used to distinguish between live, three-dimensional faces, and two-dimensional pictures of a face, as exemplified by Tanii. §XII.C.1 (motivation). And, in my opinion, a POSITA would have been motivated to look for these expected distortions as either a supplemental or alternative verification of three-dimensionality of a face. *Id*. A POSITA would have appreciated that verifying the three-dimensional nature of the face using distance-induced distortion would be accomplished by matching the positions of biometric features across the first and second images, consistent with Zhang. When modifying Zhang to look for such distortions as an indication of three-dimensionality rather than perform a

homography transformation, *see* §XII.C.1 (motivation), the comparison would evaluate whether one of the images exhibits the distance-induced distortion that would be expected when the user's face is captured at a close distance to the camera, and the other image has less of the distance-induced distortion when captured further from the camera. *Id*.

e. 1[d]: determine that the three-dimensional shape is not exhibited when the second biometric data does not have expected differences compared to the first biometric data, the expected differences comprising at least differences due to the change in the relative distance between the user's facial features and the camera when the at least one first image was captured at the first distance and the at least one second image was captured at the second distance, wherein the expected differences result from fish-eye type distortion in at least one of the at least one first image and the at least one second image and due to the three-dimensional nature of the human face and the change in distance between the camera and the user.

767. In my opinion, Zhang combined with Tanii teaches limitation 1[i].

768. Zhang discloses that "if the image differences [between the first and second image after undergoing homography transformation] does not meet the threshold value, then the face in the first and second images is determined to be a flat surface and thus a picture of a face." *Zhang*, [0034]; *see also id.*, [0031]-[0033] (explaining comparison process to identify expected differences). Zhang does not expressly disclose that the expected difference is a "fish-eye" type distortion induced by changing the distance between the camera and face.

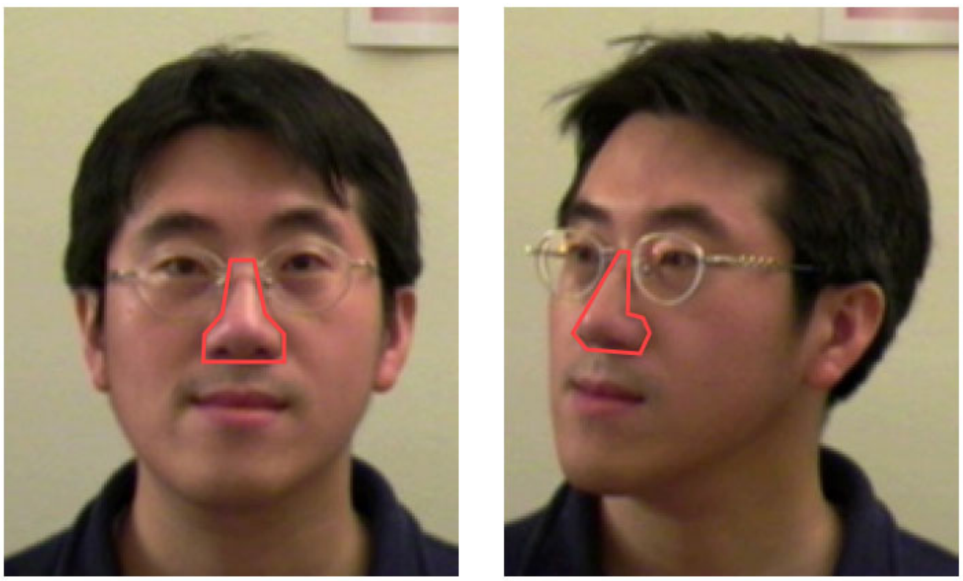
769. In my opinion, however, a POSITA would have understood that when modifying Zhang in view of Tanii, the difference being evaluated is whether facial features exhibit different degrees of distance-induced, "fish-eye" distortion that depends on the distance between the user and the camera in each image, because the existence of "fish-eye" distortion in a close image, but a lesser amount of distortion in a further image indicates the face has three-dimensional depth. *See* §XII.C.1 (motivation). For this reason, if the series of images do ***not*** contain different degrees of these types of expected, distance-induced "fish-eye" distortion when taken at different distances between the camera and face, a POSITA would have understood that it would be determined that the face is not three dimensional because only a three-dimensional face would exhibit these types of distance-induced distortions. §XII.C.1 (motivation).

3. Claim 2: The non-transient computer readable medium of claim 1 wherein the expected differences appear as changes in the relative size and shape of facial features of the user.

770. In my opinion, Zhang combined with Tanii teaches claim 2.

771. Zhang discloses that the system assumes that three three-dimensional faces will exhibit differences between two images after one undergoes homography transformation (because Zhang is looking for a planar image of a face where no differences exist after a homography transform). *Zhang*, [0031]-[0034]. Moreover, a POSITA would have understood generally that, when applying a homography

transformation to an image of a three-dimensional face from different perspectives, the transformation would induce some expected differences in the size and shape of the facial features as a result of the transformation. For instance, provided below are two images from different perspectives that show the shape of the nose changes based on the perspective. *See, e.g.*, Ex-1018, 99 (annotated).

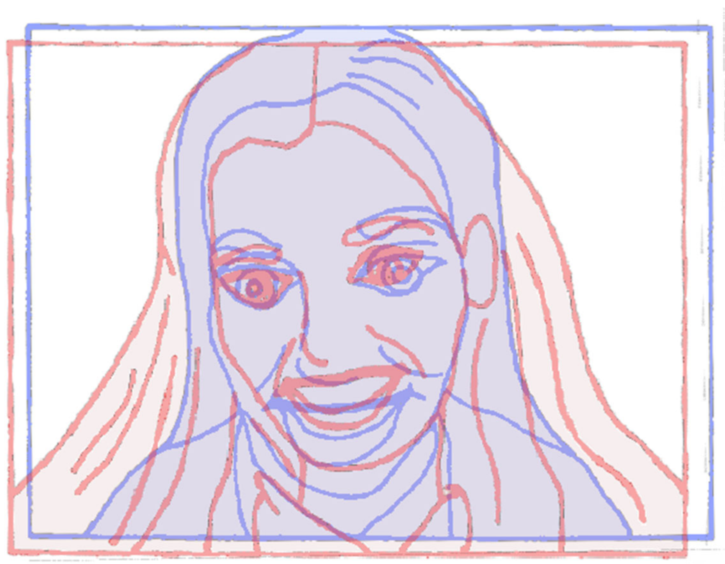


(a) (b)

When performing a homography transformation on, e.g., image (b), it would be unable to map and adequately transform features of the face that are obstructed, such as the left nostril and left ear. Any transformation of image (b) will therefore exhibit distortions of those features that would not match exhibit (a).

772. Furthermore, when modifying Zhang in view of Tanii to look for different degrees of distance-induced distortion, a POSITA would have appreciated that the expected differences in a face caused by distance-induced distortion is the

relative size and shape of facial features of the user. *Tanii*, [0047] ("where the main object 9 and the cellular phone 1 are close together…an unnatural image results in which the perspective is exaggerated."), [0056] (describing "a warp in which the peripheral areas of the main object appear reduced in size relative to the center area.").



In my opinion, a POSITA would have therefore known to look for these expected differences in size and shape of facial features to determine whether the face is three-dimensional.

4. Claim 3: The non-transient computer readable medium of claim 1 wherein determining that three-dimensionality is not exhibited happens during an authentication session.

773. In my opinion, Zhang discloses or suggests claim 3's additional limitation.

774. Zhang discloses the homography transformation process to verify the three-dimensionality of a user's face occurs during a facial authentication procedure, *Zhang*, [0012]-[0013], [0016], including determining that a face is not three-dimensional, *id.*, [0018].

5. Claim 5

a. 5[a]: The non-transient computer readable medium of claim 1 wherein the machine executable code is further configured to compare at least portions of the first data, second data, or both to enrollment data derived from an enrollment image, the enrollment image captured and stored prior to an authentication session; and

775. In my opinion, Zhang discloses or suggests 5[a]'s additional limitation.

776. Zhang discloses comparing at least portions of the first image, second image, or both to enrollment data captured and stored prior to the authentication session. *Zhang*, [0017] ("The authentication of user 108 can be performed…by comparing one or more of images 106 to previously captured images of user 108.").

777. In my opinion, a POSITA would have understood that Zhang's "previously captured images" would be taken during an enrollment session, as is conventional for biometric-authentication systems. *See* Ex-1018, 4-11 (providing overview of biometric authentication and verification).

b. 5[b]: determining the user is not authenticated when the first data, the second data, or both do not sufficiently correspond to the enrollment data.

778. In my opinion, Zhang discloses or suggests 5[b]'s additional limitation.

779. Although Zhang does not provide significant details about the overall authentication process—but instead states "a variety of different manners" can be used—Zhang's description of comparing biometric features to "previously captured images" is consistent with a conventional biometric-authentication procedure that requires a sufficient "match" above a threshold. *See Zhang*, [0017] ("The authentication of user 108 can be performed…by comparing one or more of images 106 to previously captured images of user 108."), [0038] (disclosing inter-picture matching); *see also, e.g.*, Ex-1018, 17 ("[B]iometric systems mostly decide on a person's identity based on a *close* match between the template and the query, where the strength of the match (or the degree of similarity) is represented by the match score."), 18 ("a verification system makes a decision by comparing the match score *s* to a threshold η").

780. For these reasons, in my opinion, a POSITA would have understood Zhang as disclosing a conventional facial-authentication procedure in which the first or second data (or both) must match "previously captured" enrollment data within a predetermined threshold to authenticate the identity of the user; otherwise, the user is ***not*** authenticated. In fact, authenticating a user's *identity* is a central aspect of facial authentication systems, and not just evaluating whether the face is three-dimensional or not. *Zhang*, [0001] (noting the purpose of the invention is to prevent unauthorized users from accessing secure resources); *see* Ex-1018, 259 (noting

"[l]iveness detection"—like Zhang—is just one aspect of biometric authentication systems to mitigate spoofers).

6. Claim 6: The non-transient computer readable medium of claim 1 wherein the computing device camera is part of a computing device and the computing device is a hand-held device, and the user holds the device at the first distance to capture the at least one first image and then holds the computing device at the second distance to capture the at least one second image.

781. In my opinion, Zhang alone or in combination with Tanii teaches claim 6's additional limitation.

782. Zhang discloses that the face authentication process can be implemented on a variety of different types of hand-held computing devices, such as a cellular or other wireless phone, a digital camera or video camera. *Zhang*, [0013]. Moreover, Tanii notes that distance-induced distortions often occur in mobile devices that have incorporated wide-angle lenses, and the amount of distortion is dictated by the distance between the user and the camera. *Tanii*, [0007], [0047]-[0048], Figs. 3A-B, 4A-B.

783. In my opinion, a POSITA would have also understood that, when performing a three-dimensional verification of the face on a mobile computing device, *see* §XII.C.2.c (1[b]), the user would adjust the distance by holding the mobile device and extending and retracting their arm (holding the mobile device at a first distance, then a second distance). A POSITA would have appreciated that

mobile devices are routinely held to capture images, and holding the mobile device and adjusting distance would be a convenient and obvious way of changing the distance.

7. Claim 7: The non-transient computer readable medium of claim 1 wherein the first biometric data and the second biometric data comprise image data of facial features.

784. In my opinion, Zhang discloses or suggests claim 7's additional limitation.

785. Zhang discloses the biometric data is image data of facial features. *Zhang*, [0027] ("[O]ne or more feature points are extracted from two images… A variety of different feature points can be extracted, such as a corner of an eye, a corner of a mouth, a tip of a nose, and so forth."). In my opinion, a POSITA would have understood that Zhang's identification of facial landmarks from the captured images constitutes image data of facial features. In fact, many different conventional methods were known to use image data to perform facial authentication. *See, e.g.*, Ex-1018, 109-27.

8. Independent Claim 8

a. 8[pre]: A method for determining when a user, based on images of the user's face, does not exhibit three-dimensionality, the method comprising:

786. If the preamble is limiting, in my opinion, Zhang discloses or suggests it.

787. Specifically, in addition to computer-readable code, Zhang discloses a method to verify a user's face is three-dimensional based on images of the user's face. *See* §XII.C.2.a (1[pre]).

b. 8[a]: capturing at least one first image of the user's face taken with a camera located at a first distance from the user's face, the camera associated with a computing device;

788. In my opinion, Zhang discloses or suggests 8[a] for the reasons discussed in §XII.C.2.b (1[a]).

c. 8[b]: processing the at least one first image or a portion thereof to create first data;

789. In my opinion, Zhang discloses or suggests 8[b] for the reasons discussed in §XII.C.2.b (1[a]).

d. 8[c]: moving the camera to a second distance from the user's face, where the second distance is different from the first distance;

790. In my opinion, Zhang alone or combined with Tanii teaches 8[c].

791. Specifically, Zhang (whether alone or combined with Tanii) discloses capturing a series of images as part of the three-dimensional verification of a face. *See* §XII.C.2.c (1[b]). In my opinion, a POSITA would have understood that, to capture multiple images at multiple distances, either the camera would need to be moved from the first to the second distance, or the subject would need to be moved in relation to the camera from a first distance to a second distance because these are

the only two ways to change the relative distance between the two. *See* §XII.C.2.c (1[b]).

792. Furthermore, because Zhang discloses implementing the method using mobile computing devices that incorporate a camera such as laptop computers or "wireless phone[s]," *Zhang*, [0013], in my opinion, a POSITA would have understood that Zhang at least suggests moving a hand-held device with a camera in relation to the user's face because that is the more user-convenient option than moving the user's face relative to the camera, and because there are only two possible options for changing the distance between the user's face and the camera (either moving the camera or the user).

e. 8[d]: capturing, at the second distance, at least one second image of the user's face taken with the camera associated with the computing device;

793. In my opinion, Zhang alone or combined with Tanii teaches 8[d] for the reasons discussed in §XII.C.2.c (1[b]).

f. 8[e]: processing the at least one second image or a portion thereof to create second data;

794. In my opinion, Zhang alone or combined with Tanii teaches 8[e] for the reasons discussed in §XII.C.2.c (1[b]).

g. 8[f]: examining the first data and the second data to determine whether differences between the first data and the second data indicate an expected type of distorting change in at least one image that is consistent with a real person being imaged and which is indicative of three-dimensionality;

795. In my opinion, Zhang alone or combined with Tanii teaches 8[f]. *See* §§XII.C.2.d (1[c], describing expected differences), XII.C.2.e (1[d], describing expected distorting changes from the Zhang-Tanii's distance-induced distortion approach).

h. 8[g]: determining the user's face is not three-dimensional when the first data and the second data do not have expected differences indicating the user exhibits three-dimensionality.

796. In my opinion, Zhang alone or combined with Tanii teaches 8[g] for the reasons discussed in §XII.C.2.e (1[d]).

9. Claim 9

797. In my opinion, Zhang alone or combined with Tanii teaches claim 9 for the reasons discussed below.

a. 9[a]: The method of claim 8 further comprising: capturing one or more additional images at distances from the user's face that are between the first distance and the second distance;

798. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 9[a].

799. Zhang discloses capturing a series of intermediate images between two non-adjacent (e.g., first and second) images. *Zhang*, [0035]-[0037].

800. In my opinion, a POSITA would have understood that these intermediate images would provide images at different positions (e.g., rotation or translation) of the camera between the first and second images. That is because Zhang describes them as being sequentially captured, and the purpose of a homography transform is to change the perspective of a camera during image capture. *See* §VII.C (Zhang). A POSITA would have therefore understood the series of images would be sequential as the perspective changes between two endpoints.

801. Additionally, Tanii discloses that distance-induced distortions increase as distances between the face and camera decreases. *See Tanii*, [0048]. In my opinion, a POSITA would have understood that a relationship exists between the extent of distance-induced distortion and distance based on Tanii. Thus, for any set of captured images at different distances, the closest will display the most distance-induced distortion, the furthest will display the least, and any intermediate images captured at intermediate distances will display intermediate levels of distortion. In my opinion, a POSITA would have also understood generally that capturing more images would be preferred, as it would provide the biometric-authentication system more samples to evaluate to authenticate the identity of the user. The upper limit, in

my opinion, would be dictated by user resistance to the time of the authentication procedure.

802. For this reason, in my opinion, a POSITA would have been motivated to capture additional images at least one distance between the first and second distance to determine whether the series of images at different distances exhibit an expected range of distance-induced distortion to further confirm the three-dimensionality of the user's face. A POSITA would have appreciated that capturing additional images increases accuracy, but comes at a cost of processing power and acquisition time.

b. 9[b]: for at least one of the one or more additional images, generating additional data;

803. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 9[b].

804. Zhang discloses processing an additional, intermediate image to obtain additional data. *Zhang*, [0036] ("the feature point extraction and feature point matching in acts 302 and 304 can be generated for each adjacent pair of images in the sequence, which can facilitate the feature matching process when matching features across two images with one or more intervening images.").

805. Moreover, in my opinion, when modifying Zhang in view of Tanii, a POSITA would have been motivated to acquire an additional image and extract

additional data from the additional image to compare it to the positions of the data based on the first and second images. *See* §XII.C.9.a (9[a]).

c. 9[c]: examining the additional data, the first data, and the second data, or portions thereof, to determine whether expected differences therebetween indicate the user's face exhibits three-dimensionality.

806. Zhang, alone or combined with Tanii, teaches 9[c]'s additional limitation.

807. Zhang discloses examining intermediate images and comparing them to the first and second images to determine whether expected differences exist. *Zhang*, [0036]-[0037]. And Zhang and Tanii together teach verifying the three-dimensionality of a face by evaluating for expected differences caused by distance-induced distortion. §§XII.C.2.c-XII.C.2.e (1[b]-1[d]).

808. In my opinion, a POSITA would have understood that, when capturing additional images consistent with Zhang, (§XII.C.9.a-XII.C.9.b (9[a]-9[b])), with or without Tanii, facial landmarks would be identified and compared in those additional images as well to determine whether expected differences exist to confirm whether the user's face is three dimensional. *See* §XII.C.9.a (9[a]; describing tradeoffs of accuracy versus processing power and time).

10. Claim 12: The method of claim 8 wherein the computing device is a hand-held device, and the user holds the computing device at the first distance from the user's face when capturing at least one first image and holds the computing device at the second distance from the user's face when capturing the at least one second image.

809. In my opinion, Zhang combined with Tanii teaches claim 12's additional limitation. *See* §XII.C.6 (cl.6).

810. In my opinion, a POSITA also would have understood that, when adjusting the distance of a hand-held computing device, the device would move between a first distance from the user's face and a second distance from the user's face because Zhang and Tanii are both concerned with differences in images of a user's face captured in different conditions, such as distances. *See* §XII.C.2.c (1[b]).

11. Claim 13: The method of claim 8 wherein the first data and the second data comprise at least in part biometric data.

811. Zhang, alone or combined with Tanii, teaches claim 13's additional limitation for the reasons discussed in §§XII.C.2.b-XII.C.2.c (1[a]-[b]).

812. In my opinion, a POSITA would have also understood that data created from an image (or portion of an image) of a user's face would comprise biometric data. *Id.*

12. Claim 14: The method of claim 8 wherein moving the camera comprises moving the camera linearly toward or away from the user's face.

813. In my opinion, Zhang, alone or combined with Tanii teaches claim 14's additional limitation.

814. Zhang discloses capturing "multiple images" from different perspectives. *Zhang*, [0016] ("Image capture component 102 captures multiple images"); *see* §XII.C.2.c (1[b]). In my opinion, a POSITA would have understood Zhang disclosure of capturing multiple images to perform a homography transform to mean that the camera is moved between different perspectives, which would obviously include a linear translation towards or away from the user's face to capture images at different distances. *See id.*

815. Similarly, Tanii teaches that images captured at different distances exhibit different degrees of distortion. *Tanii*, [0047]-[0048], [0056], Figs. 3A-B, 4A-B. In my opinion, a POSITA would have understood that Tanii depicts a *linear* movement of the camera towards or away from the user's face because the face shown in Figures 3A-B and 4A-B are from the same face-on perspective, but with different degrees of distortion. In my opinion, therefore, a POSITA would have understood that, when modifying Zhang in view of Tanii to look for distance-induced distortions, the camera would be moved linearly towards or away from the user's face between image captures. §XII.C.2.c (1[b]).

13. Claim 16: The method of claim 8 wherein a face of the user is held steady when capturing the at least one first image and the at least one second image and the camera moves from the first location to the second location.

816. In my opinion, Zhang, alone or combined with Tanii, teaches claim 16's additional limitation.

817. In my opinion, a POSITA would have understood that Zhang, both alone and combined with Tanii, suggests moving the camera linearly to capture images at two different distances. *See* §§XII.C.2.c (1[b]), XII.C.12 (cl.14). In my opinion, a POSITA would have understood that, when moving the camera to capture images from different distances, the user's face would be stationary (e.g., steady) both during each image capture (to ensure each image is not blurry due to the camera's exposure time) and during movement of the camera (to isolate any differences between images to those attributable to the change of distance). Moreover, a POSITA would have appreciated that holding the user's face steady and moving the camera closer and further away would be more user friendly than forcing the user to move their head closer and further from the camera while holding the camera steady.

14. Independent Claim 17

a. 17[pre]: A method, performed using a computing device, for providing authentication of a person during an authentication session, the method comprising:

818. If the preamble is limiting, in my opinion, Zhang discloses or suggests it.

819. Zhang discloses a computer-implemented authentication method for authenticating a person during an authentication session. *See, e.g.*, *Zhang*, Abstract, [0001]; §§XII.C.2.a (1[pre]), §XII.C.5 (cl.5).

b. 17[a]: capturing a first image of a head of the person with a camera at a first distance from the person, the camera associated with the computing device;

820. In my opinion, Zhang, alone or combined with Tanii, teaches 17[a] for the reasons discussed in §XII.C.2.b (1[a]).

821. Furthermore, in my opinion, a POSITA would have further understood that, when capturing an image of a user's face, the image would be of the user's head because the face exists on the head.

c. 17[b]: changing a distance between the person and the camera to a second distance, which is different from the first distance;

822. In my opinion, Zhang, alone or combined with Tanii, teaches 17[b] for the reasons discussed in §§XII.C.2.c (1[b]), XII.C.8.d (8[c]).

d. 17[c]: capturing a second image of the head of the person with the camera at the second distance from the person;

823. In my opinion, Zhang, alone or combined with Tanii, teaches 17[b] for the reasons discussed in §§XII.C.2.c (1[b]), XII.C.14.b (17[a]).

e. 17[d]: comparing one or more aspects of the head from the first image or first biometric data derived from the first image to one or more aspects of the head from the second image or second biometric data derived from the second image to determine whether expected differences are not present, wherein the expected differences:

824. In my opinion, Zhang, alone or combined with Tanii, teaches 17[d] for the reasons discussed in §XII.C.2.d (1[c]).

825. In my opinion, a POSITA would have also understood that, when comparing facial landmarks consistent with Zhang, that would be a comparison of "aspects of the head," because facial landmarks are present on the user's head.

f. 17[e]: would be present when the first image and second images of the head of the person being captured at different distances has three-dimensional characteristics but not if the head did not have three-dimensional characteristics; and

826. In my opinion, Zhang, alone or combined with Tanii, teaches 17[e]. *See* §§XII.C.1 (motivation), XII.C.2.e (1[d]), XII.C.14.e (17[d]).

g. 17[f]: the expected differences result from differences in relative dimensions of a person's face appearing different when capturing images is done close to the person's face and far from the persons face; and

827. In my opinion, Zhang, alone or combined with Tanii, teaches 17[f]. *See* §§XII.C.1 (motivation), XII.C.2.e (1[d]), XII.C.14.e (17[d]).

828. Moreover, in my opinion, a POSITA would have understood that the approach disclosed by Zhang (homography), or taught by Zhang combined with Tanii (distance-induced distortion), would be used to evaluate for expected differences in the relative dimensions of a person's face when one image is captured close to the person's face and another image is captured far from the person's face. Specifically, Zhang discloses performing a homography transformation, which a POSITA would have understood involves capturing multiple images from multiple perspectives (including distances) to determine whether a homography

transformation of one exhibits distortions of facial features attributable to the fact that the three-dimensional face does not exist in a single plane. §§XII.C.1 (motivation), XII.C.2.e (1[d]). And the Zhang-Tanii approach looks for different degrees of distance-induced distortion—which alters the relative dimensions of facial features—between a close and far image. *Id.*

h. 17[g]: if the expected differences are not present, denying authentication of the person and providing notice thereof to one or more of the person, a third party, or a software application, wherein the authentication is authentication of liveness, three-dimensionality, or both.

829. In my opinion, Zhang, alone or combined with Tanii, teaches 17[g] for the reasons discussed in §XII.C.2.e (1[d]).

830. Furthermore, when an authentication attempt is rejected as a spoof attempt, Zhang discloses providing notice of the rejection to the user, and that the authentication is one of at least three-dimensionality. *Zhang*, [0003], [0017].

15. Claim 18: The method of claim 17 wherein the steps of comparing, denying authentication, and providing notice are performed by a server that is remote from the computing device.

831. In my opinion, Zhang discloses or suggests claim 18's additional limitation.

832. Zhang discloses that the image capture component and live face detection module (104) with accompanying 3D structure determination module

(112) can communicate and send data, including biometric facial feature data, over a variety of different networks, such as the Internet, a local area network (LAN), an intranet, etc. *Zhang*, [0014].

833. Although Zhang does not expressly disclose that the data is sent to a "server," in my opinion, a POSITA would have understood based on Zhang's disclosure that the use of a server would be implicit, or at least obvious. Specifically, servers are well-known networking infrastructure, and servers were known to be used for back-end processing of biometric data. *See, e.g.*, *Derakhshani*, 9:27-58, 10:1-24; Ex-1016, Abstract, [0040]-[0043]; Ex-1012, Fig. 1A, 5:24-50. Furthermore, in my opinion, a POSITA would have understood that what types of data to send to the server is a design choice that balances the processing capabilities of the image-capture device and the transmission bandwidths available. If there is little bandwidth and ample processing power on the image capture device, a POSITA would be motivated to design the authentication on the device, so only a small amount of information needs to be transmitted to the server. If there is ample bandwidth available, but the image capture device is constrained in its processing power, a POSITA would be motivated to transmit raw image data to the server, so that the authentication could be done there. It would be obvious to a POSITA how to manage intermediate scenarios through standard engineering analysis.

834. For these reasons, in my opinion, a POSITA would have been motivated in circumstances where a back-end server is used for biometric authentication to configure the server to compare, deny authentication, and provide notice from the remote server.

16. Claim 19: The method of claim 17 wherein the authentication is authentication of three-dimensionality.

835. Zhang discloses or suggests claim 19's additional limitation.

836. Zhang discloses an authentication process that includes verifying the three-dimensionality of a user. *Zhang*, [0003], [0017]-[0018].

17. Independent Claim 20

a. 20[pre]: A method for determining whether a user exhibits three-dimensionality, the method comprising:

837. If the preamble is limiting, in my opinion, Zhang discloses or suggests it for the reasons discussed in §§XII.C.2.a (1[pre]), XII.C.8.a (8[pre]).

b. 20[a]: capturing at least one first image of a user's face taken with a camera located a first distance from the user, the camera associated with a computing device;

838. In my opinion, Zhang discloses or suggests limitation 20[a] for the reasons discussed in §XII.C.2.b (1[a]).

c. 20[b]: processing the at least one first image or a portion thereof to create first data, the first data derived from the user's face;

839. In my opinion, Zhang discloses or suggests limitation 20[b] for the reasons discussed in §XII.C.2.b (1[a]).

d. 20[c]: intentionally moving the camera from the first location to a second location, the second location being a second distance from the user, or the user moving to change a distance between the user and the camera from the first distance to the second distance;

840. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 20[c] for the reasons discussed in §§XII.C.2.c (1[b]), XII.C.8.d (8[c]).

841. Furthermore, in my opinion, a POSITA would have understood that any movement of the camera to change perspective (Zhang) or create a distance-induced-distortion effect (Zhang combined with Tanii) would involve *intentional* movement of the camera to induce these effects because both Zhang and Tanii suggest capturing images as nontrivially different distances, which would obviously call for intentional movement of the camera.

e. 20[d]: capturing at least one second image of the user's face taken with the camera located a second distance from the user, the second distance being different than the first distance;

842. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 20[d] for the reasons discussed in §XII.C.2.c (1[b]).

f. 20[e]: processing the at least one second image or a portion thereof to create second data, the second data derived from the user's face;

843. In my opinion, Zhang alone or in combination with Tanii teaches limitation 20[e] for the reasons discussed in §XII.C.2.c (1[b]).

g. 20[f]: analyzing the first data to determine at least if the first data exhibits first characteristics that indicate the first data was derived from an image of the user captured at the first distance;

844. In my opinion, Zhang alone or combined with Tanii teaches limitation 20[f].

845. Zhang discloses that, as part of the verification process, "sub regions within a face (such as eyes, mouth, nose, and so forth)" are detected. *Zhang*, [0032].

846. In my opinion, a POSITA would have understood that, when analyzing the images for facial landmarks, the landmarks would exhibit characteristics (e.g., size, distance from other landmarks, or other features) that indicates the image was taken at some distance, because—when all else is held equal, such as the camera system, lenses, and focus settings used—those types of characteristics of the landmarks depend on the distance between the user and the camera. *See, e.g.*, Ex-1018, 107 (Fig. 3.9 providing an example of how distance from the camera affects resolution of facial features).

847. Moreover, a POSITA would have understood that, when evaluating images for specific types of distortion (like distance-induced distortion taught by

Zhang combined with Tanii), it is especially important for the three-dimensional verification process to be able to orient which image is taken at which distance for the focus-distance and distance-induced distortion approaches in order to assess the three-dimensionality of the face. For instance, for the distance-induced distortion approach, it is important to characterize which image is the “close” image (e.g., Tanii’s Fig. 4B) to assess whether it specifically exhibits expected degrees of distance-induced distortion.

h. 20[g]: analyzing the second data to determine at least if the second data exhibits second characteristics that indicate the second data was derived from an image the user captured at the second distance, wherein the first characteristics or the second characteristics include at least distortion within the at least one first image or the at least one second image;

848. In my opinion, Zhang alone or combined with Tanii teaches limitation 20[g] for the reasons discussed in §XII.C.17.g (20[f]).

849. Moreover, when applying the Zhang-Tanii distance-induced distortion approach, a POSITA would have understood that the images are analyzed to identify specific distortion characteristics (e.g., barrel or fish-eye distortions for distance-induced distortion). §§XII.C.1 (motivation), XII.C.2.e (1[d]).

i. 20[h]: determining the user does not exhibit the expected degree of three-dimensionality when either or both of the following occur:

850. In my opinion, Zhang combined with Tanii teaches limitation 20[h].

851. As explained in further detail below, Zhang determines that the user's face is not three-dimensional when there is no mismatch between the biometric landmarks and no expected differences exist. *Zhang*, [0031]-[0033].

j. 20[h1]: the step of analyzing the first data determines the first data does not exhibit first characteristics that indicate the first data was derived from an image of the user captured at the first distance; or

852. In my opinion, Zhang combined with Tanii teaches limitation 20[h1].

853. Zhang discloses that a face is determined to not be three-dimensional when expected differences do not exist in the facial landmarks (e.g., biometric data) of the first and second images. *See* §XII.C.2.e (1[d]). In my opinion, a POSITA would have understood that, if the first and second images match after one of the images undergoes Zhang's homography transformation, a determination is made that the face is planar and not three dimensional. *See* §VII.C (Zhang; explaining how a match after a homography transformation is an indication that the face is planar).

854. In my opinion, a POSITA would have understood that when combining Zhang and Tanii, a POSITA would have understood that if neither the first nor second images display any differences in the degree of distance-induced distortion (e.g, the claimed "characteristics"), then that is an indication that either the first or second image was not taken at a closer (e.g., first or second) distance from the face, because such distortions would be expected to be induced in at least one image. In other words, a POSITA would have appreciated that if the "first" image is intended

to be the “closer” image, then a lack of distortionsin the “first” data would indicate the first data was not derived from an image of the user captured at the first (closer) distance.

k. 20[h2]: the step of analyzing the second data determines the second data does not exhibit second characteristics that indicate the second data was derived from an image of the user captured at the second distance.

855. In my opinion, Zhang combined with Tanii teaches limitation 20[h2]. for the reasons discussed in §XII.C.17.j (20[h1]).

856. Similarly, a POSITA would have appreciated that if the “second” image is intended to be the “closer” image, then a lack of distortions (or characteristics, as claimed) in the “second” data would indicate the first data was not derived from an image of the user captured at the first (closer) distance.

18. Claim 22: The method of claim 20 wherein the at least one first image and the at least one second image are captured with a hand-held computing device, and the user holds the computing device at the first distance when capturing at least one first image and at the second distances when capturing the at least one second image.

857. In my opinion, Zhang combined with Tanii and Tahk teaches claim 22’s additional limitation for the reasons discussed in §XII.C.6 (cl.6).

19. Claim 23: The method of claim 20 wherein the first data and the second data comprise at least in part biometric data.

858. In my opinion, Zhang alone or combined with Tanii teaches claim 20 for the reasons discussed in §XII.C.11 (cl.13).

20. Claim 24: The method of claim 20 wherein the first data and the second data comprise at least in part image data of facial features.

859. In my opinion, Zhang discloses or suggests claim 24's additional limitation for the reasons discussed in §XII.C.7 (cl.7).

D. Ground 2B: Zhang, Tanii, and Tahk (Claims 4, 10-11, 21)

1. Motivation to Combine

860. In my opinion, a POSITA would have been motivated to modify Zhang, with or without Tanii, in view of Tahk because Tahk provides a user-friendly way of ensuring that a face presented for facial authentication is properly framed. In my opinion, a POSITA would have understood that providing a real-time preview of what an image would look like prior to capturing the image, as well as providing express prompts—such as written instructions or oval shapes on the live-preview screen—as taught by Tahk, §VII.D (Tahk), would have been particularly useful for authentication procedures that require capturing multiple images of a face, as taught by Zhang and Tanii.

861. For instance, providing a user a real-time preview would allow the user to actively adjust the position of the camera and/or their orientation of their face to

properly frame their face for the image capture. Moreover, providing express prompts and oval shapes sized to guide the user to properly position their face would ensure images best suited for facial recognition can be captured, and that the faces would be captured from sufficiently different perspectives and/or distances to ensure the three-dimensional verification taught by Zhang, alone or in combination with Tanii, could be performed. This is particularly important for facial-recognition systems, which generally are known to have difficulty matching faces across different views. *See, e.g.*, Ex-1018, 32.

862. Moreover, a POSITA would have known that providing real-time image feedback, written instructions, and oval shapes to frame a face during a facial authentication process were all well-known and conventional techniques to provide user feedback during image capture as of the time of the invention. *See, e.g.*, Ex-1034, 7:16-8:7, Figs. 6B-7C (providing "an example of an interface used upon registering a facial image of a person to be authenticated" in which "the image of this person is displayed on a monitor" during registration, and oval-shaped prompts to indicate the size/distance of the user from the face authentication sensor); Ex-1035, 5:31-32 ("The computing device may present prompts that instruct the user to perform one or more liveness gestures"), 6:3-4 (same).

2. Claim 4: The non-transient computer readable medium of claim 1 wherein the computing device camera is part of a computing device and the machine executable code is configured to display an interface on the computing device's screen to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

863. In my opinion, Zhang combined with Tanii and Tahk teaches claim 4's additional limitation.

864. Zhang discloses taking a series of images sufficient to calculate a homography matrix. *See, e.g.*, *Zhang*, [0026], Figs. 1, 3. In my opinion, a POSITA would have understood that Zhang already discloses, or that Zhang combined with Tanii teaches, taking a series of images at different distances between the face and the camera. *See* §§XII.C.1 (motivation), XII.C.2.c (1[b]). However, Zhang and Tanii do not expressly teach providing a series of prompts to a user to guide them through different camera positions that would enhance calculations of the homography matrix.

865. Tahk, however, teaches that using one or more prompts on a screen ensures images of the face are captured at the correct distances. *See, e.g.*, Tahk, Figs. 8A-B ("Please step further back" and "Please step further forward," and presenting an oval to frame the face at the correct distance). In my opinion, a POSITA would have been motivated by Tahk to modify Zhang, whether alone or in combination with Tanii, to expressly prompt a user to alter the distance of the camera in order to

either capture sufficiently different images to perform a homography transformation (Zhang) or to capture an image with distance-induced distortions (Tanii) to ensure the images could be used to distinguish live from two-dimensional images of faces. *See also* §XII.D.1 (motivation).

3. Claim 10: The method of claim 8 further comprising displaying one or more prompts on a screen associated with the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

866. In my opinion, Zhang combined with Tanii and Tahk teaches claim 10's additional limitation for the reasons discussed in §XII.D.2 (cl.4).

4. Claim 11: The method of claim 10 wherein the one or more prompts are an on the screen shape within which an image of a face of the user is aligned during capture the at least one first image and the at least one second image.

867. In my opinion, Zhang combined with Tanii and Tahk teaches claim 11.

868. Although Zhang and Tanii do not expressly describe using prompts to guide a user during the facial-authentication process, Tahk teaches this feature, including the use of shape (oval) prompts to frame a user's face. *See* §§VII.D (Tahk), XII.D.1 (motivation).

869. In my opinion, a POSITA would have been motivated to modify Zhang, alone or combined with Tanii, to provide such oval-shaped prompts because they are a natural shape to appropriately size and frame a face at different distances. *See* §§VII.D (Tahk), XII.D.1 (motivation).

5. **Claim 21: The method of claim 20 further comprising displaying one or more prompts on a screen associated with the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.**

870. In my opinion, Zhang combined with Tanii and Tahk teaches claim 21's additional limitation for the reasons discussed in §XII.D.2 (cl.4).

E. Ground 2C: Zhang, Tanii, and Suzuki (Claim 15)

1. Motivation to Combine

871. Zhang discloses implementing a process to verify the three-dimensionality of a user's face by capturing a series of images of a user using a mobile computing device, such as a laptop or phone. *See, e.g.*, *Zhang*, [0013]. User-facing webcams have existed since the mid 1990s, and the front-facing smartphone camera became popularized with the launch of the iPhone 4 in 2010. *See* Ex-1037. Thus, although Zhang does not expressly say so, a POSITA reading Zhang would have understood that Zhang's process—when carried out on a mobile computing device such as a laptop or phone—would have utilized a user-facing camera because by then they had become common. And a POSITA would have further understood that a user-facing camera would be used for facial authentication so that the user does not need to turn the device around—hiding the display of the device—to capture an image of themselves without any feedback or guidance.

872. A POSITA would have also understood that cameras often utilize illumination sources, such as camera-flash modules or other lighting to improve the

quality of the captured images (especially in low-light conditions). In fact, camera flashes have been used in conjunction with cameras since the late 1800s; using the two together is ubiquitous.

873. When providing a user-facing camera on a mobile computing device, however, Suzuki recognized that, rather than provide an independent camera-flash module, the device's display could serve as the flash module. *Suzuki*, [0009], [0019], [0021], [0024]-[0025]. That way, fewer electronic components would be required to provide better lighting for a user-facing camera to maintain the device's small size. *Id.*, [0005]-[0006].

874. A POSITA would have therefore been motivated by Suzuki to modify Zhang to use the device's existing, user-facing display to provide the illumination source for a user-facing camera. That way, capture-image quality would be improved without requiring a separate user-facing flash module. Moreover, a POSITA would have had a reasonable expectation of success in carrying out this modification because mobile-device displays already provided for controlled illumination of the contents of the display screen; all that would be required would be to configure the display to output white pixels at maximum brightness at the time of image capture.

2. Claim 15: The method of claim 8 further comprising illuminate a screen of the computing device while capturing the at least one first image and/or the at least one second image to improve quality of an image being captured.

875. In my opinion, Zhang in view of Tanii and Suzuki teaches claim 15's additional limitation.

876. Neither Zhang nor Tanii discuss using the device's screen as an illumination source (e.g., the camera's flash). However, Suzuki taught using a user-facing screen to serve as a camera's illumination source (or "flash"). *Suzuki*, [0009], [0019], [0021], [0024]-[0025]. In my opinion, a POSITA would have been motivated to modify Zhang to use a user-facing display (as Zhang already envisioned) as a light source for capturing an image to improve the quality of the image. §XII.E.1 (motivation).

XIII. '910 PATENT: DETAILED EXPLANATION OF GROUNDS

A. Ground 1A: Derakhshani and Tanii (Claims 1-13, 15-24)

1. Motivation to Combine

877. In my opinion, a POSITA would have been motivated to combine Derakhshani and Tanii because both concern identifying and accounting for the three-dimensional nature of a face when capturing an image. They differ, however, in what principles are used to account for the face's three-dimensionality. Derakhshani, for instance, uses changes in focus distance (e.g., image resolution for structures imperfectly in focus) and/or parallax effect to determine whether a face

has depth. *See* §VII.A (Derakhshani). And although Tanii is not expressly directed to *evaluating* whether a face has depth, Tanii exemplifies the well-known distortions caused by the interaction between the camera's lens and the three-dimensional nature of the face at different distances, *see* §VII.B (Tanii). A POSITA would have appreciated, therefore, that Tanii recognizes another alternative to evaluating the depth of a face, consistent with Derakhshani's existing two approaches.

878. A POSITA would have recognized, for instance, that Derakhshani's focus-distance approach and Tanii's evaluation of distance-induced distortions are both attributable to classical optical effects such as refraction and diffraction caused by (among other factors) different distances between the camera and the object(s) being captured. *Derakhshani*, 16:57-60 ("Degree of focus is a measure of the extent to the image of the landmark is blurred by optical effects … (e.g., due to *diffraction* and convolution with the aperture shape."); *Tanii*, [0048] (noting the "unnatural image" is caused by the angles of the face relative to the angle of the camera lens).

879. Derakhshani and Tanii differ, however, in the type of effect that is occurring. Specifically, Derakhshani takes advantage of the blurring of objects that are at distances *other than* the camera's focal plane (referred to by photographers as a "bokeh effect"), which makes those objects appear unfocused. *Derakhshani*, 16:54-57; §VII.A (Derakhshani). By adjusting the focus distance (or position of the focal plane by moving the camera) and evaluating when objects (or features of an

object) in an image are clear versus when they are blurry, distance information can be derived. *Derakhshani*, 16:51-63; §VII.A (Derakhshani).

880. Tanii is more specifically concerned with a type of radial distortion that arises due to the interaction of certain (e.g., wide-angle) lenses and the three-dimensional nature of the face. §VII.B (Tanii). As Tanii explains, the convex shape of a three-dimensional face, when placed near the lens, exacerbates this type of distortion. *Tanii*, [0048]; §VII.B (Tanii). Thus, particularly when a camera incorporates a wide-angle lens, images of a face close to the camera will exhibit significant radial distortion in-part because of the distances between different facial features and the lens, and in-part because the face occupies both the center and the periphery of the camera's field of view so differences in radial distortion are more apparent. *Tanii*, [0047]; §VII.B (Tanii). But when the face is further from the camera and occupies less of the image, the distortion will be less apparent because the face is more centered on the region of the lens where radial distortion is not as severe, and there is sufficient distance for the light rays from the face to strike this central portion of the lens. *Tanii*, [0047]; §VII.B (Tanii).

881. In my opinion, a POSITA would have appreciated that when evaluating multiple images taken at either different *focus* distances or *actual* distances, these different effects serve to provide information about an object's depth. In other words, a POSITA would have understood that Tanii teaches another obvious alternative to

Derakhshani's existing two approaches to evaluate whether a face being captured is three-dimensional or not.

882. That said, a POSITA would have also had specific reasons to substitute Derakhshani's existing approaches with Tanii's distance-induced distortion analysis in certain circumstances. A POSITA would have understood, for instance, that implementing Derakhshani's focus-distance approach requires a camera with a sufficiently sized sensor and lens that could provide enough sensitivity to distinguish small differences in depth on the scale of a few centimeters when trying to evaluate the depth of a face. *See Derakhshani*, 16:48-51; Ex-1029, 3 (A 200mm lens focused at 12ft will have a smaller depth of field compared to a 20mm lens focused at 12ft).

883. But a POSITA would have also understood that the cameras typically found in mobile devices—especially around the 2014 timeframe—do not have this ability; mobile devices typically incorporate wide-angle lenses to capture a wide field of view, with a fixed focal length and a large depth of field because of their small size. Tanii, [0007]; Ex-1030 ("Other features of a smartphone are obvious but worth stating, they almost always are fixed focal length, fixed aperture, with no shutter, sometimes with an ND filter (neutral density) and generally not very low F-number. In addition to keep modules thin, focal length is usually very short, which results in wide angle images with lots of distortion."). With such limited-capability cameras, it was known that distortions would therefore largely be a product of the

lens shape and distance between the object and the lens. *See* Ex-1017, 177 ("The amount of spherical aberration, when the aperture and focal length are fixed, varies with both the ***object distance*** and the lens shape."). In other words, there is not enough room in mobile devices to incorporate large image sensors with small F-numbers (a measure of light-gathering ability of the camera) to allow these cameras to fine-tune the focus distance and induce blurring of out-of-plane objects. That is why, for instance, the iPhone introduced its "Portrait Mode" (in 2016, a few years after the earliest possible effective date) as a software-based *simulation* of the blurring effect that can only be achieved by much larger cameras. Ex-1031 (noting how blurring backgrounds was "previously only capable on DSLR cameras" prior to the iPhone's software-based "bokeh" effect).

884. For this reason, in my opinion, a POSITA would have been motivated to modify Derakhshani to capture at least two images at different *actual* distances and evaluate whether one exhibits more distance-induced distortion than the other, as suggested by Tanii. A POSITA would have been especially motivated to make this change when implementing biometric authentication in a mobile device as Derakhshani already envisions. *Derakhshani*, 5:23-26. A POSITA would have found such a modification obvious because both techniques merely involve the application of different well-known optics principles relating camera design and object's distance from the camera, and would have had a reasonable expectation of

success in doing so because Tanii already taught a mechanism to identify such distance-induced distortions and thus indicate when the face being captured has depth. *See, e.g.*, *Tanii*, [0056].

885. Although Derakhshani separately discloses a process to verify the three-dimensionality of a face using parallax, in my opinion, a POSITA would have understood that evaluating for distance-induced distortion consistent with Tanii would be easier for users on a mobile device. Specifically, a POSITA would have naturally understood that mobile devices such as phones or laptops typically capture images of users at arm's length distances because that is how these devices are used (at arm's length). Moreover, a POSITA would have appreciated that facial features do not have *significant* differences in their depth (on the order of a few centimeters, as opposed to meters between the face and a background). Thus, to evaluate for parallax at hand-held distances with suitable accuracy, a POSITA would expect that the user would need to move their device around their head, or could simulate a parallax effect by rotating their head around a stationary camera to create substantial differences in perspective and thus more parallax to more accurately verify the face as three-dimensional. But to do so would have involved moving the device out of the user's line of sight, meaning the user could not see exactly what they are capturing or know if what they were capturing is sufficient.

Evaluating for distance-induced distortions when the camera is held at different distances consistent with Tanii, however, could be accomplished while keeping the device directly in the user's direct line of sight, and would therefore be easier for users to verify that their face is, in fact, three dimensional. But, in my opinion, a POSITA would have also appreciated that biometric security is always subject to spoofing, and thus would have known that evaluating for distance-induced distortion consistent with Tanii could be *supplemented* by also evaluating for any parallax.

2. Independent Claim 1

a. 1[pre]: A computing device for verifying three-dimensionality of a user via a user's camera equipped computing device, the computing device comprising:

886. If the preamble is limiting, in my opinion, Derakhshani discloses or suggests it.

887. Derakhshani discloses systems and methods for using a camera-equipped computing device for "biometric authentication." *See Derakhshani*, 1:11-25, 2:4-30, 5:22-27, 6:3-5, 9:10-22, 18:1-3. Although Derakhshani uses the eye as the primary means of authentication, *see, e.g.*, *id.*, Abstract, as part of the ocular-authentication process, Derakhshani also verifies that the user's face is three-dimensional by capturing multiple images of a user's face at different focus distances

or from different perspectives to calculate a "spatial metric" representing the face's three-dimensionality. *Id.*, 1:11-25, 3:14-15, 16:44-18:4.

b. 1[a]: a processor configured to execute machine executable code;

888. In my opinion, Derakhshani discloses or suggests 1[a].

889. Derakhshani discloses that the invention can be implemented in computing devices such as a "smart phone, a tablet computer, a television, a laptop computer, or a personal computer" (*Derakhshani*, 5:22-27), which incorporate a processor configured to execute machine-readable code (*see, e.g.*, *id.*, Fig. 9, 2:4-12, 2:31-38, 7:15-20, 22:12-44, 23:26-37, 24:49-25:8).

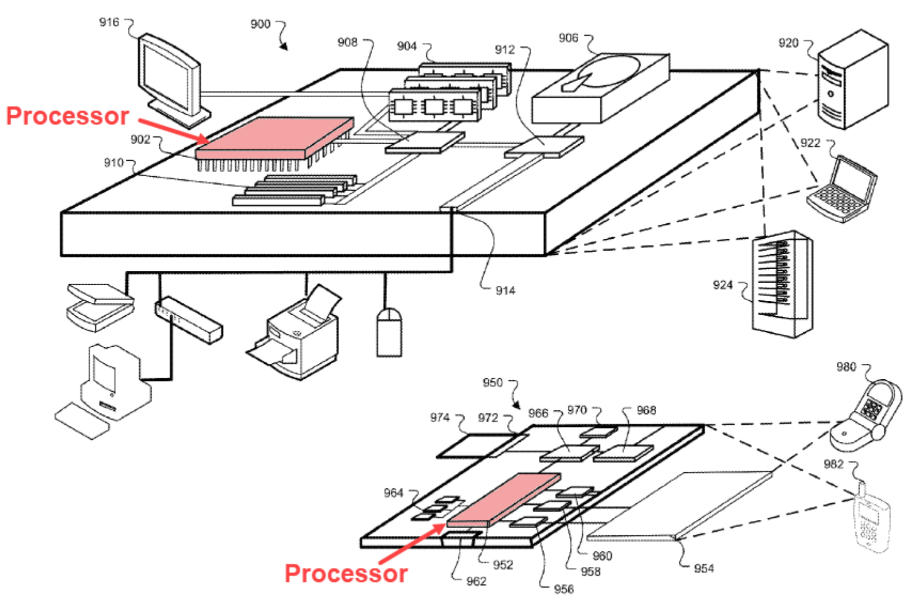


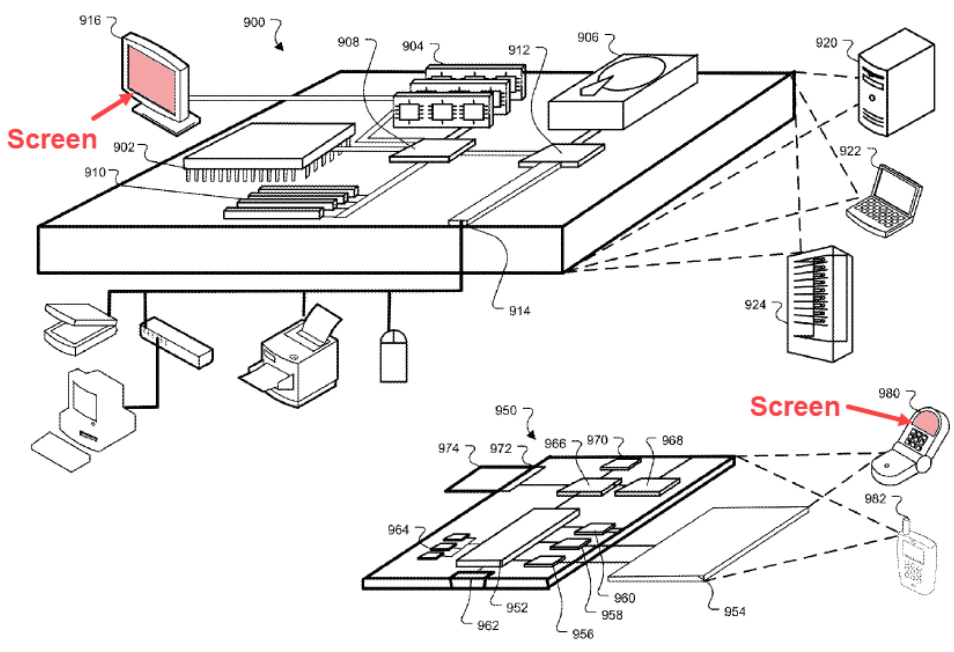
Fig. 9

Id., Fig. 9 (annotated). Derakhshani also discloses the computing device may be a server that also comprises a processor. *Derakhshani*, 7:38-50, 8:48-9:4, 9:27-31, 10:16-19 ("the server system 514 is a data processing apparatus that includes one or more processors."), 23:14-44.

c. 1[b]: a screen configured to provide a user interface to the user;

890. In my opinion, Derakhshani discloses or suggests 1[b].

891. Derakhshani discloses that the computing device incorporates a screen to provide a user interface to the user. *See, e.g.*, *id.* 6:8-11, 9:22-24, 14:35-37, 22:33-38 23:48-52.



Id., Fig. 9 (annotated). Derakhshani also discloses the computing device may be a server that also comprises a screen. *Derakhshani*, 23:14-44, Fig. 9.

d. 1[c]: a camera configured to capture images;

892. In my opinion, Derakhshani discloses or suggests 1[c].

893. Derakhshani discloses that the computing device incorporates a camera configured to capture images. *See e.g.*, *Derakhshani*, 5:23-27, 6:3-10.

e. 1[d]: one or more memories configured to store machine readable instructions that are stored on the memory of the authentication server which when executed by the processor, cause the computing device to:

894. In my opinion, Derakhshani discloses or suggests 1[d].

895. As I mentioned previously, in my opinion, it is somewhat unclear which structure is intended to be the "computing device": (1) either a user-facing computing device that engages with a back-end authentication server; or (2) the authentication server itself. *See* §VI.D.3 (claim construction). However, in my opinion, Derakhshani discloses the claims under either interpretation.

896. First, Derakhshani discloses computing devices (e.g., a personal computer or phone) with memory that stores machine-readable instructions that are executed by the processor. *Derakhshani*, 2:4-12, 2:31-38, 22:26-44, 24:49-25:8.

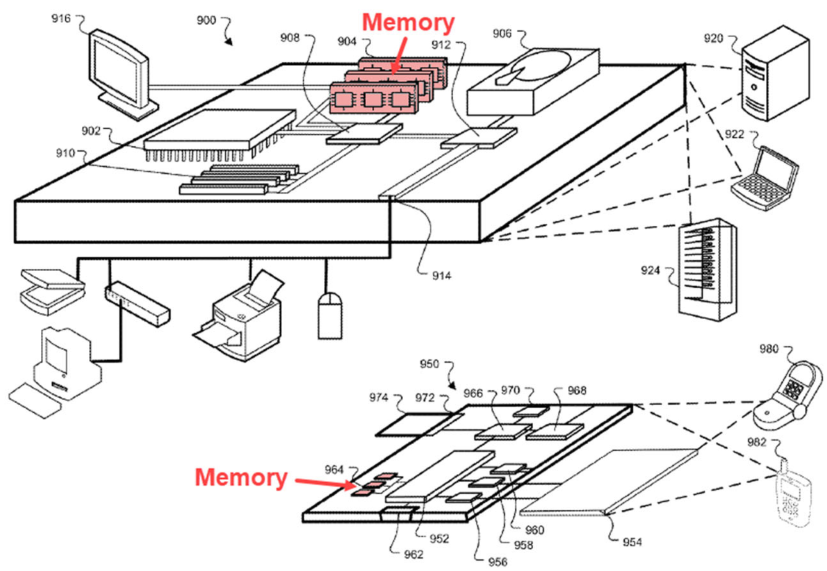


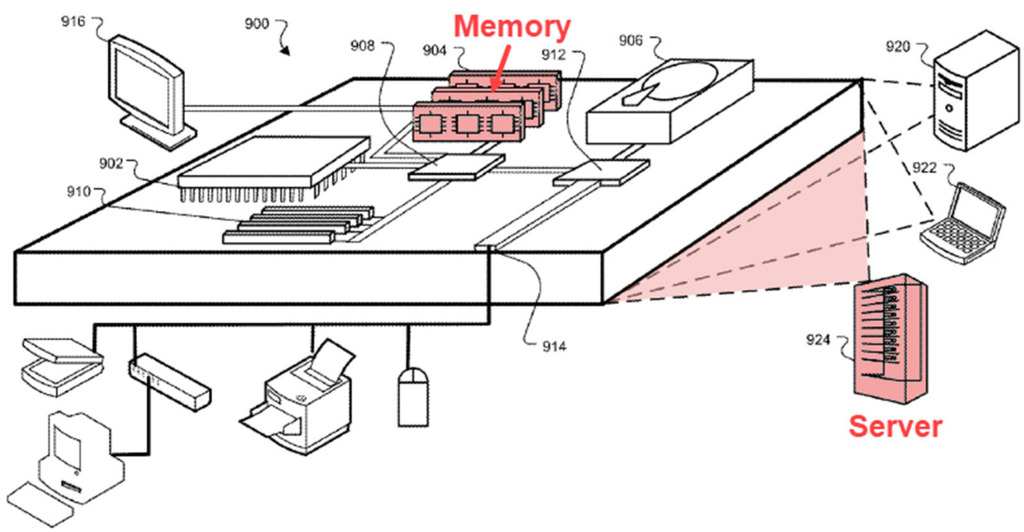
Fig. 9

Id., Fig. 9 (annotated). Derakhshani also discloses different embodiments of networked authentication systems, including: (1) a user-facing computing device that interacts with a "secure transaction service 523 hosted, e.g., by [a remote] server system" with "an authentication module 525 that coordinates authentication of users from the secured server's side of the interaction" (*Derakhshani*, 8:29-39); and (2) a user-facing computing device that hosts a local application that interacts with a remote authentication server (*id.*, 9:10-34). And Derakhshani discloses more generally that "authentication functions may be distributed between the client and the server side processes in a manner suited [to] a particular application." *Id.*, 9:27-58, 10:1-24.

897. In my opinion, a POSITA would have understood that Derakhshani teaches different system configurations, including one in which a remote

authentication server drives all aspects of the authentication process for a user-facing device such that the server directs the user-facing device to carry out certain aspects of Derakhshani's procedure (e.g., user image capture) using local memory that stores machine readable instructions originating from the memory of the authentication server.

898. Second, Derakhshani alternatively discloses that the computing device itself can be a server containing the components depicted in Figure 9 (such as the display 916 and processor 902). *Derakhshani*, 22:12-18, 23:14-25.



899. Although Derakhshani does not show a camera as part of the server system, a POSITA would have understood that Derakhshani at least contemplates the server itself needing biometric protection. A POSITA would have understood that, just like personal computers, servers were known to store sensitive information—from user profiles for websites, employment or medical records, and

more. A POSITA seeking to prevent unauthorized access to reconfigure servers or access their files would have therefore understood that the server itself may be provided with biometric authentication—as Derakhshani at least suggests, *Derakhshani*, 22:12-18, 23:14-25—and that in such cases it would include a camera (1[c]) to carry that authentication out.

f. 1[d1]: capturing at least one first image of the user taken with the camera of the computing device at a first location which is a first distance from the user;

900. In my opinion, Derakhshani discloses or suggests 1[d1].

901. Derakhshani discloses that, as part of the process to verify that the face is in fact three-dimensional, "two or more images of a subject" are captured using the camera of the computing device. *Derakhshani*, 1:44-46, 16:44-17:11, 17:45-18:4.

902. In my opinion, a POSITA would have understood that Derakhshani captures an image at a first distance. *Derakhshani*, 16:44-17:11. Specifically, in my opinion, a POSITA would have understood that there *must* be some distance between the camera and the face to capture the first image—enough so that the camera's field of view encompasses the face. If there were no distance between the camera and the face (e.g., if the camera were pressed up against the user's skin), then the field of view would be limited to just that patch of skin and the face would not be captured in the image, which would be useless for the three-dimensional verification process.

g. 1[d2]: processing the at least one first image or a portion thereof to create first data;

903. In my opinion, Derakhshani discloses or suggests limitation 1[d2].

904. Derakhshani discloses that, as part of the verification process, "a landmark (e.g., an iris, an eye corner, a nose, an ear, or a background object) may be identified and located in the plurality of images." *Derakhshani*, 16:44-54 (focus-distance approach), 17:45-64 (parallax approach).

905. In my opinion, a POSITA would have understood that Derakhshani's identification of facial landmarks constitutes data—and more specifically biometric data—because the identification involves using a computer (which operates on data) to characterize the unique physical characteristics of an individual, which would include the positions of "landmarks" such as a user's eyes, nose, ears, and other such features. *See, e.g.*, Ex-1018, 2 ("biometric recognition can be defined as the science of establishing the identity of an individual based on the physical and/or behavioral characteristics of the person."), 100-103 (describing the types of biometric data about a face used for facial-recognition systems, "such as the structure of the face components (e.g., eyes), [and] the relationship between facial components").

h. 1[d3]: capturing at least one second image of the user taken with the camera of the computing device is at a second distance from the user, the second distance being different than the first distance, the capturing at least one second image of the user occurring after movement of the camera or the user to establish the camera at the second distance from the user;

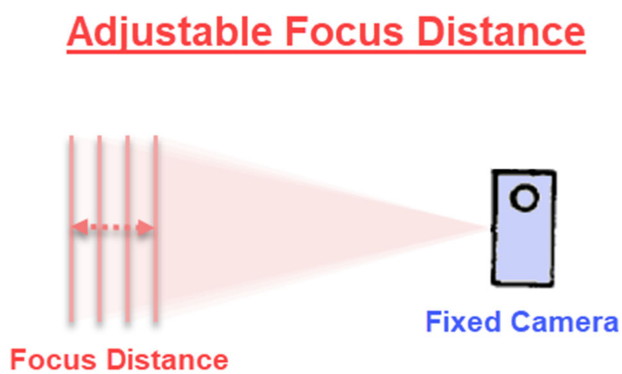
906. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 1[d3].

907. To start, there appears to be a typographical error in this claim. Specifically, the claim recites "capturing at least one second image of the user ***taken*** with the camera…***is*** at a second distance," which does not make sense grammatically. I have therefore read the claim to read "capturing at least one second image of the user with the camera…at a second distance," without the words "taken," and "is" to make it grammatically legible. In my opinion, these two words do not impact the claim as it relates to the application of the prior art.

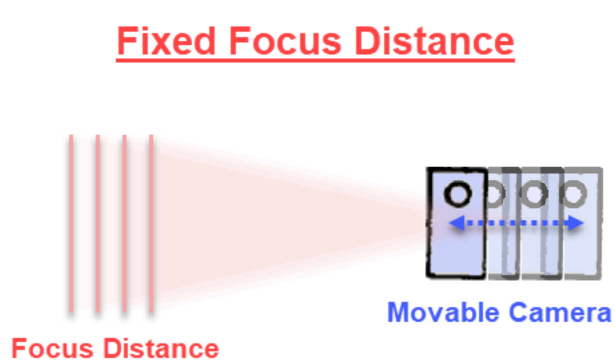
908. Derakh*s*hani discloses capturing "two or more images of a subject" using the camera. *Derakhshani*, 1:44-46, 16:44-17:11, 17:45-18:4; §XIII.A.2.f (1[d1]).

909. When utilizing Derakhshani's focus-distance approach to evaluate depth, a POSITA would have understood that the distance between the user and the camera would need to change only if the camera had a fixed focus distance.

910. When utilizing Derakhshani's focus-distance approach to evaluate depth, however, a POSITA would have understood that adjusting the focus distance of the camera does not require changing the actual distance between the camera if a stationary camera is capable of adjusting its lens position with respect to the image sensor. *See* §VII.A (Derakhshani explaining operation of the focus-distance approach).



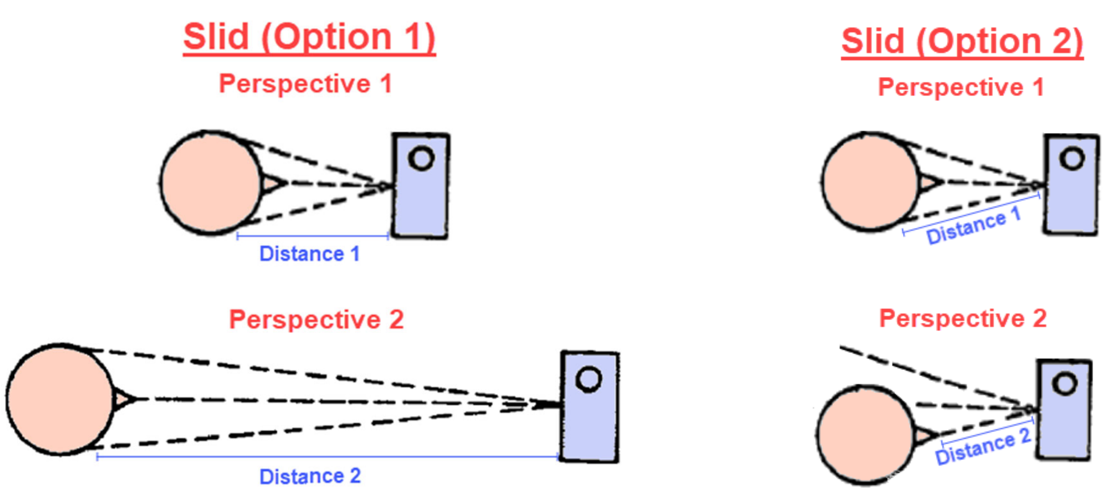
But if the camera has a fixed focus distance (i.e. position of the lens with respect to the image sensor), as is found in many mobile devices (*see* §XIII.A.1), a POSITA would have been motivated to instead implement Derakhshani's focus-distance approach by changing *actual* distance to capture multiple images, as shown below:



In other words, even if the focus distance of the camera cannot be changed, the "slices" of a face at different depths can be evaluated by moving the camera.

911. Regardless, in my opinion, a POSITA would have understood that Derakhshani's parallax approach captures multiple images from multiple distances, because Derakhshani discloses that "[a] plurality of images [are] taken from different perspectives on the subject," such as: (1) when "a single camera [is] rotated *or slid* slightly"; (2) "a user is prompted to move" between image captures; or (3) the sensor moves naturally, such as "where the sensor is a camera in a hand-held user device (e.g., a smartphone or tablet) [that] may naturally move relative to the users face due to involuntary haptic motion." *Derakhshani*, 17:45-18:4.

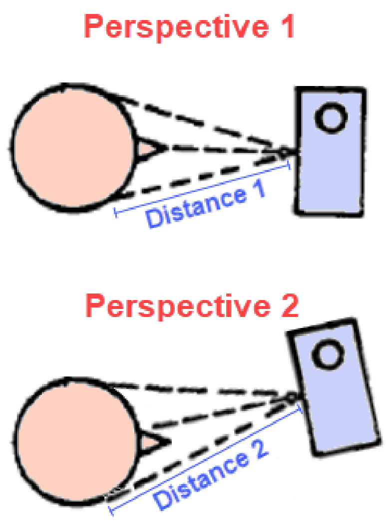
912. In my opinion, a POSITA would have understood that Derakhshani's use of the term "slid" means either of two things: (1) the camera is displaced front-to-back to increase or decrease the distance from the face; or (2) the camera is displaced side-to-side, both of which are depicted below:



913. In either case, a parallax effect would be evident if the face were three-dimensional because of the different perspectives of the face captured in each. For instance, a POSITA would have recognized that, with a front-to-back translation, more of the periphery of the face would be captured by the camera, and there may be other optical effects (e.g., distance-induced distortion) that are more apparent appear in the closer image than the further one. And with side-to-side translation, more features on the side of the face the camera favors would be captured, but features on the other side of the face may be obstructed due to the face's three-dimensionality.

914. Moreover, a POSITA would have understood that Derakhshani's use of the term "rotated" means the camera itself is rotated relative to the face. I have provided an example of rotation below that also includes some side-to-side translation to keep the face centered on the camera.

Rotated



As these exemplary figures demonstrate, however, a POSITA would have understood that, regardless of whether the camera is "slid" or "rotated," distances between facial landmarks and the camera will change. In my opinion, a POSITA would have understood that any of these options results in "capturing at least one second image of the user … at a second distance from the user, the second distance being different than the first distance," as claimed, because there is no one single "distance" between the camera and a three-dimensional user when changing the position/perspective of the camera; some distances will always change. However, even if the claims were limited to a front-to-back translation to change the overall distance between the camera and the user, a POSITA would have understood that Derakhshani discloses or suggests as much.

915. But even if Derakhshani does not expressly disclose taking two images at different distances, in my opinion, a POSITA would have been motivated to look to the differences in degree of distance-induced distortions exemplified by Tanii as an alternative or additional evaluation of the three-dimensionality of the face besides Derakhshani's focus-distance and parallax approaches. §XIII.A.1 (motivation). When making this modification, a POSITA would have been motivated to modify Derakhshani in view of Tanii to expressly capture a second image at a second distance, and look for more distance-induced distortions in one image compared to the other to determine whether the face has depth. §XIII.A.1 (motivation). Moreover, a POSITA would have had a reasonable expectation of success in making this modification because Tanii already taught a mechanism to identify such distance-induced distortions and thus indicate when the face being captured has depth. *See, e.g.*, *Tanii*, [0056]. Accordingly, because Derakhshani, whether alone or combined with Tanii, teaches capturing a plurality of images at different distances between the user and the camera, in my opinion, a POSITA would have understood that either the camera or user must move relative to the other in between image captures; there are no other ways to change the distance between the two.

i. 1[d4]: processing the at least one second image or a portion thereof to create second data;

916. In my opinion, Derakhshani discloses or suggests limitation 1[d4].

917. Derakhshani discloses processing the captured images to identify biometric data "landmarks" in the face as part of the three-dimensional verification process. *Derakhshani*, 17:45-52; §XIII.A.2.g (1[d2]).

j. 1[d5]: comparing the first data to the second data to determine whether expected differences exist between the first data and the second data which indicates three-dimensionality of the user;

918. In my opinion, Derakhshani, alone or in combination with Tanii, teaches 1[d5].

919. Derakhshani discloses that, regardless of whether the focus-distance approach or parallax approach is used, biometric features are identified and compared across each of the images. *Derakhshani*, 16:66-17:2 ("comparing the degree of focus for a landmark in images with different focus distances."); 17:45-64 (evaluating relative displacement of identified landmarks across images). In my opinion, a POSITA would have also appreciated that, when modifying Derakhshani to evaluate for distance-induced distortions exemplified by Tanii, biometric data would also be compared across images to then determine whether they exhibit distance-induced distortion relative to each other, consistent with Derakhshani. In other words, a POSITA would have understood that each of the approaches to evaluate an object's depth taught by Derakhshani (focus distance or parallax), whether alone or combined with Tanii (distance-induced distortion), would require comparing biometric data points across multiple images. For this reason, a POSITA

would have known, or at least been motivated to, match the biometric data between each of the images (such as matching the ears, eyes, and nose in one image to those same features in another) to evaluate the differences between them in different images.

920. Derakhshani discloses that, when comparing the first biometric data to the second biometric data, a determination is made whether differences between the two exist. *Derakhshani*, 16:66-17:2 (for focus distance, "[b]y comparing the degree of focus for a landmark in images with different focus distances, the distance from the sensor to the landmark may be estimated."), 17:55-59 (for parallax, "[i]f all the landmarks in the images undergo the same apparent displacement due to the relative motion of the sensor…then the subject viewed by the camera has a likelihood of being a two-dimensional spoof attack.").

921. In my opinion, Derakhshani describes a comparison between images that looks for "expected" differences consistent with how the '910 Patent uses the term because one would expect that following either the focus-distance or parallax approaches Derakhshani discloses would produce specific differences: the focus-distance approach would capture some images where certain facial features are blurred and others where those same features are clear, and the parallax approach would produce expected relative displacements of certain facial features depending on the change of perspective and distance between the specific features and the

camera lens. Moreover, Derakhshani's focus-distance approach looks for expected differences in the blurriness or clearness of facial landmarks by changing the actual distance (for fixed-focus cameras), and Derakhshani's parallax approach looks for expected differences in the relative displacement of different facial landmarks by changing the actual distance alone. *See* §XIII.A.2.h (1[d3]).

922. Relatedly, a POSITA would have been particularly motivated to configure Derakhshani to capture images with specific, pre-defined configurations (e.g., a specific set of focus distances, or a specific position of the camera relative to the face) to minimize the variability between the images used for facial recognition and specifically tailor the system to look for expected changes between images. For example, Derakhshani's focus-distance approach (with its loss of spatial frequency) would improve its performance if images were acquired with the face at different distances from the camera. Doing this with two or more distances would remove range ambiguity and decrease the variance in estimates of the distance from the camera to particular features. *See, e.g.*, Ex-1018, 32 (noting how facial-recognition systems often require controlling conditions such as a "fixed and simple background with controlled illumination' because "systems … have difficulty in matching face images captured from two different views, under different illumination conditions, and at different times."). In other words, rather than permitting users to change the focus distance or perspective of the camera any way they wish, which would require

a system that could account for such variabilities, having the user follow a pre-determined protocol to capture images at set focus distances or perspectives would simplify the matching process.

923. In my opinion, however, a POSITA would have also understood that, when utilizing the distance-induced distortion approach exemplified by Tanii, the images captured from that process would also exhibit expected distortion based on the distance between the camera and the face. §XIII.A.1 (motivation). In my opinion, a POSITA would have been motivated to look for and utilize these expected differences in distortion as an alternative or supplemental verification of three-dimensionality of a face in Derakhshani, particularly in mobile devices that incorporate wide-angle lenses. §XIII.A.1 (motivation). In doing so, a POSITA would have understood that verifying a three-dimensional face using distance-induced distortion would be accomplished by matching the positions of biometric features across the first and second images—as Derakhshani already discloses—but rather than look for blurriness/clearness or parallax of those biometric features, the images would instead be evaluated for expected differences in the distortion of those features caused by the distance-induced distortion. §XIII.A.1 (motivation).

k. 1[d6]: verifying the images of the user exhibit three dimensional traits when the expected differences exist between the first data and the second data as a result of capturing the at least one first image and the at least one second image at different distances from the user.

924. In my opinion, Derakhshani, alone or in combination with Tanii, teaches 1[d7].

925. Derakhshani discloses a process that determines a face is three-dimensional when expected differences exist between the biometric landmarks (e.g., data) using at least the focus-distance approach. §§XIII.A.2.h (1[d3]), XIII.A.2.j (1[d5]). Specifically, Derakhshani discloses that a face is determined to be three-dimensional when mismatches exist between the biometric landmarks (e.g., biometric data) using either the focus-distance or parallax approach. §§XIII.A.2.h (1[d3]), XIII.A.2.j (1[d5]). The focus-distance approach evaluates whether facial landmarks are blurry in one image and clear in another, indicating depth. §§XIII.A.2.h (1[d3]), XIII.A.2.j (1[d5]). And the parallax approach evaluates whether different facial landmarks are displaced by different amounts, also indicating depth. §§XIII.A.2.h (1[d3]), XIII.A.2.j (1[d5]).

926. If Derakhshani's parallax approach for some reason cannot be considered to already disclose this limitation, however, in my opinion, Derakhshani combined with Tanii does. *See* §§XIII.A.2.h (1[d3]),XIII.A.2.j (1[d5]). Specifically, in my opinion, a POSITA would have appreciated that when modifying Derakhshani

to specifically look for differences caused by distance-induced distortions (consistent with Tanii), a three-dimensional face would be indicated when one of the two sets of biometric data exhibits expected distance-induced distortions due to the change in distance of the camera. *See* §§XIII.A.2.h (1[d3]), XIII.A.2.j (1[d5]). And a POSITA would have been motivated to utilize this expected distortion as an alternative or supplemental verification of three-dimensionality of a face in Derakhshani because it provided a user-friendly way of verifying three-dimensionality using well-understood optical effects common to widely used camera systems (e.g., wide-angle lens in mobile devices). §XIII.A.1 (motivation).

927. Thus, Derakhshani, whether alone or combined with Tanii, looks for mismatches between first and second biometric data to indicate whether a face is three dimensional.

3. Claim 2

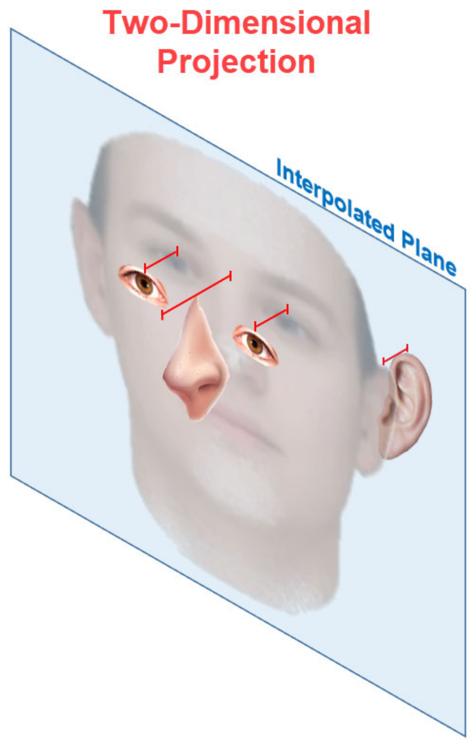
a. 2[a]: The system according to claim 1, further comprising: interpolating the first data and the second data to obtain estimated intermediate data;

928. In my opinion, Derakhshani, alone or combined with Tanii, teaches 2[a]'s additional limitation.

929. Derakhshani discloses that one optional way the three-dimensionality of a face can be verified is by fitting "the location of multiple landmarks…to the

closest two dimensional plane and the average distance of the landmarks from this fit plane can be determined as the spatial metric." *Derakhshani*, 17:12-26.

930. In my opinion, a POSITA would have understood that Derakhshani identifies the position of this two-dimensional plane relative to the facial landmarks by matching up the landmarks that appear across different images, and then calculating an average distance between the various landmarks at their identified three-dimensional positions based on the series of images. If this average distance between the plane and landmarks is sufficiently large, the face is determined to be three dimensional. I have provided a graphic depiction of this process for demonstration purposes, with the plane being identified in blue, the position of the facial landmarks determined by analyzing the series of images, and the red line between the facial landmarks and the plane representing the distance between the two:



931. A POSITA would have therefore understood that the plane constitutes “interpolated” biometric data, because the term “interpolated” is generally understood to mean “to insert between other things” or “estimate values of (data or a function) between two known values. Ex-1026, 654.

932. A POSITA would have further understood that a similar projection could be determined from Derakhshani’s parallax process. However, rather than use a two-dimensional plane projection based on images taken from the same perspective (like Derakhshani’s focus-distance approach), images taken from two different perspectives to evaluate for parallax would be better suited by using a three-dimensional model approach. For instance, it was well-known that three-

dimensional modeling of the perspective of a face could be estimated using images of a face. *See, e.g.*, Ex-1018, 117 (noting use of 2D and 3D modeling techniques to account for variations in perspective); Ex-1027, 8 (describing a method that involves estimating the position of a face to locate and match facial features); Ex-1015 (describing the generation of a three-dimensional model of a face based on two-dimensional images). In fact, Derakhshani expressly recognizes that the spatial metric can be determined by determining deviations between the images captured and a three-dimensional model of the face. *Derakhshani*, 17:27-44.

933. A POSITA would have understood that these three-dimensional models to which images are compared would be an "interpolation." In other words, a POSITA would have recognized, or at least been motivated to implement Derakhshani's parallax approach by constructing a three-dimensional, interpolated model based on the series of images captured to either: (1) compare it to an existing three-dimensional model generated during enrollment; or (2) determine whether the series of images can create a suitable three-dimensional model, which itself would indicate that the imaged face has three dimensions.

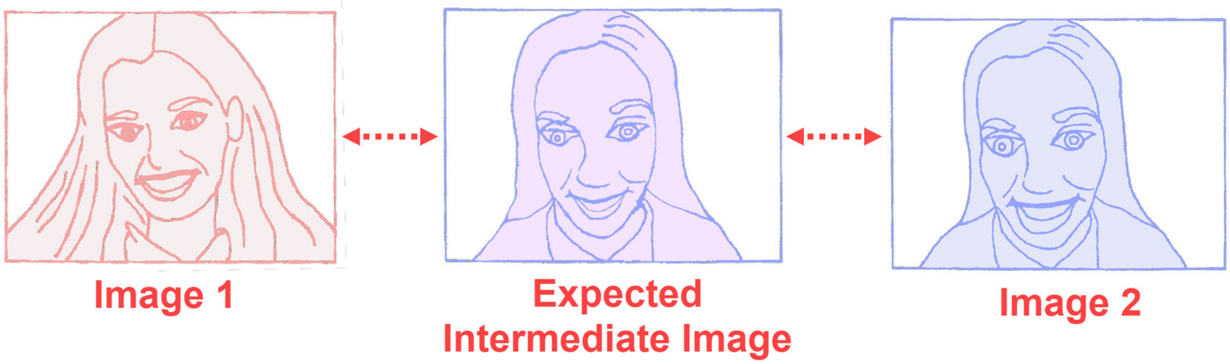
934. Based on a POSITA's understanding of Derakhshani, a POSITA would have further been motivated to derive interpolated data based on the combination of Derakhshani and Tanii. Specifically, Tanii discloses that distance-induced distortions increase as distance between the face and camera decreases. *See Tanii*,

[0048]. A POSITA would have therefore understood that, all else being equal, distance-induced distortion depends on the distance between the user and the camera, and thus any set of images as the distance between the user and camera changes will have different degrees of facial warping, similar to the gradual changes in facial distortion that appears in the series of images below (although these images also altered the focal length of the camera to ensure the face remains a constant size in the frame, rather than just distance:

Ex-1022.

935. A POSITA reading Derakhshani—which discloses generating intermediate projections to evaluate depth—in view of Tanii therefore would have been motivated to interpolate intermediate data with an intermediate, interpolated amount of distance-induced distortion based on the two images captured to create an array of potential distance-induced distortions that would indicate depth in a three-dimensional face, such a gradual projection from the lines of the facial landmarks

shown in blue (the distorted image) to the lines of the facial landmarks shown in red (the distorted image).



Performing this type of interpolation between the distorted and undistorted images would allow for further comparison with additional images, for instance, to ensure the distance-induced distortion matches expectations across a range of distances if the face were truly three-dimensional, as depicted below:

In my opinion, a POSITA would have understood that this would be akin to Derakhshani's modeling approach, but rather than build a model based on a two-dimensional projected plane or three-dimensional model of a head, the model would be of various degrees of expected distance induced distortion with which the captured images could be compared.

b. 2[b]: capturing at least one third image of the user taken with the camera of the computing device at a third distance from the user, the third distance being between the first distance and the second distances;

936. In my opinion, Derakhshani, alone or combined with Tanii, teaches 2[b]'s additional limitation.

937. Derakhshani discloses that, as part of the three-dimensional verification process, "a plurality" of images may be captured. *Derakhshani*, 16:44-46 (focal-distance embodiment), 17:45-47 (parallax embodiment).

938. In my opinion, a POSITA would have also understood generally that capturing more images would provide increased accuracy in verifying a three-dimensionality of the face because there would be more samples to evaluate, with the trade-off being an increase in processing demands. For instance, taking four images using the focus-distance approach would enable precise depth information of at least four facial landmarks that sit on different planes, such as the ears, eyes, mouth, and nose. For the parallax approach, fewer images would likely be necessary depending on how significant the change of perspective is—e.g., rotating the camera may reveal parallax in as little as two images, whereas sliding may benefit from an additional image—since that approach looks for displacement of facial landmarks

due to the change in perspective, which does not depend on taking images at different "slices" of depth like the focus-distance approach.

939. A POSITA would have also appreciated as a general matter that, in any set of images with more than two images captured at different distances, *see* §XIII.A.2.h (1[d3]; capturing images at different distances), one would have a minimum distance and one would have a maximum, with the rest existing in the range in between. For instance, in a set of distances of 10cm, 50cm, and 1m, 10cm would be the minimum, and 1m would be the maximum, with 50cm existing in between the two.

940. Moreover, when modifying Derakhshani in view of Tanii to interpolate intermediate biometric data attributable to distance-induced distortions, a POSITA would have been further motivated to capture a third image at a distance that correlates to one of the interpolated data sets for further confirmation of distortion (and therefore authentication) of three-dimensional depth of the face in the captured images. *See* §§XIII.A.2.h (1[d3]), XIII.A.3.a (2[a]).

c. 2[c]: processing the at least one third image or a portion thereof to obtain third data; and

941. In my opinion, Derakhshani, alone or combined with Tanii, teaches 2[c]'s additional limitation.

942. Derakhshani discloses processing the images to identify feature landmarks in each of the images, *see* Sections §§XIII.A.2.g (1[d2]), XIII.A.2.i

(1[d4]). Therefore, in my opinion, a POSITA would have found it obvious to process any images captured by the camera to derive biometric data so that the biometric data could be compared between images, consistent with Derakhshani.

943. Furthermore, when modifying Derakhshani in view of Tanii, a POSITA would have found it obvious to acquire a third image and extract biometric data from the third image to compare it to the interpolated positions of the biometric data based on the first and second images. *See* §XIII.A.3.b (2[b]).

d. 2[d]: comparing the estimated intermediate data with the third data to determine whether the third data matches the estimated intermediate data.

944. In my opinion, Derakhshani, alone or combined with Tanii, teaches 2[d]'s additional limitation.

945. Derakhshani discloses comparing multiple images to the two-dimensional projection interpolated from those images. *Derakhshani*, 17:12-26; §XIII.A.3.a (2[a]). In my opinion, a POSITA would have therefore understood that once an interpolated projection or model of the face is generated consistent with Derakhshani, if a third image is captured, that too would be compared to the projection or model to estimate the distance or deviation of any facial landmarks in that image from the projection or model. §XIII.A.3.b (2[a]).

946. Furthermore, in my opinion, a POSITA Derakhshani in view of Tanii, a POSITA would have found it obvious to acquire a third image and extract data

from the third image to compare it to the interpolated positions of the data based on the first and second images to determine if there is a match between the two. *See* §§XIII.A.3.a, (2[a]), XIII.A.3.b (2[b]).

4. Claim 3: The system according to claim 1, further comprising verifying the presence of one or more features on a side of a user's head in the at least one first image, and verifying the absence or reduced visibility of the one or more features on the side of the user's head in the at least one second image due to image capture at different distances from the user's head, wherein the first distance is larger than the second distance.

947. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 3's additional limitation.

948. Derakhshani discloses that "a landmark…an ear…may be identified and located." *Derakhshani*, 16:51-54; *see also* 17:14-19.

949. In my opinion, a POSITA would have appreciated that, when following Derakhshani's focus-distance approach, in some captured images, the ear would have reduced visibility (i.e., it is blurry) when it does not lie in the focal plane, and would be clear (e.g., a verified presence) when it does lie in the focal plane. *See* §XIII.A.2.h (1[d3]). In my opinion, a POSITA would have appreciated that an ear is a "feature[] on a side of a user's head" as claimed. And POSITA would have also appreciated that, in some circumstances, distances in which the ear would be clear would be greater than those with reduced visibility, such as when the focal plane is aligned behind the ears.

950. And when following Derakhshani's parallax approach, a POSITA would have also appreciated that some perspectives would obviously capture one or more ears (when both are exposed, such as a front-facing image from sufficient distance), and other perspectives would only capture one (when the other is obstructed by the head), which would indicate that the user's face is three-dimensional. *See* §XIII.A.2.h (1[d3]) (providing an example figure in which a camera rotation would obfuscate one ear). If the object being captured were a two-dimensional picture of a face with ears, however, any perspective would capture both ears because the ears exist on a single plane of the picture.

951. Similarly, when modifying Derakhshani in view of Tanii to use distance-induced distortions to verify the three-dimensional nature of a face, Tanii teaches that the absence and presence of an ear is a natural result of the distance between the user and camera. Specifically, when a sufficient distance between the face and camera exists, the ears are captured because there is enough distance for the light rays from the ears to strike the camera's lens. *See Tanii*, [0048], Figs. 3A-3B.

Fig.3A

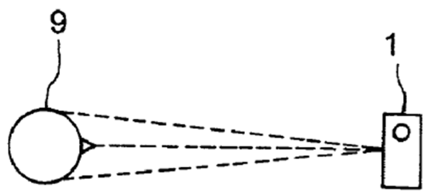
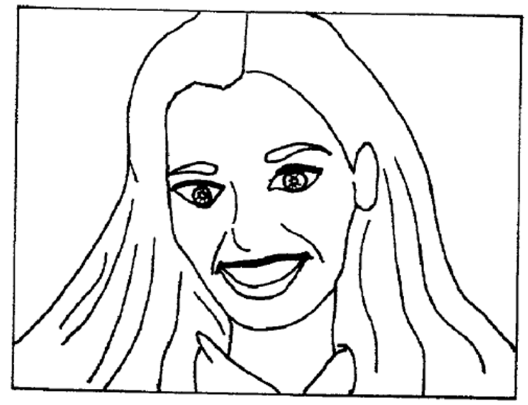


Fig.3B



But when a face is too close, the ears will not be captured because there is insufficient distance for the light rays from the ears to strike the camera's lens. *See Tanii*, [0047]-[0048], Figs. 4A-4B.

Fig.4A

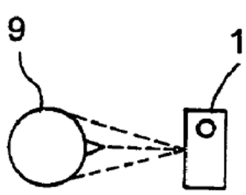
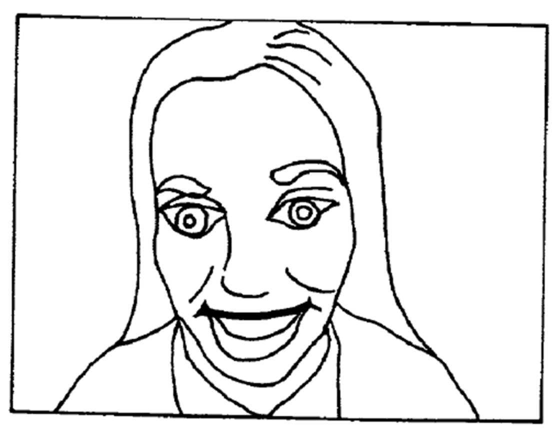
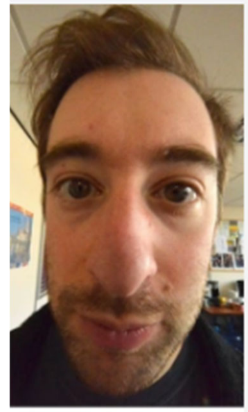


Fig.4B



This effect was well known and demonstrated in actual applications, as shown below.

85mm @ 200cm



8mm @ 20cm

952. In my opinion, therefore, a POSITA would have appreciated based on at least Tanii that the presence of a user's ears in one image at a sufficient distance, but absence of a user's ears in another image at a closer distance would be indicative of a three-dimensional face, and would have been motivated to modify Derakhshani to verify the presence and absence of the ears between images as yet another indicator of the three-dimensional nature of the face.

5. Claim 4: The system according to claim 1, wherein the machine readable instructions is configured to display one or more prompts on the screen of the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

953. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 4's additional limitation.

954. Derakhshani discloses that the invention can be implemented in computing devices such as a "smart phone, a tablet computer, a television, a laptop computer, or a personal computer," *Derakhshani*, 5:22-27, which incorporate a

camera, *id*., 5:23-27, 6:3-10, and a display. *Id*., 6:8-11, 9:22-24, 14:35-37, 22:33-38, 25:9-15. Derakhshani discloses displaying prompts to a user to guide the user to capture images of the user's face for authentication, *Derakhshani*, 5:23-32, 6:8-16, 9:22-26, including at more than once distance, *id.*, 17:64-66; §XIII.A.2.h (1[d3]).

955. But even if Derakhshani does not expressly disclose taking two images at different distances, doing so would have been obvious in view of Tanii to identify distance-induced distortions that indicate depth of a three-dimensional face. §XIII.A.2.h (1[d3]). When modifying Derakhshani to look for distance-induced distortions consistent with Tanii by capturing images at different distances, a POSITA would have found it obvious to also provide prompts to a user to ensure the images are captured at the correct distances, as Derakhshani already discloses providing prompts to correctly orient the user relative to the camera.

6. Claim 5

a. 5[a]: The system according to claim 1, further comprising comparing the first data, second data, or both to enrollment data derived from an enrollment image, the enrollment image captured and stored prior to an authenticating; and

956. In my opinion, Derakhshani discloses or suggests 5[a]'s additional limitation.

957. Derakhshani discloses capturing and analyzing multiple images of a user and comparing the user's features to a previously stored "reference record" to

authenticate the user. *Derakhshani*, 4:19-24; 7:20-34; 8:60-64; 9:31-34. In my opinion, a POSITA would have understood the "reference record" to be "enrollment data" because the process Derakhshani describes to generate and then use the "reference record" for authentication is consistent with a typical biometric-authentication enrollment procedure. *See* §V.A (biometric security overview). Specifically, Derakhshani discloses that the system captures one or more initial reference images of the user during a registration process, extracts features from the reference images, stores the extracted features as the reference record, and then subsequently compares later-captured images to the reference record. *Id*., 7:19-34 ("To create a reference record *for a new user* and enrollment or registration process may be carried out."); 9:31-34 ("The collection of image data from user may also facilitate authentication against a reference record for a user identity."); 13:62-14:9 (describing authentication matching against a reference record). Then, during the authentication process, Derakhshani compares the extracted features from the captured images (i.e., portions of the first data, second data, or both) to the user's enrollment reference record to determine a match score. *Id*. 9:59-67; 13:62-14:9; 17:32-36. This is consistent with a conventional biometric enrollment and authentication process. *See, e.g.*, Ex-1018, 4-11 (providing overview of biometric authentication and verification).

b. 5[b]: only authenticating the user when the first data, the second data, or both match the enrollment data within a predetermined threshold.

958. In my opinion, Derakhshani discloses or suggests 5[b]'s additional limitation.

959. Derakhshani discloses calculating a match score during the authentication process based on the comparison of features extracted from the first and second image to the corresponding features in an enrollment reference record. *Derakhshani*, 13:62-14:9. Derakhshani also discloses that, only when the match score is above a threshold—because the first or second data, or both, sufficiently correspond to the enrollment data—it is determined the user is authenticated. *Id.*, 14:25-35. This is consistent with conventional biometric-authentication processes. *See, e.g.*, Ex-1018, 17 ("[B]iometric systems mostly decide on a person's identity based on a *close* match between the template and the query, where the strength of the match (or the degree of similarity) is represented by the match score."), 18 ("a verification system makes a decision by comparing the match score s to a threshold η").

7. Claim 6: The system according to claim 1, wherein the computing device is a hand-held device, and the user holds the device at the first and second distance to capture the at least one first image and the at least one second image.

960. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 6's additional limitation.

961. Derakhshani discloses that the biometric-authentication process can be implemented on a variety of different type of hand-held computing devices, such as "a laptop computer, a handheld computer…, a tablet computing device, a personal digital assistant (PDA), a cellular telephone…, a camera, a smart phone," and more. *See, e.g.*, *Derakhshani*, 8:11-28, 18:1-4. Moreover, Derakhshani recognizes that, to verify three-dimensionality of the face, "a single camera may be rotated or slide slightly," or that, when the device is hand-held, "the [camera] sensor may naturally move relative to the users face due to involuntary haptic motion" that may sufficiently to capture a parallax effect. *Id.*, 17:59-18:4. Moreover, Tanii notes that distance-induced distortions often occur in mobile devices that have incorporated wide-angle lenses, and the amount of distortion is dictated by the distance between the user and the camera. *Tanii*, [0007], [0047]-[0048], Figs. 3A-B, 4A-B.

962. When implementing a three-dimensional verification process on a handheld mobile computing device consistent with Derakhshani, alone or in combination with Tanii (*see* §§XIII.A.2.h (1[d3]), XIII.A.5 (claim 4)), it is my opinion that a POSITA would have further understood that the user would hold the computing device at a first distance for the first image, and a second distance for the second image (e.g., by extending and retracting the user's arm), because that is a convenient and obvious way of changing the distance between a hand-held device and the user's face, and because Derakhshani already envisions evaluating depth

based on displacement of the user's arm holding the device. §VII.A (Derakhshani); *Derakhshani*, 16:44-11, 17:45-18:4.

8. Claim 7: The system according to claim 1, wherein the first data and the second data comprise biometric data.

963. In my opinion, Derakhshani discloses or suggests claim 7's additional limitation for the reasons discussed in §XIII.A.2.g (1[d2]); §XIII.A.2.j (1[d5]).

964. In my opinion, a POSITA would have also understood that data created from an image (or portion of an image) of a user's face would comprise biometric data. *Id.*

9. Claim 8: The system according to claim 1, wherein the first data and the second data comprise a mapping of facial features.

965. In my opinion, Derakhshani discloses or suggests claim 8's additional limitation.

966. Derakhshani discloses processing the captured images to identify and locate facial biometric "landmarks" (e.g., an iris, an eye corner, a nose, a mouth, an ear) in a three-dimensional verification process. *Derakhshani*, 16:44-54. In my opinion, a POSITA would have understood that the identification of facial landmarks would include their locations relative to one another, thus constituting a mapping of facial features. In fact, processing image data to map facial features was a conventional aspect of facial-recognition systems. *See, e.g.*, Ex-1018, 103 (Fig. 3.5(b) describing how "Level 2 features require detailed processing for face recognition.

Information regarding the structure and the specific shape and texture of local regions in a face is used to make an accurate determination of the subject's identity."). Moreover, a POSITA would have understood that the process of converting facial features in an image to computer-readable data conventionally involves mapping those features to data. *See, e.g.*, *id.*, 116-17 (noting how "appearance-based techniques generate a compact representation of the entire face region in the acquired image by *mapping* the high-dimensional face image into lower dimensional sub-space.").

10. Claim 9: The method according to claim 1, wherein the first image and the second image is of the user's face and the user's head and facial features are held steady and without movement during capture of the first image and the second image.

967. Derakhshani, alone or combined with Tanii, teaches claim 9's additional limitation.

968. Derakhshani and Tanii both teach or suggest moving the camera to capture images at two different distances. *See* §XIII.A.2.h (1[d3]).

969. In my opinion, a POSITA would have understood that, when moving the camera to capture images from different distances, the user's face would be stationary (e.g., steady) both during each image capture (to ensure each image is not blurry due to the camera's exposure time) and during movement of the camera (to isolate any differences between images to those attributable to the change of

distance). Moreover, a POSITA would have appreciated that holding the user's face steady and moving the camera closer and further away would be more user friendly than requiring the user to move their head closer and further from the camera while holding the camera steady.

970. Furthermore, in my opinion, a POSITA would have further understood that, when capturing an image of a user's face, the image would be of the user's head because the face exists on the head.

11. Independent Claim 10

a. 10[pre]: A method for evaluating three-dimensionality of a user, the method comprising:

971. If the preamble is limiting, in my opinion, Derakhshani discloses or suggests it.

972. Derakhshani discloses systems and methods for using a camera-equipped computing device for "biometric authentication." *See Derakhshani*, 1:11-25, 5:22-27, 6:3-5, 9:10-22, 18:1-3. Although Derakhshani uses the eye as the primary means of authentication, *see, e.g.*, *id.*, Abstract, as part of the ocular-authentication process, Derakhshani also verifies that the user's face is three-dimensional by capturing multiple images of a user's face at different focus distances or from different perspectives to calculate a "spatial metric" representing the face's three-dimensionality. *Id.*, 1:11-25, 3:14-15, 16:44-18:4. *See* §XIII.A.2.a (1[pre]).

b. 10[a]: capturing at least one first image of the user taken with a camera at a first location which is a first distance from the user;

973. In my opinion, Derakhshani discloses or suggests limitation 10[a] for the reasons discussed in §XIII.A.2.f (1[d1]).

c. 10[b]: processing the at least one first image or a portion thereof to create first data;

974. In my opinion, Derakhshani discloses or suggests limitation 10[b] for the reasons discussed in §XIII.A.2.g (1[d2]).

d. 10[c]: moving the camera from the first location to a second location, the second location being a second distance from the user, or the user moving to change the distance between the user and the camera from the first distance to the second distance;

975. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 10[c] for the reasons discussed in §XIII.A.2.h (1[d3]).

e. 10[d]: capturing at least one second image of the user taken with the camera when the camera is the second distance from the user, the second distance being different than the first distance;

976. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 10[d] for the reasons discussed in §XIII.A.2.h (1[d3]).

f. 10[e]: processing the at least one second image or a portion thereof to create second data;

977. In my opinion, Derakhshani discloses or suggests limitation 10[e] for the reasons discussed in §XIII.A.2.i (1[d4]).

g. 10[f]: comparing the first data to the second data to determine whether expected differences exist between the first data and the second data which indicate three-dimensionality of the user;

978. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 10[f] for the reasons discussed in §XIII.A.2.j (1[d5]) (describing comparison to look for expected differences), XIII.A.2.k (1[d6]) (describing expected differences as distorting changes from Derakhshani's focus-distance approach and Derakhshani-Tanii's distance-induced distortion approach).

h. 10[g]: verifying the images of the user exhibit three-dimensional traits when the first data and the second data have expected differences resulting from the at least one first image being captured with the camera at a different distance from the user than when the at least one second image is captured.

979. In my opinion, Derakhshani, alone or in combination with Tanii, teaches limitation 10[g] for the reasons discussed in §XIII.A.2.k (1[d6]).

12. Claim 11

a. 11[a]: The method according to claim 10, further comprising: interpolating the first data and the second data to obtain estimated intermediate data;

980. In my opinion, Derakhshani, alone or in combination with Tanii, discloses or suggest limitation 11[a] for the reasons discussed in §XIII.A.3.a (2[a]).

b. 11[b]: capturing at least one third image of the user taken with the camera at a third distance from the user, the third distance being between the first distance and the second distances;

981. In my opinion, Derakhshani, alone or in combination with Tanii, discloses or suggest limitation 11[b] for the reasons discussed in §XIII.A.3.b (2[b]).

c. 11[c]: processing the at least one third image or a portion thereof to obtain third data; and

982. In my opinion, Derakhshani, alone or in combination with Tanii, discloses or suggest limitation 11[c] for the reasons discussed in §XIII.A.3.c (2[c]).

d. 11[d]: comparing the estimated intermediate data with the third data to determine whether the third data matches the estimated intermediate data.

983. In my opinion, Derakhshani, alone or in combination with Tanii, discloses or suggest limitation 11[d] for the reasons discussed in §XIII.A.3.d (2[d]).

13. Claim 12: The method according to claim 10, further comprising verifying the presence of ears of the user in the at least one first image, and verifying the absence or reduced visibility of the ears in the at least one second image, wherein the first distance is larger than the second distance.

984. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 12's additional limitations for the reasons discussed in §XIII.A.4 (cl.3).

14. Claim 13: The method according to claim 10, further comprising one or more prompts on a screen to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

985. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 13's additional limitations for the reasons discussed in §XIII.A.5 (cl.4).

15. Claim 15: The method according to claim 10, wherein the camera is part of a computing device is a hand-held device, and the user holds the computing device at the first distance when capturing at least one first image and at the second distances when capturing the at least one second image.

986. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 15's additional limitations for the reasons discussed in §XIII.A.7 (cl.6).

16. Claim 16: The method according to claim 10, wherein the first data and the second data comprise biometric data.

987. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 16's additional limitation for the reasons discussed in §XIII.A.8 (cl.7).

17. Claim 17: The method according to claim 10, wherein the first data and the second data comprise a map of facial features.

988. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 17's additional limitation for the reasons discussed in §XIII.A.9 (cl.8).

18. Claim 18: The method according to claim 10, further comprising illuminate a screen of a computing device while capturing the at least one first image and/or the at least one second image, and processing the at least one first image and/or the at least one second image to detect a reflection of the illumination from a face of the user.

989. In my opinion, Derakhshani discloses or suggests claim 18's additional limitation.

990. Derakhshani discloses, in addition to a "spatial metric," calculating a separate "reflectance metric" that measures changes in surface glare on the eye due to changes in a light source such as the illumination of the screen to further verify the "liveness" of the user's face. *Derakhshani*, 18:8-19 ("The reflectance metric may be a measure of changes in glare or specular reflection patches… [from] a dynamic light source (e.g.,… LCD screen…."). In my opinion, a POSITA would have understood that the "LCD screen" would obviously be the screen of the computing device. Derakhshani also discloses processing images of the user to detect a reflection of the illumination from a face of the user. *Id.* ("a reflectance metric is determined 716 based on detected change in surface glare or specular reflection patterns on a surface of the eye as the eye appears in a plurality of images."). In my opinion, a POSITA would have therefore understood that illuminating the screen while capturing the first and second images, and measuring the reflectance from the illumination in the captured images, serves as another liveness measure to ensure the user's face is from a real user, and not a spoofer.

19. Claim 19: The method according to claim 10, wherein a face of the user is held steady when capturing the at least one first image and the at least one second image and the camera moves from the first location to the second location.

991. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 19's additional limitation for the reasons discussed in §XIII.A.10 (cl.9).

20. Claim 20: The method according to claim 10, wherein the first data and the second data are maintained on a computing device.

992. In my opinion, Derakhshani discloses or suggests claim 20's additional limitation.

993. Derakhshani discloses that the biometric-authentication process can be performed locally on the device, on a server, or split between the two. *See, e.g.*, *Derakhshani*, 9:27-58, 10:1-24. In my opinion, a POSITA would have understood that Derakhshani's three-dimensional verification—which is part of the biometric-authentication process—would be configured in some circumstances to perform locally on the device. In such instances, in my opinion, a POSITA would have understood that the biometric data would be maintained on the computing device so that all processes would be performed using a single device. If a single, local device is being used for all authentication procedures, there would be no need to store any data on a separate device. Moreover, a POSITA would have understood that biometric data is extremely sensitive, and transmitting that data over networks

presents a security risk of being intercepted. A POSITA would have understood that there would be no need to take any security risks by transmitting such data if the single, local device is handling the entire biometric-authentication procedure.

21. Claim 21: The method of claim 10 wherein the camera is part of is one of a smartphone, tablet, laptop, or desktop computer.

994. In my opinion, Derakhshani, alone or in combination with Tanii, teaches claim 21's additional limitation.

995. Derakhshani discloses that images "may be captured with a sensor (e.g., a camera) that is integrated into a computing device such as, for example, a smart phone, a tablet computer, a television, a laptop computer, or a personal computer." *Derakhshani*, 5:22-27; *see also, e.g.*, 9:18-22, 18:1-4.

22. Independent Claim 22

a. 22[pre]: A method, performed by a user using a user's computer device, for verifying three-dimensionality of the user, the method comprising:

996. If the preamble is limiting, in my opinion, Derakhshani discloses or suggests it for the reasons discussed in §XIII.A.2.a (1[pre]).

b. 22[a]: capturing a first image of the user's head with a camera at a first distance from the user, the camera associated with the user's computing device;

997. In my opinion, Derakhshani discloses or suggests limitation 22[a] for the reasons discussed in §XIII.A.2.f (1[d1]).

c. 22[b]: changing a distance between the user and the camera to a second distance by the user moving the camera, or the user moving relative to the camera, or both;

998. In my opinion, Derakhshani discloses or suggests limitation 22[b] for the reasons discussed in §XIII.A.2.h (1[d3]).

d. 22[c]: capturing a second image of the user's head with the camera when the camera is at the second distance from the user, the second distance being different than the first distance;

999. In my opinion, Derakhshani discloses or suggests limitation 22[c] for the reasons discussed in §XIII.A.2.h (1[d3]).

e. 22[d]: comparing one or more aspects of the user's head from the first image to one or more aspects of the user's head from the second image to determine whether expected differences, between the first image and the second image, exist which indicates three-dimensionality of the user, such that the expected differences between the first image and the second image result from the first image being captured when the camera is at a different distance from the user than when the second image is captured; and

1000. In my opinion, Derakhshani discloses or suggests limitation 22[d] for the reasons discussed in §§XIII.A.2.j (1[d5]), XIII.A.2.k (1[d6]).

f. 22[e]: responsive to the comparing determining that expected differences between the first image and the second image exist, providing notice to the user, a third party, or both that the three-dimensionality of the user is verified.

1001. In my opinion, Derakhshani discloses or suggests limitation 22[e].

1002. Derakhshani discloses that, when an authentication attempt is rejected as a spoof attempt, the system provides notice of the rejection to the user or a third party, and that the authentication is one of at least liveliness and/or three-dimensionality. *Id*., 8:67-9:4; 11:17-26. Separately, Derakhshani discloses providing notice to a user that their identify has been authenticated. *Derakhshani*, 14:45-58 (explaining the user may be "informed of the acceptance 632 through a message that is shown on a display or played through a speaker…[or] transmit[ed]…through a network.").

1003. Although Derakhshani does not expressly disclose notifying the user when their face has been verified as three-dimensional, in my opinion, providing such intermediate, additional notice would have been obvious to a POSITA. For instance, if the biometric authentication system comprised separate liveness evaluations performed in series, as Derakhshani discloses (such as behavioral, spatial, and reflectance), (*Derakhshani*, 19:46-48; *see also generally id.* 15:26-19:48) a POSITA would have been motivated to provide notice after each separate evaluation to indicate the next had begun. Implementing such notifications would have involved routine computer coding, since Derakhshani already envisions providing notifications during the authentication process.

23. Claim 23: The method of claim 22 wherein the one or more aspects of the user's head from the first image is first data resulting from processing the first image and the one or more aspects of the user's head from the second image is second data resulting from processing the second image.

1004. In my opinion, Derakhshani, alone or in combination with Tanii, teaches this claim for the reasons discussed in §§XIII.A.2.g (1[d2]), XIII.A.2.i (1[d4]).

24. Claim 24: The method of claim 22 wherein the user's head is the user's face.

1005. In my opinion, Derakhshani, alone or in combination with Tanii, teaches this claim for the reasons discussed in §§XIII.A.2.g (1[d2]), XIII.A.10 (cl.9).

B. Ground 1B: Derakhshani, Tanii, and Tahk (Claim 14)

1. Motivation to Combine

1006. In my opinion, a POSITA would have been motivated to modify Derakhshani, with or without Tanii, in view of Tahk because Tahk provides a user-friendly way of ensuring that a face presented for facial authentication is properly framed. In my opinion, a POSITA would have understood that providing a real-time preview of what an image would look like prior to capturing the image, as well as providing express prompts—such as written instructions or oval shapes on the live-preview screen—as taught by Tahk, §VII.D (Tahk), would have been particularly useful for authentication procedures that require capturing multiple images of a face, as taught by Derakhshani and Tanii.

1007. For instance, providing a user a real-time preview would allow the user to actively adjust the position of the camera and/or their orientation of their face to properly frame their face for the image capture. Moreover, providing express prompts and oval shapes sized to guide the user to properly position their face would ensure images best suited for facial recognition can be captured, and that the faces would be captured from sufficiently different perspectives and/or distances to ensure the three-dimensional verification taught by Derakhshani, alone or in combination with Tanii, could be performed. This is particularly important for facial-recognition systems, which generally are known to have difficulty matching faces across different views. *See, e.g.*, Ex-1018, 32.

1008. Moreover, a POSITA would have known that providing real-time image feedback, written instructions, and oval shapes to frame a face during a facial authentication process were all well-known and conventional techniques to provide user feedback during image capture as of the time of the invention. *See, e.g.*, Ex-1034, 7:16-8:7, Figs. 6B-7C (providing "an example of an interface used upon registering a facial image of a person to be authenticated" in which "the image of this person is displayed on a monitor" during registration, and oval-shaped prompts to indicate the size/distance of the user from the face authentication sensor); Ex-1035, 5:31-32 ("The computing device may present prompts that instruct the user to perform one or more liveness gestures"), 6:3-4 (same).

2. Claim 14: The method according to claim 13, wherein the one or more prompts are an oval shape guide on the screen within which an image of a face of the user is aligned to capture the at least one first image and the at least one second image.

1009. In my opinion, Derakhshani, combined with Tanii and/or Tahk, teaches claim 14's additional limitation.

1010. As I have previously explained, Derakhshani, alone or in combination with Tanii, teaches providing prompts to user to properly frame themselves at different distances to capture images for biometric authentication. *See* §XIII.A.14 (claim 13). But Derakhshani and Tanii do not expressly describe using oval-shaped prompts to guide a user during the facial-authentication process.

1011. In my opinion, however, a POSITA would have been motivated to provide such oval-shaped prompts (as well as express written instructions) in view of Tahk. *See, e.g.*, *Tahk*, Figs. 8A-B ("Please step further back" and "Please step further forward," and presenting an oval to frame the face at the correct distance). A POSITA would have been motivated to modify Derakhshani, alone or in combination with Tanii, to provide such oval-shaped prompts because they are a natural shape to appropriately size and frame a face at different distances. *See* §VII.D (Takh); §XIII.B.1 (motivation).

C. Ground 2A: Zhang and Tanii (Claims 1-3, 6-12, 15-17, 19-24)

1. Motivation to Combine

1012. In my opinion, a POSITA would have been motivated to combine Zhang and Tanii because both concern identifying and accounting for the three-dimensional nature of a face when capturing an image. They differ, however, in what principles are used to account for the face's three-dimensionality. Zhang, for instance, looks to dissimilarities in two images after one undergoes a mathematical homography. *See* §VII.C (Zhang). And although Tanii is not expressly directed to *evaluating* whether a face has depth like Zhang, Tanii exemplifies the well-known distortions caused by the interaction between the camera's lens and the three-dimensional nature of the face, *see* §VII.B (Tanii). A POSITA would have appreciated, therefore, that Tanii recognizes another clear alternative to evaluating the depth of a face, consistent with Zhang's existing homography transformation.

1013. A POSITA would have recognized, as Tanii does, that distance-induced distortions occur because of the interactions between the shape of the camera lens and shape of the face, and the distortion in part depends on the distance between the face and the camera. §VII.B (Tanii); *Tanii*, [0048]. Accordingly, a POSITA would have understood from Tanii that, by taking two images from two different distances, a larger amount of distortion in the closer of the two images indicates whether a face is three-dimensional or not.

1014. In my opinion, a POSITA would have therefore appreciated from Tanii that images captured by Zhang—without any modification—may exhibit distance-induced distortions based on the particular camera used to perform Zhang's process (e.g., particularly when a wide-angle lens with significant barrel distortion is used, as is common in computers and mobile devices). However, a POSITA would have also appreciated that any distance-induced distortions would further enhance Zhang's homography-transformation process because a homography transformation cannot correct for these distortions.

1015. For instance, if a homography transformation were applied to Tanii's Figure 4B (serving as Zhang's "first image") to compare to Figure 3B (serving as Zhang's "second image"), the transformation would not account for differences between the images caused by the distance-induced distortion.

Fig.4B

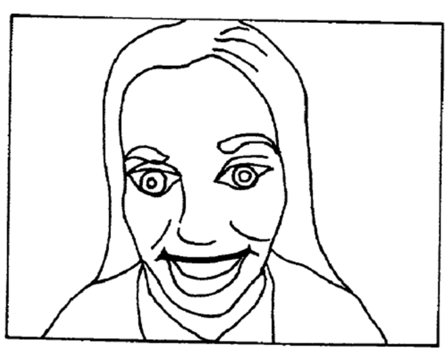
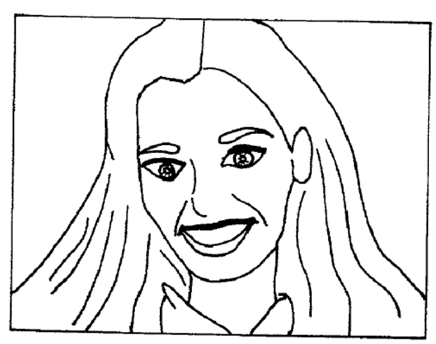


Fig.3B



Tanii, Figs. 3B, 4B. That is because Zhang relies on a mathematical principle that enables transforming the *perspective* of a planar object, such as a photograph being used to spoof the authentication procedure to a different *perspective*, §VII.C (Zhang), whereas the distortion identified by Tanii is *radial* and a byproduct of the lens' imperfections and the change in magnification with distance. A homography transform does not account for such radial distortions, but would instead transform the perspective of Tanii with its distortions intact. In other words, in a transformation of perspective with a three-dimensional object such as a real face, Tanii's distance-induced distortions would remain. Ultimately, however, when comparing the two images once one is transformed into the perspective of the other, there would remain differences attributable to the distance-induced distortion which, in my opinion, a POSITA would have understood would result in Zhang identifying the face as three-dimensional.

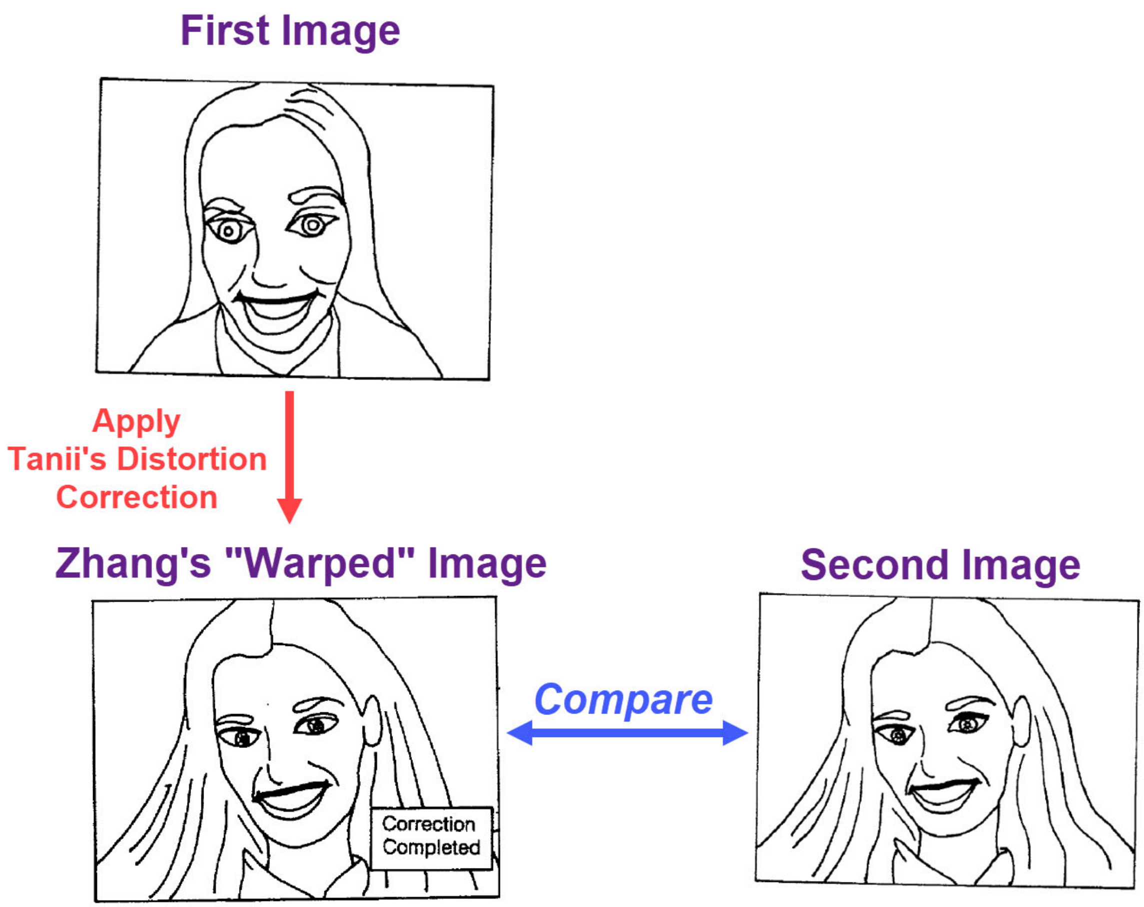
1016. In my opinion, a POSITA would have therefore recognized that Zhang's existing process would be *enhanced* by prompting a user to capture two images and two distances—one of which would have increased distance-induced distortion—because if the face were three-dimensional, Zhang's existing procedure would identify the two images as different and indicate a three-dimensional face. The lack of a match between the two images would likely be enhanced by changes in radial distortion: it makes them even less like data from two planar objects which would produce a match.

1017. However, in my opinion, a POSITA would have also been motivated to modify Zhang's process in view of Tanii in either of two additional ways.

1018. First, in my opinion, a POSITA would have been motivated to follow Zhang to verify the three dimensionality of a face during a facial authentication procedure by taking two or more images, applying mathematics to one of the images, and comparing the mathematically altered image to a second (unaltered) image. But instead of the mathematics applied being a homography transformation, in my opinion, a POSITA would have been motivated to *substitute* Zhang's mathematics for those taught by Tanii to correct for distance-induced distortion. In other words, rather than change the perspective of one image to match the second image, a POSITA would correct the distortion of one image (to create what Zhang refers to as its "warped" image[7]) and compare the result to another image taken further away

[7] Zhang and Tanii both use the term "warped" to refer to different effects, but they are not inconsistent with one another. Specifically, Zhang uses the term "warped" to refer to the resulting image that has undergone homography transformation because the original relationship between the pixels in the image are modified. Tanii uses the term "warped" to refer to the distortions in an image of a face induced by the image-capture conditions (e.g., distance and lens geometry). When I refer to Zhang's

that does not exhibit the same degree of distance-induced distortions.



Tanii, Figs. 3B, 4B, 9.

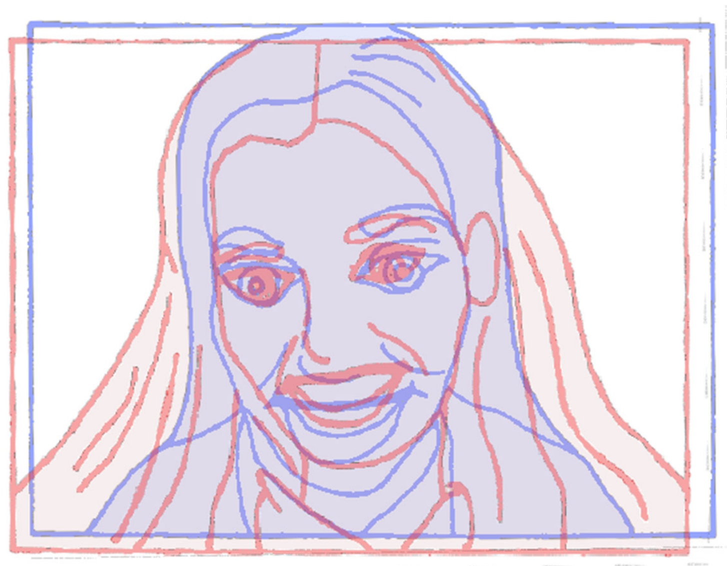
1019. A POSITA would have appreciated that if the "warped" (distortion-corrected) image and second image are sufficiently similar, that indicates a three-dimensional face because Tanii is correcting for distortions attributable the three-

---

"warping," I am referring to the result of a mathematical application to an image; and when I refer to Tanii's warping, I am referring to distance-induced distortion.

dimensionality of the user's face. By following this approach, a POSITA would have recognized that the only difference (besides the mathematics) is that the comparison between the Zhang-Tanii "warped" (distortion-corrected) image would look for a match with the second image.

1020. Alternatively, a POSITA would have appreciated that Zhang and Tanii could be modified to eliminate the mathematical transformation of a first image entirely. Once again, a POSITA would have been motivated to follow Zhang to verify the three dimensionality of a face during a facial authentication procedure by taking two or more images, but rather than apply mathematics to "warp" one of the images (e.g., using either a homography transform or distortion-correction procedure), the facial features would be mapped in each image, matched between the two images, and evaluated to determine whether differences attributable to distance-induced distortion appear (e.g., does the shape of the nose, size of the mouth or forehead, or do facial features shift by expected degrees relative to one another?). For instance, I have overlayed Tanii's two images to show how one (in blue) exhibits expected distortions while the other (in red) does not, resulting in various misalignments in facial features (assuming the faces are normalized in size):



In such circumstances, a POSITA would understand that two images would still be required, rather than just evaluating one image for distance-induced distortion. Otherwise, an imposter could provide a picture of a user with distance-induced distortion already applied to spoof the system; the need for a more-distance, undistorted image of the user for comparison would still be required.

1021. In my opinion, a POSITA would have been motivated to make either of these two modifications for two reasons. First, a POSITA would have appreciated that Zhang's homography-transformation process may be spoofed by presenting a non-planar picture of a face, because Zhang's homography transformation process specifically looks for a *planar* structure in images that can be transformed nearly identically from one perspective to another. Thus, Zhang's system could possibly be subverted by bending the picture in a way to trick the system, or applying the picture to a three-dimensional shape. For this reason, in my opinion, a POSITA would have

therefore been motivated to look for other methods to ensure the user's face is from the user, and not a spoofer. A POSITA would have also appreciated that distance-induced distortion is more difficult to spoof, because it is induced by the interactions of geometries between the user's face and the camera's lens, and therefore could not be circumvented as easily. Second, a POSITA would have appreciated that either of the processes suggested by Tanii offers a potentially less computationally demanding than the homography mathematics proposed by Zhang, which may be more suitable for a low-power portable device.

1022. In my opinion, a POSITA would have had a reasonable expectation of success in making this modification to Zhang because Tanii already taught a mechanism to identify (and correct) distance-induced distortions, *see, e.g.*, *Tanii*, [0056], and it was already well-known to use depth information about a face derived from a series of images to distinguish between live faces and two-dimensional images of faces. *See, e.g.*, Ex-1014, Abstract, [0031], [0036].

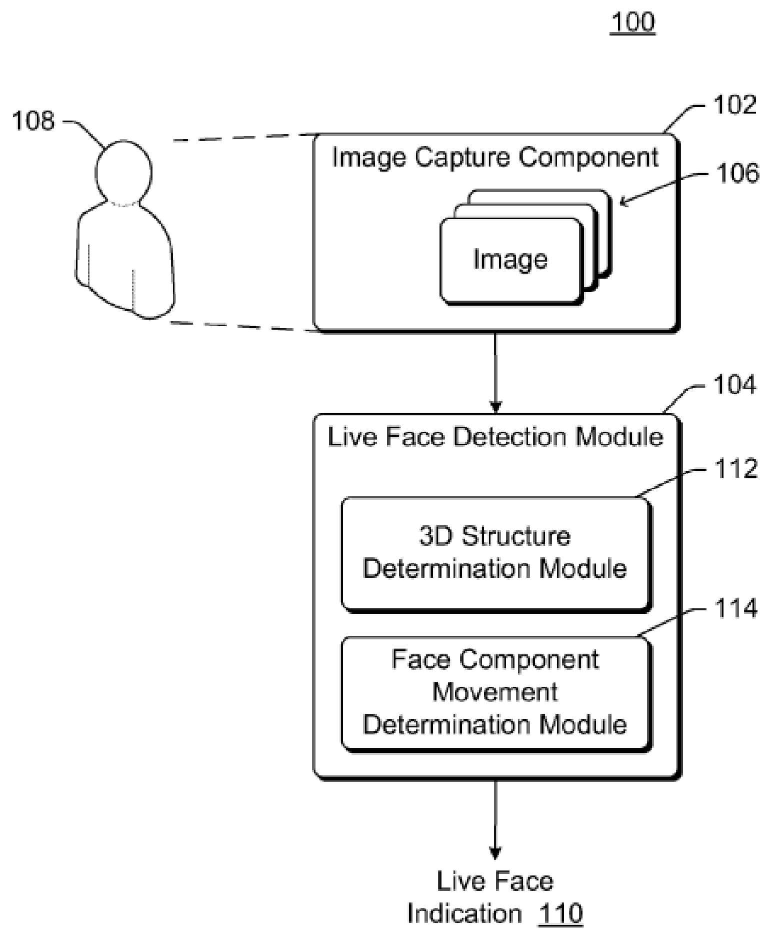
2. Independent Claim 1

a. 1[pre]: A computing device for verifying three-dimensionality of a user via a user's camera equipped computing device, the computing device comprising:

1023. If the preamble is limiting, in my opinion, Zhang discloses or suggests it.

1024. Zhang discloses a system "to determine whether a face in multiple images is a 3D structure or a flat surface" (*Zhang*, [0026], Figs. 1-3; *see also, e.g.*, *id.*, Abstract, [0003], [0013]) to "authenticate a user for particular access" (*id.*, [0012]). To accomplish this, Zhang captures and analyzes multiple images of a user's face using the image capture component 102 implemented in a computing device (e.g., "a desktop computer, a laptop or notebook computer…[or] a cellular or other wireless phone"). *Zhang*, [0012]-[0013], [0016].

1025. In my opinion, a POSITA would have understood that the "image capture component 102" would be a "camera," because cameras are conventionally used to capture images, especially in computing devices. In fact, the "CCDs" and "CMOS" sensors Zhang references are the types of sensors commonly used in cameras. *Zhang*, [0016]; *see also, e.g.*, *Suzuki*, [0019] ("The camera unit includes solid-state image pickup elements such as CCD or CMOS"); Ex-1028, 3 ("Presently, there are two main technologies that can be used for the image sensor ***in a camera***, i.e., CCD (Charge-coupled Device) and CMOS (Complementary Metal-oxide Semiconductor).

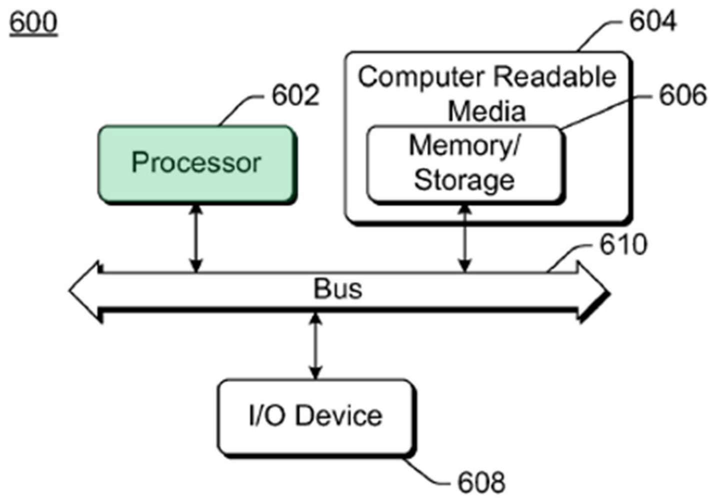


Zhang, Fig. 1.

b. 1[a]: a processor configured to execute machine executable code;

1026. In my opinion, Zhang discloses or suggests limitation 1[a].

1027. Zhang discloses a computing device that contains a processor to execute machine-readable software instructions.



Zhang, Fig. 6, [0063]-[0067] (annotated).

c. 1[b]: a screen configured to provide a user interface to the user;

1028. In my opinion Zhang discloses or suggests limitation 1[b].

1029. Zhang discloses a computing device that contains a display that allows a user to interact with the device and presents information to the user. *Zhang*, [0067].

d. 1[c]: a camera configured to capture images;

1030. In my opinion, Zhang discloses or suggests limitation 1[c].

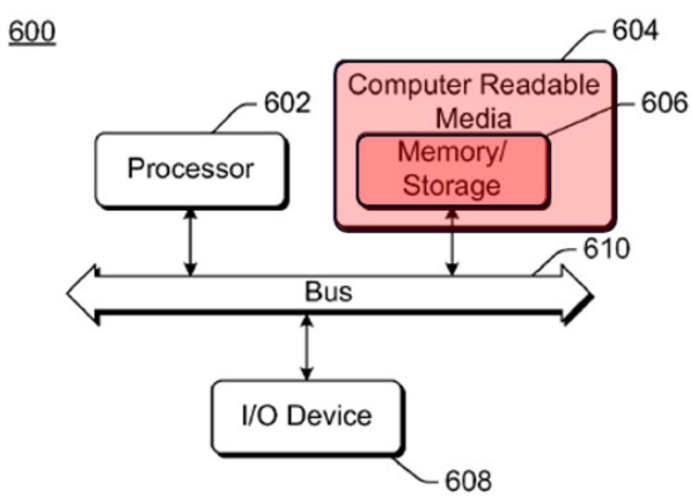
1031. Zhang discloses a computing device that contains an image capture component. *Zhang*, [0012]-[0013]. As I explained previously, a POSITA would have understood that Zhang's "image capture component" is a camera. §XIII.C.2.a (1[pre]).

e. 1[d]: one or more memories configured to store machine readable instructions that are stored on the memory of the authentication server which when executed by the processor, cause the computing device to:

1032. In my opinion, Zhang discloses or suggests limitation 1[d].

1033. As I mentioned previously, in my opinion, it is somewhat unclear which structure is intended to be the "computing device": (1) either a user-facing computing device that engages with a back-end authentication server; or (2) the authentication server itself. *See* §VI.A.3 (claim construction). However, in my opinion, at a minimum, Zhang discloses the former, and the latter would have generally been obvious to a POSITA.

1034. Zhang discloses a computing device that contains computer-readable media (e.g., memory) storing software instructions. *Zhang*, Fig. 6, [0063]-[0067], Fig. 6.



Zhang also discloses that the image capture component and live face detection module (104) can be separate computing devices that communicate and send data, including biometric facial feature data, over a variety of different networks, such as the Internet, a local area network (LAN), an intranet, etc. *Zhang*, [0014].

1035. Although Zhang does not expressly mention that the data is sent to a "server," in my opinion a POSITA would have found it obvious that Zhang's separate computing device would be a server because servers were well-known networking infrastructure, and servers were known to be used for back-end processing of biometric data. *See, e.g.*, *Derakhshani*, 9:27-58, 10:1-24; Ex-1016, Abstract, [0040]-[0043]; Ex-1012, Fig. 1A, 5:24-50. Furthermore, in my opinion, a POSITA would have further understood that, when using a server as the separate computing device running live face detection module (104), the server would store the machine-readable instructions to carry out Zhang's disclosed process and would send instructions to the separate image-capture component that would be stored (even if temporarily) in memory provided in the image-capture component to process and execute those instructions to carry out facial recognition.

1036. Alternatively, a POSITA would have considered it obvious that a server itself may utilize biometric protection. A POSITA would have understood that, just like personal computers, servers were known to store sensitive information—from user profiles for websites, employment or medical records, and more. A POSITA seeking to prevent unauthorized access to reconfigure servers or access their files would have therefore understood that the server itself may be provided with biometric authentication and that in such cases it would include a camera (1[c]) to carry that authentication out.

f. 1[d1]: capturing at least one first image of the user taken with the camera of the computing device at a first location which is a first distance from the user;

1037. In my opinion, Zhang discloses or suggests limitation 1[d1].

1038. In my opinion, Zhang discloses capturing a first image of a user as part of the authentication method. *Zhang*, [0016] ("user 108 presents himself or herself to image capture component 102, allowing component 102 to capture images 106 of user 108."), [0021].

1039. In my opinion, a POSITA would have understood that Zhang's process captures an image at a first distance because some distance exists between the user and image capture component 102 in order to capture a picture of the user's face. *Zhang*, [0016]. Specifically, a POSITA would have understood that there *must* be some distance between the camera and the face to capture the first image so that the camera's field of view encompasses the face. If there were no distance between the camera and the face (e.g., if the camera were pressed up against the user's skin), then the field of view would be limited to just that patch of skin and the face would not be captured in the image, which would be useless for the three-dimensional verification process.

g. 1[d2]: processing the at least one first image or a portion thereof to create first data;

1040. In my opinion, Zhang discloses or suggests limitation 1[d2].

1041. Zhang discloses processing the first image to extract "feature points" from the image. *Zhang*, [0027] ("[O]ne or more feature points are extracted from two images…A variety of different feature points can be extracted, such as a corner of an eye, a corner of a mouth, a tip of a nose, and so forth."), [0026] (disclosing "software, firmware, hardwire, or combin[ed]" implementations).

1042. In my opinion, a POSITA would have understood that Zhang's extracted feature points constitute "biometric data" because "biometric data" generally refers to unique physical characteristics of an individual, which would include the positions of "feature points" such as a user's eyes, nose, mouth, and other such features. *See, e.g.*, Ex-1018, 2 ("biometric recognition can be defined as the science of establishing the identity of an individual based on the physical and/or behavioral characteristics of the person."), 100-103 (describing the types of biometric data about a face used for facial-recognition systems, "such as the structure of the face components (e.g., eyes), [and] the relationship between facial components").

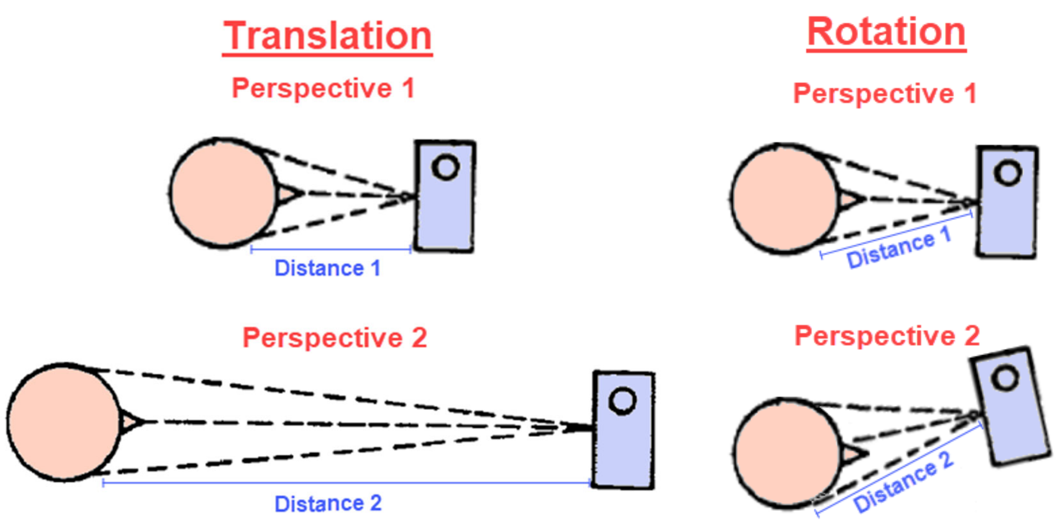
h. 1[d3]: capturing at least one second image of the user taken with the camera of the computing device is at a second distance from the user, the second distance being different than the first distance, the capturing at least one second image of the user occurring after movement of the camera or the user to establish the camera at the second distance from the user;

1043. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 1[d3].

1044. To start, there appears to be a typographical error in this claim. Specifically, the claim recites "capturing at least one second image of the user ***taken*** with the camera…***is*** at a second distance," which does not make sense grammatically. I have therefore read the claim to read "capturing at least one second image of the user with the camera…at a second distance," without the words "taken," and "is" to make it grammatically legible. In my opinion, these two words do not impact the claim as it relates to the application of the prior art.

1045. Zhang discloses capturing a second image of a user as part of the authentication method. *Zhang*, [0016] ("Image capture component 102 captures multiple images").

1046. Zhang does not expressly disclose that the second image is captured at a second distance different from the first distance of the first image. But, in my opinion, a POSITA would have understood that Zhang at least implicitly requires *some* change of distance. §VII.C (Zhang). Specifically, Zhang discloses a "3D structure determination module 112" that uses a "homography" technique to distinguish between a real face and a picture of a face by, *inter alia*, transforming a first image to the perspective of a second image and comparing the two. *Zhang*, [0024], [0026]-[0035]; §VII.C (Zhang). In my opinion, a POSITA would have understood from Zhang that—like Derakhshani's parallax approach—the distances between the camera and at least some facial landmarks would change in order to obtain an image from a different perspective than the first, and would obviously also encompass changing the overall distance between the camera and face as well. *See, e.g.*, §XIII.A.2.h (in the context of Derakhshani, discussing changes of distance for parallax).



Moreover, a POSITA would have not only understood that providing images at different distances allows for a greater understanding of depth between objects in the scene, as exemplified in the paper Zhang references; Ex-1013, 22-25, but that taking pictures at different distances may induce distance-based distortion that would enhance the accuracy of Zhang's homography transformation to detect a three-dimensional face. §XIII.C.1 (motivation).

1047. Even if Zhang cannot be considered to disclose or suggest taking two images at different distances, however, a POSITA would have been motivated to do so in view of other prior art. For instance, a POSITA would have understood that distortions caused by camera lenses can indicate depth in the object being captured, as exemplified by Tanii. §XIII.C.1 (motivation). Thus, even if Zhang does not already disclose this limitation, a POSITA would have been motivated to modify Zhang in view of Tanii to capture a second image at a second distance and evaluating the images for different degrees of distance-induced distortions to distinguish between live, three-dimensional faces and two-dimensional pictures of a face. §XIII.C.1 (motivation).

1048. Accordingly, because Zhang, whether alone or combined with Tanii, teaches capturing a plurality of images at different distances between the user and the camera, in my opinion, a POSITA would have understood that either the camera or user must move relative to the other in between image captures; there are no other ways to change the distance between the two.

i. 1[d4]: processing the at least one second image or a portion thereof to create second data;

1049. In my opinion, Zhang discloses or suggests limitation 1[d4].

1050. Zhang discloses processing the second image to obtain second feature-point biometric data from the image. *Zhang*, [0026]-[0027]; §XIII.C.2.g (1[d2]).

j. 1[d5]: comparing the first data to the second data to determine whether expected differences exist between the first data and the second data which indicates three-dimensionality of the user;

1051. In my opinion, Zhang discloses or suggests limitation 1[d5].

1052. Zhang discloses that "[t]he feature points extracted…are matched across the first and second images (act 304)," and those feature points constitute data. *Zhang*, [0028]; §XIII.C.2.i (1[d4]). Furthermore, Zhang discloses that the matching process may also "determine[] whether the first and second images include the same face," including "during the matching of feature points in 304, if all (or at least a threshold number) of the feature points cannot be matched then it is determined that the first and second images are of different faces." *Zhang*, [0038].

1053. Zhang discloses that, after the homography matrix between the first and second image is determined, a "warped" version of the first image is created and then compared to the second image to determine whether differences exist. *Zhang*, [0025], [0031]. Zhang also discloses that, as part of the comparison, "any of a variety of conventional face detection algorithms or face recognition algorithms can be used to detect the face within each image, and the selected locations are the locations that are part of a face within at least one of the warped and second images." *Zhang*, [0032].

1054. In my opinion, a POSITA would have therefore understood that Zhang discloses comparing a first biometric data (e.g., the facial-feature locations in the first "warped" (transformed) image) and second biometric data (e.g., the facial-feature locations in the second image) to determine whether differences between the two exist, in which it would be expected that a live face would have sufficient differences between the two images due to movement of the image capture component 102 (camera).

1055. However, a POSITA would have also been aware that differences between two images—one with lens-induced distortions and one without—can also be used to distinguish between live, three-dimensional faces, and two-dimensional pictures of a face, as exemplified by Tanii. §XIII.C.1 (motivation). And, in my opinion, a POSITA would have been motivated to look for these expected distortions as either a supplemental or alternative verification of three-dimensionality of a face. *Id*. A POSITA would have appreciated that verifying the three-dimensional nature of the face using distance-induced distortion would be accomplished by matching the positions of biometric facial features across the first and second images, consistent with Zhang. But rather than using that comparison to calculate a homography matrix, the comparison would evaluate whether one of the images exhibits the distance-induced distortion that would be expected when the user's face is captured at a close distance to the camera, and the other image does not exhibit similar lens-induced distortion when captured further from the camera. *Id*. In my opinion, a POSITA would have appreciated that, when modifying Zhang to evaluate differences caused by distance-induced distortions, a three-dimensional face would be indicated when one of the two sets of data exhibits expected distance-induced distortions due to the change in distance of the camera. *Zhang*, [0025], [0034].

k. 1[d6]: verifying the images of the user exhibit three dimensional traits when the expected differences exist between the first data and the second data as a result of capturing the at least one first image and the at least one second image at different distances from the user.

1056. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 1[d6].

1057. Zhang discloses that captured images are determined to be of a live, three-dimensional face when differences in the image data exist after undergoing a homography transformation, *Zhang*, [0031], including when first biometric data (the position of facial features in the first "warped" (transformed) image) does not match the second biometric data, *Zhang*, [0032]-[0034]. These differences would be expected due to a change in perspective (rotation and/or distance) of the camera between the two images. *See* §§XIII.C.2.h (1[d3]), XIII.C.2.j (1[d5]). By matching data across the first and second images, Zhang also discloses determining whether the first and second images include the same face for the purpose of authenticating the user. *Zhang*, Fig. 2, [0017], [0038]. Conversely, if "all (or at least a threshold number) of the feature points cannot be matched then it is determined that he first and second images are of different faces," and the user is not authenticated. *Zhang*, Fig. 2, [0017], [0038].

1058. Moreover, to the extent Zhang does not disclose this limitation, Zhang combined with Tanii does. *See* §§XIII.C.2.h (1[d3]), XIII.C.2.j (1[d5]). Specifically, a POSITA would have appreciated that, when modifying Zhang to evaluate differences arising from distance-induced distortions in Tanii, a three-dimensional face would be indicated when one of the two sets of biometric data exhibits expected distance-induced distortions due to the change in distance of the camera, allowing Zhang's authentication process to proceed. *See* §§XIII.C.2.h (1[d3]), XIII.C.2.j (1[d5]).

3. Claim 2

a. 2[a]: The system according to claim 1, further comprising: interpolating the first data and the second data to obtain estimated intermediate data;

1059. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 2[a].

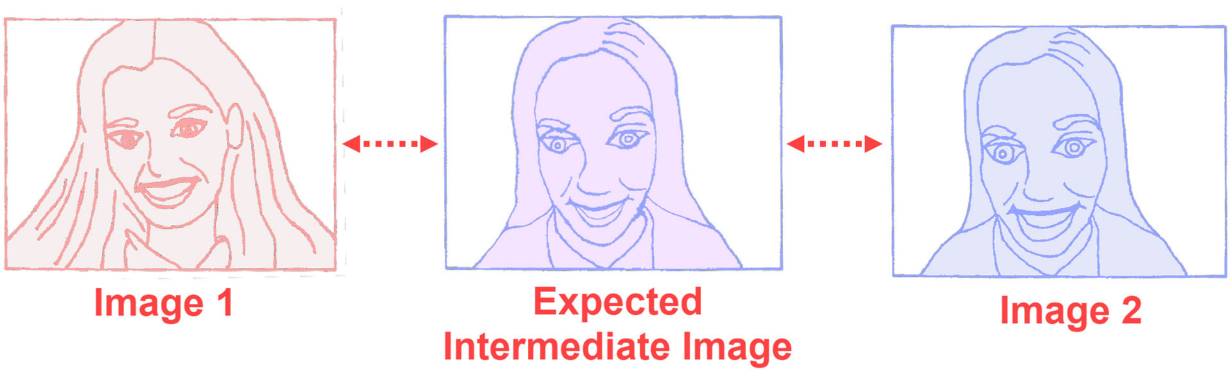
1060. Zhang discloses that, as part of the authentication process, the two images being compared may be "non-adjacent." Zhang, [0036]. Zhang explains that images are "non-adjacent" when additional images exist between the two images being compared for authentication. *Id*. In such instances, Zhang discloses performing some of the processes, such as "feature point extraction and feature point matching" using the intermediate images to "facilitate the feature matching process when matching features across two images with one or more intervening images." *Id*. Zhang also discloses that the homography-transformation process can be applied to multiple pairs of images, whether the images are adjacent or non-adjacent. Zhang, [0037].

1061. When a set of intermediate images exist *between* the first and second images, as Zhang discloses, in my opinion, a POSITA would have been motivated to generate predictions (i.e., interpolations) of what those intermediate images should look like based on Zhang's first and second images because using static images to build models or predictions of the face as a means of identifying a user was well-known in the art. Ex-1015, Abstract; *Derakhshani*, 17:24-44 (interpolating two-dimensional and three-dimensional models for comparison to acquired biometric data); Ex-1036, 8:19-27 (describing capturing one or more biometric features and calculating "change parameters" to evaluate whether the changes match expectations, or predictions of what the biometric features should look like). And a POSITA would have understood that building models or predictions of what Zhang's intermediate images *should* look like would further ensure against spoofing because a spoofer could not rely on artificial differences between the first and second images to have Zhang's system authenticate a face; the differences would also have to match what is expected *between* the two images.

1062. Based on a POSITA's understanding of Zhang, a POSITA would have further been motivated to derive interpolated biometric data based on the combination of Zhang and Tanii. Specifically, Tanii discloses that distance-induced distortions increase as distance between the face and camera decreases. *See Tanii*, [0048]. A POSITA would have therefore understood that, all else being equal, distance-induced distortion depends on the distance between the user and the camera, and thus any set of images as the distance between the user and camera changes will have different degrees of facial warping, similar to the gradual changes in facial distortion that appears in the series of images below (although these images also altered the focal length of the camera to ensure the face remains a constant size in the frame, rather than just distance:

Ex-1022.

1063. In my opinion, a POSITA reading Zhang—which discloses processing, interpolating, and evaluating intermediate images—in view of Tanii therefore would have been motivated to interpolate intermediate biometric data with an intermediate, interpolated amount of distance-induced distortion based on the two non-adjacent images to create an array of intermediate distance-induced distortions that would indicate depth in a three-dimensional face, such a gradual projection from the lines of the facial landmarks shown in blue (the distorted image) to the lines of the facial landmarks shown in red (the distorted image).



Performing this type of interpolation between the distorted and undistorted images would allow for further comparison with additional images, for instance, to ensure the distance-induced distortion matches expectations if the face were truly three-dimensional, as depicted below:

b. 2[b]: capturing at least one third image of the user taken with the camera of the computing device at a third distance from the user, the third distance being between the first distance and the second distances;

1064. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 2[b].

1065. Zhang discloses capturing a series of intermediate images between two non-adjacent images. *Zhang*, [0035]-[0037]; *see* §XIII.C.3.a (2[a]). A POSITA would have understood that these intermediate images would provide images at different positions (e.g., rotation or translation) of the camera between the first and second images. *See* §XIII.C.3.a (2[a]).

1066. Moreover, when modifying Zhang in view of Tanii to interpolate intermediate biometric data attributable to distance-induced distortions, a POSITA would have been further motivated to capture a third image at a distance that correlates to one of the interpolated data sets for further authentication of three-dimensional depth of the face in the captured images. *See* §XIII.C.3.a (2[a]).

c. 2[c]: processing the at least one third image or a portion thereof to obtain third data; and

1067. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 2[c].

1068. Zhang discloses processing a third (intermediate) image to obtain third data. *Zhang*, [0036] ("the feature point extraction and feature point matching in acts 302 and 304 can be generated for each adjacent pair of images in the sequence, which can facilitate the feature matching process when matching features across two images with one or more intervening images.").

1069. Moreover, when modifying Zhang in view of Tanii, a POSITA would have found it obvious to acquire a third image and extract biometric data from the third image to compare it to the interpolated positions of the biometric data based on the first and second images. *See* §§XIII.C.3.a (2[a]), XIII.C.3.b (2[b]).

d. 2[d]: comparing the estimated intermediate data with the third data to determine whether the third data matches the estimated intermediate data.

1070. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 2[d].

1071. Zhang discloses tracking and comparing biometric features between the non-adjacent and intermediate images. *See Zhang*, [0036]-[0037]; §§XIII.C.3.a (2[a]).

1072. In my opinion, a POSITA would have understood that, when interpolating what the intermediate images *should* look like based on the first and second images, a POSITA would have understood that the estimated, interpolated biometric data would be compared to the intermediate images to determine whether the intermediate images match what was predicted. *See* §XIII.C.3.b (2[b]).

1073. Moreover, when modifying Zhang in view of Tanii, a POSITA would have found it obvious to acquire a third image and extract data from the third image to compare it to the interpolated positions of the data based on the first and second images to determine if there is a match between the two. *See* §XIII.C.2.j (1[d5]).

4. Claim 3: The system according to claim 1, further comprising verifying the presence of one or more features on a side of a user's head in the at least one first image, and verifying the absence or reduced visibility of the one or more features on the side of the user's head in the at least one second image due to image capture at different distances from the user's head, wherein the first distance is larger than the second distance.

1074. In my opinion, Zhang combined with Tanii teaches claim 3's additional limitations.

1075. Zhang does not expressly disclose a process of verifying the presence in one image and absence in another of a feature on the side of a user's head.

1076. However, when modifying Zhang in view of Tanii to use distance-induced distortions to verify the three-dimensional nature of a face, Tanii teaches that the absence and presence of an ear is a natural result of the distance between the user and camera. In my opinion, a POSITA would have appreciated that an ear is a "feature[] on a side of a user's head" as claimed. When a sufficient distance between the face and camera exists, the ears are captured because there is enough distance for the light rays from the ears to strike the camera's lens. *See Tanii*, [0048], Figs. 3A-3B.

Fig.3A

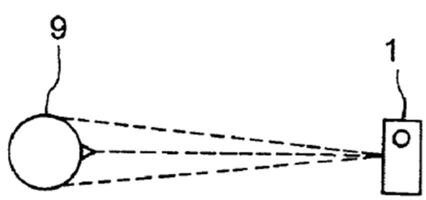
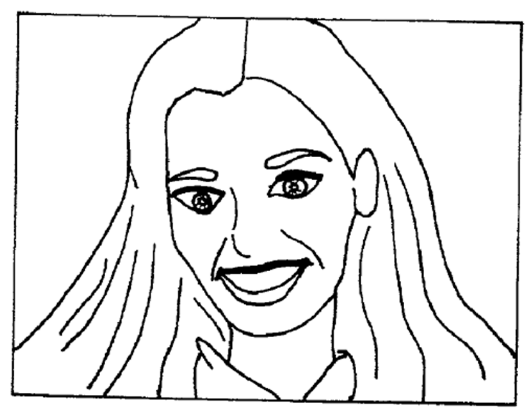


Fig.3B



But when a face is too close, the ears will not be captured because there is insufficient distance for the light rays from the ears to strike the camera's lens. *See Tanii*, [0047]-[0048], Figs. 4A-4B.

Fig.4A

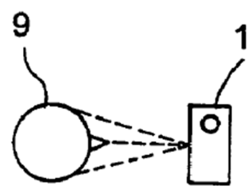
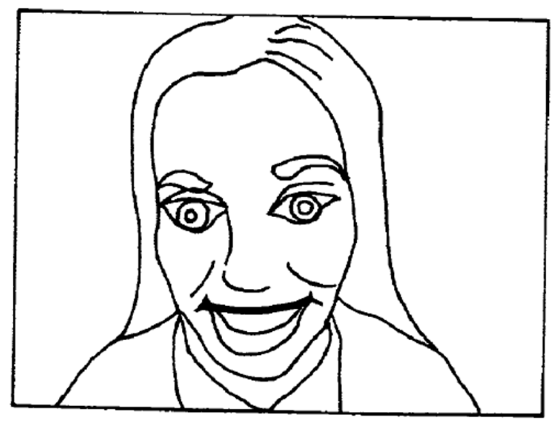
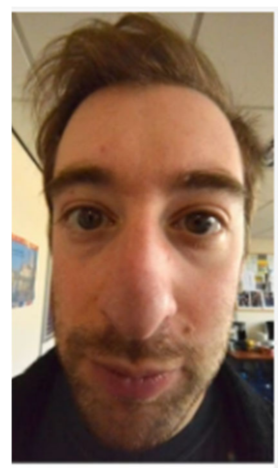


Fig.4B



This effect was well-known and demonstrated in actual applications, as shown below.

85mm @ 200cm



8mm @ 20cm

In my opinion, therefore, a POSITA would have appreciated based on at least Tanii that the presence of a user's ears in one image at a sufficient distance, but absence of a user's ears in another image at a closer distance would be indicative of a three-dimensional face, and would have been motivated to modify Zhang to verify the presence and absence of the ears between images as yet another indicator of the three-dimensional nature of the face.

5. Claim 5

a. 5[a]: The system according to claim 1, further comprising comparing the first data, second data, or both to enrollment data derived from an enrollment image, the enrollment image captured and stored prior to an authenticating; and

1077. In my opinion, Zhang discloses or suggests 5[a]'s additional limitation.

1078. Zhang discloses comparing at least portions of the first image, second image, or both to enrollment data captured and stored prior to the authentication session. *Zhang*, [0017] ("The authentication of user 108 can be performed…by comparing one or more of images 106 to previously captured images of user 108.").

1079. In my opinion, a POSITA would have understood that Zhang's "previously captured images" would be taken during an enrollment session, as is conventional for biometric-authentication systems. *See* Ex-1018, 4-11 (providing overview of biometric authentication and verification).

b. 5[b]: only authenticating the user when the first data, the second data, or both match the enrollment data within a predetermined threshold.

1080. In my opinion, Zhang discloses or suggests 5[b]'s additional limitation.

1081. Although Zhang does not provide significant details about the overall authentication process—but instead states "a variety of different manners" can be used—Zhang's description of comparing biometric features to "previously captured images” is consistent with a conventional biometric-authentication procedure that requires a sufficient “match” above a threshold. *See Zhang*, [0017] (“The authentication of a user 108 can be performed…by comparing one or more of images 106 to previously captured images of user 108.”), [0038] (disclosing inter-picture matching); *see also, e.g.*, Ex-1018, 17 (“[B]iometric systems mostly decide on a person’s identity based on a *close* match between the template and the query, where the strength of the match (or the degree of similarity) is represented by the match score.”), 18 (“a verification system makes a decision by comparing the match score s to a threshold η”). In fact, thresholds were used because facial authentication must account for many different conditions that cause differences between image captures of a face, such as the perspective of the face in the image, lighting, facial accessories, facial hair, and more. *See, e.g.*, *id.*, 99, Fig. 3.1 (noting “[t]he problem of intra-class (i.e., intra-user) variation is quite pronounced in the context of face recognition. The face image of an individual can exhibit a wide variety of changes that make automated face recognition a challenging task” such as differences in “pose, illumination, and expression…aging,” and facial accessories). In other words, facial authentication looks for matches to a prescribed certainty (e.g., the threshold), rather than an exact match.

1082. For these reasons, in my opinion, a POSITA would have understood Zhang as disclosing a conventional facial-authentication procedure in which the first or second data (or both) must match "previously captured" enrollment data within a predetermined threshold to authenticate the identity of the user. In fact, authenticating a user's *identity* is a central aspect of facial authentication systems, and not just evaluating whether the face is three-dimensional or not. *Zhang*, [0001] (noting the purpose of the invention is to prevent unauthorized users from accessing secure resources); *see* Ex-1018, 259 (noting "[l]iveness detection"—like Zhang—is just one aspect of biometric authentication systems to mitigate spoofers).

6. Claim 6: The system according to claim 1, wherein the computing device is a hand-held device, and the user holds the device at the first and second distance to capture the at least one first image and the at least one second image.

1083. In my opinion, Zhang combined with Tanii teaches claim 6's additional limitations.

1084. Zhang discloses that the face authentication process can be implemented on a variety of different type of hand-held computing devices, such as a cellular or other wireless phone, a digital camera or video camera. *Zhang*, [0013]. Moreover, Tanii notes that distance-induced distortions often occur in mobile devices that have incorporated wide-angle lenses, and the amount of distortion is dictated by the distance between the user and the camera. *Tanii*, [0007], [0047]-[0048], Figs. 3A-B, 4A-B.

1085. In my opinion, a POSITA would have also understood that, when performing a three-dimensional verification of the face on a mobile computing device, *see* §XIII.C.2.h (1[d3]), the user would adjust the distance by holding the mobile device and extending and retracting their arm (holding the mobile device at a first distance, then a second distance). A POSITA would have appreciated that mobile devices are routinely held to capture images, and holding the mobile device and adjusting distance would be a convenient and obvious way of changing the distance.

7. Claim 7: The system according to claim 1, wherein the first data and the second data comprise biometric data.

1086. In my opinion, Zhang discloses or suggests claim 7's additional limitation for the reasons discussed in §§XIII.C.2.g (1[d2]), XIII.C.2.i (1[d4]).

1087. In my opinion, a POSITA would have also understood that data created from an image (or portion of an image) of a user's face would comprise biometric data. *Id.*

8. Claim 8: The system according to claim 1, wherein the first data and the second data comprise a mapping of facial features.

1088. In my opinion, Zhang discloses or suggests claim 8's additional limitation.

1089. Zhang discloses processing the multiple images to extract "feature points" from the image that correspond to characteristics of a user's face. *Zhang*, [0027], ("[O]ne or more feature points are extracted from two images… A variety of different feature points can be extracted, such as a corner of an eye, a corner of a mouth, a tip of a nose, and so forth.”). In my opinion, a POSITA would have understood that the identification of facial landmarks would include their locations relative to one another, thus constituting a mapping of facial features. In fact, processing image data to map facial features was a conventional aspect of facial-recognition systems. *See, e.g.*, Ex-1018, 103 (Fig. 3.5(b) describing how “Level 2 features require detailed processing for face recognition. Information regarding the structure and the specific shape and texture of local regions in a face is used to make an accurate determination of the subject’s identity.”). Moreover, a POSITA would have understood that the process of converting facial features in an image to computer-readable data conventionally involves mapping those features to data. *See, e.g.*, Ex-1018, 116-17 (noting how “appearance-based techniques generate a compact representation of the entire face region in the acquired image by *mapping* the high-dimensional face image into lower dimensional sub-space.”).

9. Claim 9: The method according to claim 1, wherein the first image and the second image is of the user's face and the user's head and facial features are held steady and without movement during capture of the first image and the second image.

1090. In my opinion, Zhang, alone or in combination with Tanii, teaches claim 9’s additional limitation.

1091. Zhang and Tanii both teach or suggest moving the camera to capture images at two different distances. *See* §XIII.C.2.h (1[d3]).

1092. In my opinion, a POSITA would have understood that, when moving the camera to capture images from different distances, the user's face would be stationary (e.g., steady) both during each image capture (to ensure each image is not blurry due to the camera's exposure time) and during movement of the camera (to isolate any differences between images to those attributable to the change of distance). Moreover, a POSITA would have appreciated that holding the user's face steady and moving the camera closer and further away would be more user friendly than requiring the user to move their head closer and further from the camera while holding the camera steady.

10. Independent Claim 10

a. 10[pre]: A method for evaluating three-dimensionality of a user, the method comprising:

1093. If the preamble is limiting, in my opinion, Zhang discloses or suggests it.

1094. Specifically, Zhang discloses a method "to determine whether a face in multiple images is a 3D structure or a flat surface," *Zhang*, [0026], Figs 2-3; *see also, e.g.*, *id.*, Abstract, [0003].

b. 10[a]: capturing at least one first image of the user taken with a camera at a first location which is a first distance from the user;

1095. In my opinion, Zhang discloses or suggests limitation 10[a] for the reasons discussed in §XIII.C.2.f (1[d1]).

c. 10[b]: processing the at least one first image or a portion thereof to create first data;

1096. In my opinion, Zhang discloses or suggests limitation 10[b] for the reasons discussed in §XIII.C.2.g (1[d2]).

d. 10[c]: moving the camera from the first location to a second location, the second location being a second distance from the user, or the user moving to change the distance between the user and the camera from the first distance to the second distance;

1097. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 10[c] for the reasons discussed in §XIII.C.2.h (1[d3]).

e. 10[d]: capturing at least one second image of the user taken with the camera when the camera is the second distance from the user, the second distance being different than the first distance;

1098. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 10[d] for the reasons discussed in §XIII.C.2.h (1[d3]).

f. 10[e]: processing the at least one second image or a portion thereof to create second data;

1099. In my opinion, Zhang discloses or suggests limitation 10[e] for the reasons discussed in §XIII.C.2.i (1[d4]).

g. 10[f]: comparing the first data to the second data to determine whether expected differences exist between the first data and the second data which indicate three-dimensionality of the user;

1100. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 10[f] for the reasons discussed in §XIII.C.2.j (1[d5]).

h. 10[g]: verifying the images of the user exhibit three-dimensional traits when the first data and the second data have expected differences resulting from the at least one first image being captured with the camera at a different distance from the user than when the at least one second image is captured.

1101. In my opinion, Zhang, alone or in combination with Tanii, teaches limitation 10[g] for the reasons discussed in §XIII.C.2.k (1[d6]).

11. Claim 11

a. 11[a]: The method according to claim 10, further comprising: interpolating the first data and the second data to obtain estimated intermediate data;

1102. In my opinion, Zhang, alone or combined with Tanii, teaches limitation 11[a] for the reasons discussed in §XIII.C.3.a (2[a]).

b. 11[b]: capturing at least one third image of the user taken with the camera at a third distance from the user, the third distance being between the first distance and the second distances;

1103. In my opinion, Zhang, alone or combined with Tanii, teaches limitation 11[b] for the reasons discussed in §XIII.C.3.b (2[b]).

c. 11[c]: processing the at least one third image or a portion thereof to obtain third data; and

1104. In my opinion, Zhang, alone or combined with Tanii, teaches limitation 11[c] for the reasons discussed in §XIII.C.3.c (2[c]).

d. 11[d]: comparing the estimated intermediate data with the third data to determine whether the third data matches the estimated intermediate data.

1105. In my opinion, Zhang, alone or combined with Tanii, teaches limitation 11[d] for the reasons discussed in §XIII.C.3.d (2[d]).

12. Claim 12: The method according to claim 10, further comprising verifying the presence of ears of the user in the at least one first image, and verifying the absence or reduced visibility of the ears in the at least one second image, wherein the first distance is larger than the second distance.

1106. In my opinion, Zhang, alone or in combination with Tanii, teaches claim 12's additional limitations for the reasons discussed in §XIII.C.4 (cl.3).

13. Claim 15: The method according to claim 10, wherein the camera is part of a computing device is a hand-held device, and the user holds the computing device at the first distance when capturing at least one first image and at the second distances when capturing the at least one second image.

1107. In my opinion, Zhang, alone or in combination with Tanii, teaches claim 15's additional limitations for the reasons discussed in §XIII.C.6 (cl.6).

14. Claim 16: The method according to claim 10, wherein the first data and the second data comprise biometric data.

1108. In my opinion, Zhang, alone or in combination with Tanii, teaches claim 16's additional limitation for the reasons discussed in §XIII.C.7 (cl.7).

15. Claim 17: The method according to claim 10, wherein the first data and the second data comprise a map of facial features.

1109. In my opinion, Zhang, alone or in combination with Tanii, teaches claim 17's additional limitation for the reasons discussed in §XIII.C.8 (cl.8).

16. Claim 19: The method according to claim 10, wherein a face of the user is held steady when capturing the at least one first image and the at least one second image and the camera moves from the first location to the second location.

1110. In my opinion, Zhang, alone or in combination with Tanii, teaches claim 19's additional limitation for the reasons discussed in §XIII.C.9 (cl.9).

17. Claim 20: The method according to claim 10, wherein the first data and the second data are maintained on a computing device.

1111. In my opinion, Zhang discloses or suggests claim 20's additional limitation.

1112. Zhang discloses an image capture component (102) and a live face detection module (104) that can both be implemented on the same computing device. *Zhang*, [0014]. In such instances, in my opinion, a POSITA would have understood that the biometric data would be maintained on the computing device so that all processes would be performed using a single device. If a single, local device is being used for all authentication procedures, there would be no need to store any data on a separate device. Moreover, a POSITA would have understood that biometric data is extremely sensitive, and transmitting that data over networks presents a security risk of being intercepted. A POSITA would have understood that there would be no need to take any security risks by transmitting such data if the single, local device is handling the entire biometric-authentication procedure.

18. Claim 21: The method of claim 10 wherein the camera is part of is one of a smartphone, tablet, laptop, or desktop computer.

1113. In my opinion, Zhang, alone or in combination with Tanii, teaches claim 21's additional limitation.

1114. Zhang discloses an image capture component (102) and a live face detection module (104) that can both be implemented on the same computing device. *Zhang*, Fig. 6, [0014].

19. Independent Claim 22

a. 22[pre]: A method, performed by a user using a user's computer device, for verifying three-dimensionality of the user, the method comprising:

1115. If the preamble is limiting, in my opinion, Zhang discloses or suggests it for the reasons discussed in §XIII.C.2.a (1[pre]).

b. 22[a]: capturing a first image of the user's head with a camera at a first distance from the user, the camera associated with the user's computing device;

1116. In my opinion, Zhang discloses or suggests limitation 22[a] for the reasons discussed in §XIII.C.2.f (1[d1]).

c. 22[b]: changing a distance between the user and the camera to a second distance by the user moving the camera, or the user moving relative to the camera, or both;

1117. In my opinion, Zhang discloses or suggests limitation 22[b] for the reasons discussed in §XIII.C.2.h (1[d3]).

d. 22[c]: capturing a second image of the user's head with the camera when the camera is at the second distance from the user, the second distance being different than the first distance;

1118. In my opinion, Zhang discloses or suggests limitation 22[c] for the reasons discussed in §XIII.C.2.h (1[d3]).

e. 22[d]: comparing one or more aspects of the user's head from the first image to one or more aspects of the user's head from the second image to determine whether expected differences, between the first image and the second image, exist which indicates three-dimensionality of the user, such that the expected differences between the first image and the second image result from the first image being captured when the camera is at a different distance from the user than when the second image is captured; and

1119. In my opinion, Zhang discloses or suggests limitation 22[d] for the reasons discussed in §§XIII.C.2.j (1[d5]), XIII.C.2.k (1[d6]).

f. 22[e]: responsive to the comparing determining that expected differences between the first image and the second image exist, providing notice to the user, a third party, or both that the three-dimensionality of the user is verified.

1120. In my opinion, Zhang discloses or suggests limitation 22[e].

1121. After Zhang's live face detection module analyzes two images, the module "outputs an indication 110 of whether images 106 include a live face or a picture of a face." Zhang, [0017]. The indication may be, e.g., "yes" and "authenticate" or "no" and "do not authenticate." *Id*.

20. Claim 23: The method of claim 22 wherein the one or more aspects of the user's head from the first image is first data resulting from processing the first image and the one or more aspects of the user's head from the second image is second data resulting from processing the second image.

1122. In my opinion, Zhang, alone or in combination with Tanii, teaches claim 23's additional limitation for the reasons discussed in §§XIII.C.2.g (1[d2]), XIII.C.2.i (1[d4]).

21. Claim 24: The method of claim 22 wherein the user's head is the user's face.

1123. In my opinion, Zhang, alone or in combination with Tanii, teaches claim 24's additional limitation for the reasons discussed in §XIII.C.9 (cl.9).

D. Ground 2B: Zhang, Tanii, and Tahk (Claims 4, 13, 14)

1. Motivation to Combine

1124. In my opinion, a POSITA would have been motivated to modify Zhang, with or without Tanii, in view of Tahk because Tahk provides a user-friendly way of ensuring that a face presented for facial authentication is properly framed. In my opinion, a POSITA would have understood that providing a real-time preview of what an image would look like prior to capturing the image, as well as providing express prompts—such as written instructions or oval shapes on the live-preview screen—as taught by Tahk, §VII.D (Tahk), would have been particularly useful for authentication procedures that require capturing multiple images of a face, as taught by Zhang and Tanii.

1125. For instance, providing a user a real-time preview would allow the user to actively adjust the position of the camera and/or their orientation of their face to properly frame their face for the image capture. Moreover, providing express prompts and oval shapes sized to guide the user to properly position their face would ensure images best suited for facial recognition can be captured, and that the faces would be captured from sufficiently different perspectives and/or distances to ensure the three-dimensional verification taught by Zhang, alone or in combination with Tanii, could be performed. This is particularly important for facial-recognition systems, which generally are known to have difficulty matching faces across different views. *See, e.g.*, Ex-1018, 32.

1126. Moreover, a POSITA would have known that providing real-time image feedback, written instructions, and oval shapes to frame a face during a facial authentication process were all well-known and conventional techniques to provide user feedback during image capture as of the time of the invention. *See, e.g.*, Ex-1034, 7:16-8:7, Figs. 6B-7C (providing "an example of an interface used upon registering a facial image of a person to be authenticated" in which "the image of this person is displayed on a monitor" during registration, and oval-shaped prompts to indicate the size/distance of the user from the face authentication sensor); Ex-1035, 5:31-32 ("The computing device may present prompts that instruct the user to perform one or more liveness gestures"), 6:3-4 (same).

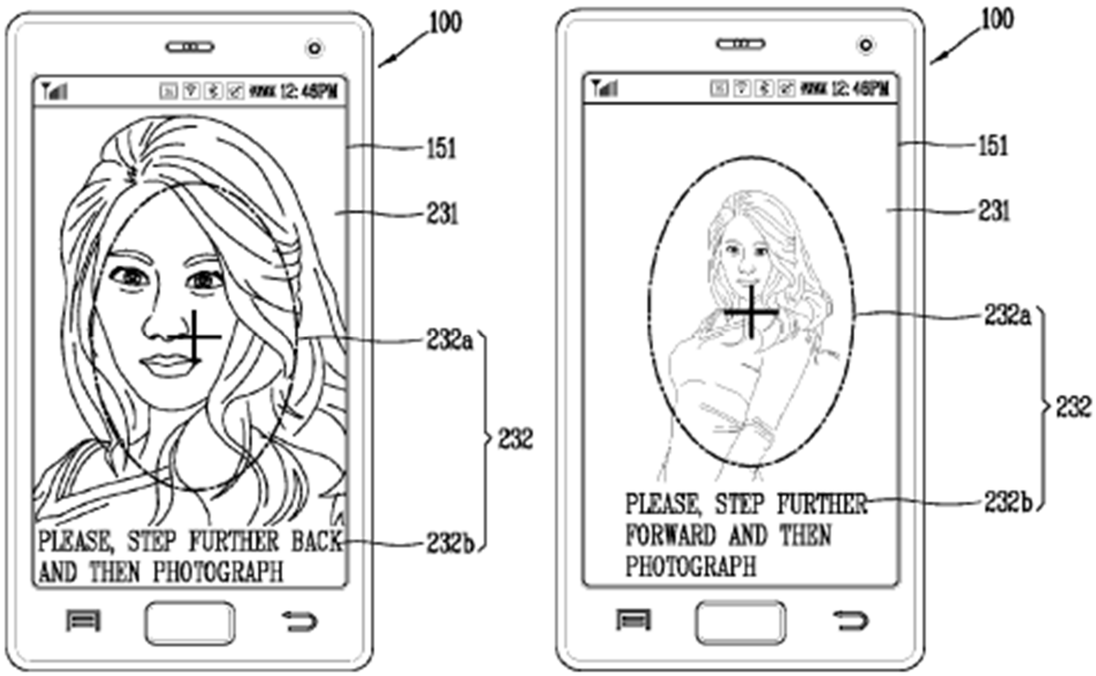
2. Claim 4: The system according to claim 1, wherein the machine readable instructions is configured to display one or more prompts on the screen of the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

1127. In my opinion, Zhang combined with Tanii and Tahk teaches claim 4's additional limitations.

1128. Zhang discloses taking a series of images sufficient to calculate a homography matrix. *See, e.g.*, *Zhang*, [0026], Figs. 1, 3. A POSITA would have understood that Zhang already discloses, or that Zhang combined with Tanii teach, taking a series of images at different distances between the face and the camera. *See* §§XIII.D.1 (motivation), XIII.C.2.h (1[d3]). However, Zhang and Tanii do not expressly teach providing a series of prompts to a user to guide them through different camera positions that would enhance calculations of the homography matrix.

1129. Tahk, however, teaches that using one or more prompts on a screen ensures images of the face are captured at the correct distances. *See, e.g.*, Tahk, Figs.

8A-B ("Please step further back" and "Please step further forward," and presenting an oval to frame the face at the correct distance).



In my opinion, a POSITA would have been motivated by Tahk to modify Zhang, whether alone or in combination with Tanii, to expressly prompt a user to alter the distance of the camera in order to either capture sufficiently different images to perform a homography transformation (Zhang) or to capture an image with distance-induced distortions (Tanii) to ensure the images could be used to distinguish live from two-dimensional images of faces. *See also* §XIII.D.1 (motivation).

3. Claim 13: The method according to claim 10, further comprising one or more prompts on a screen to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

1130. In my opinion, the combination of Zhang, Tanii, and Tahk teaches claim 13's additional limitations for the reasons discussed in §XIII.D.2 (cl.4).

4. Claim 14: The method according to claim 13, wherein the one or more prompts are an oval shape guide on the screen within which an image of a face of the user is aligned to capture the at least one first image and the at least one second image.

1131. In my opinion, the combination of Zhang, Tanii, and Tahk teaches claim 14's additional limitations.

1132. Neither Zhang nor Tanii expressly teach using prompts to guide a user during the facial-authentication process. Tahk, however, teaches using oval prompts to frame a user's face. *See* §XIII.D.2 (claim 4). A POSITA would have been motivated to modify Zhang, alone or in combination with Tanii, to provide oval-shaped prompts because they are a natural shape to appropriately size and frame a face at different distances. *See* §XIII.D.1 (motivation).

E. Ground 2C: Zhang, Tanii, and Hoyos (Claim 18)

1. Motivation to Combine

1133. Zhang discloses implementing a process to verify the three-dimensionality of a user's face by capturing a series of images of a user using a mobile computing device, such as a phone or laptop. *See, e.g.*, *Zhang*, [0013]. In my opinion, a POSITA would have understood that mobile devices are often provided with user-facing cameras, particularly in mobile devices.

1134. Although Zhang discloses using homography transformation to distinguish real, three-dimensional faces from pictures of a face, *see* §VII.C (Zhang), in my opinion, a POSITA would have appreciated that Zhang's process may be spoofed by presenting a non-planar picture of a face, because Zhang's homography transformation process specifically looks for a *planar* structure in images that can be transformed nearly identically from one perspective to another. *See id.* Thus, Zhang's system could possibly be subverted by bending the picture in a way to trick the system, or applying the picture to a three-dimensional shape. For this reason, in my opinion, a POSITA would have therefore been motivated to look for secondary methods to ensure the user's face is from the user, and not a spoofer. In my experience, biometric systems often included multiple independent checks to ensure the liveness of the user (Derakhshani, for instance, is one example that provided separate spatial, behavioral, and reflectance metrics to each independently confirm liveness of the user).

1135. A POSITA would have known that the use of reflectance of light off a face was a well-known liveness check. *See, e.g.*, *Derakhshani*, 18:5-19:11; *Hoyos*, [0018]-[0019], [0033]-[0036], Figs. 2-3. And Hoyos exemplifies this process by disclosing the use of patterned images intended to reflect off of the user's face.

Hoyos, [0018]-[0019], [0033]-[0036], Figs. 2-3. In my opinion, a POSITA would have been motivated to incorporate a secondary liveness check based on reflectance, consistent with Hoyos, to ensure a user is not attempting to spoof Zhang's homography-transformation process.

1136. In my opinion, a POSITA would have had a reasonable expectation of success in making this modification because Zhang's homography transformation and Hoyos's reflectance measure operate on two distinct, modular principles that can be operated together; Zhang requires two images from two different perspectives, and Hoyos requires reflecting different light patterns during image capture.

2. Claim 18: The method according to claim 10, further comprising illuminate a screen of a computing device while capturing the at least one first image and/or the at least one second image, and processing the at least one first image and/or the at least one second image to detect a reflection of the illumination from a face of the user.

1137. In my opinion, Zhang combined with Tanii and Hoyos teaches claim 18's additional limitation.

1138. Zhang does not expressly disclose displaying an image on the device's screen when capturing images, or detecting a reflection of the displayed image off of the user's face in the captured images. Hoyos, however, teaches that measuring reflectance of displayed images is a well-known method to verify the liveness of the user. *Hoyos*, [0018]-[0019]; [0033]-[0035]. In my opinion, a POSITA would have been motivated to modify Zhang to incorporate Hoyos's reflectance detection to provide an additional verification that the user is presenting a real, three-dimensional face. *See* §XIII.E.1 (motivation).

XIV. SECONDARY CONSIDERATIONS

1139. At this time, I am not aware of any secondary considerations of non-obviousness, or evidence sufficient to change my opinions that each of the Challenged Patents are obvious in view of the prior art. To the extent Patent Owner identifies any evidence of secondary considerations of non-obviousness, I reserve the right to respond to such evidence at that time.

XV. ADDITIONAL REMARKS

1140. I, Chris Daft, 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.

Dated: November 6, 2024

Respectfully submitted,

CHW Daft

____

Chris Daft, D. Phil.

'471 PATENT - LISTING OF CHALLENGED CLAIMS

Reference	Claim Limitations
Claim 1	
1[pre]	A system for authenticating three-dimensionality of a user via a user's camera equipped computing device, the computing device comprising:
1[a]	a processor configured to execute machine executable code;
1[b]	a screen configured to provide a user interface to the user;
1[c]	a camera configured to capture images;
1[d]	one or more memories configured to store machine readable instructions that are stored on the memory of the authentication server which when executed by the processor, cause the computing device to:
1[d1]	capturing at least one first image of the user taken with the camera of the computing device at a first location which is a first distance from the user;
1[d2]	processing the at least one first image or a portion to create first data;
1[d3]	moving the camera from the first location to a second location, the second location being a second distance from the user, or the user moving from the first location to the second location to change the distance between the user and the camera from the first distance to a second distance;
1[d4]	capturing at least one second image of the user taken with the camera of the computing device at the second distance from the user, the second distance being different than the first distance;
1[d5]	comparing the first data to the second data to determine whether expected differences exist between the first data and the second data which indicated three-dimensionality of the user;
1[d6]	authenticating the user when differences between the first data and the second data have expected distortion resulting from movement of the camera from the first location to the second location or movement of the user from the first location to the second location,

	which causes the change in distance between the user and the camera.
Claim 2	
2[a]	The system according to claim 1, further comprising: interpolating the first data and the second data to obtain estimated intermediate data;
2[b]	capturing at least one third image of the user taken with the camera of the computing device at a third distance from the user, the third distance being between the first distance and the second distances;
2[c]	processing the at least one third image or a portion thereof to obtain third data; and
2[d]	comparing the estimated intermediate data with the third data to determine whether the third data matches the estimated intermediate data.
Claim 3	
3	The system according to claim 1, further comprising verifying the presence of the user's ears in the at least one first image, and verifying the absence or reduced visibility of the user's ears in the at least one second image, wherein the first distance is larger than the second distance.
Claim 4	
4	The system according to claim 1, wherein the computing device is configured to display one or more prompts on a screen of the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.
Claim 5	
5[a]	The system according to claim 1, further comprising comparing the first data, second data, or both to enrollment data derived from an enrollment image, the enrollment image captured and stored prior to an authenticating; and
5[b]	only authenticating the user when the first data, the second data, or both match the enrollment data within a predetermined threshold.

Claim 6	
6	The system according to claim 1, wherein the computing device is a hand-held device, and the user holds the device at the first and second distance to capture the at least one first image and the at least one second image.
Claim 7	
7	The system according to claim 1, wherein the first data and the second data comprise biometric data.
Claim 8	
8	The system according to claim 1, wherein the first data and the second data comprise a mapping of facial features.
Claim 9	
9	The method according to claim 1, wherein the first image and the second image is of the user's face and the user's face is held steady and without movement during capture of the first image and the second image.
Claim 10	
10[pre]	A method for authenticating three-dimensionality of a user via a user's camera equipped computing device, the method, during an authentication session comprising:
10[a]	capturing at least one first image of the user taken with the camera of the computing device at a first location which is a first distance from the user;
10[b]	processing the at least one first image or a portion to create first data;
10[c]	moving the camera from the first location to a second location, the second location being a second distance from the user, or the user moving from the first location to the second location to change the distance between the user and the camera from the first distance to a second distance;
10[d]	capturing at least one second image of the user taken with the camera of the computing device at the second distance from the user, the second distance being different than the first distance;

10[e]	processing the at least one second image or a portion thereof to create second data;
10[f]	comparing the first data to the second data to determine whether expected distortion exist between the first data and the second data which indicated three-dimensionality of the user;
10[g]	authenticating the user when the differences between the first data and the second data have expected distortion resulting from movement of the camera from the first location to the second location or movement of the user from the first location to the second location, which causes the change in distance between the user and the camera.
Claim 11	
11[a]	The method according to claim 10, further comprising: interpolating the first data and the second data to obtain estimated intermediate data;
11[b]	capturing at least one third image of the user taken with the camera of the computing device at a third distance from the user, the third distance being between the first distance and the second distances;
11[c]	processing the at least one third image or a portion thereof to obtain third data; and
11[d]	comparing the estimated intermediate data with the third data to determine whether the third data matches the estimated intermediate data.
Claim 12	
12	The method according to claim 10, further comprising verifying the presence of the user's ears in the at least one first image, and verifying the absence or reduced visibility of the user's ears in the at least one second image, wherein the first distance is larger than the second distance.
Claim 13	
13	The method according to claim 10, wherein the computing device is configured to display one or more prompts on a screen of the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.

Claim 14	
14	The method according to claim 13, wherein the one or more prompts are ovals on the screen within which the face of the user is placed to capture the at least one first image and the at least one second image.
Claim 15	
15	The method according to claim 10, wherein the computing device is a hand-held device, and the user holds the device at the first and second distances to capture the at least one first image and the at least one second image.
Claim 16	
16	The method according to claim 10, wherein the first data and the second data comprise biometric data.
Claim 17	
17	The method according to claim 10, wherein the first data and the second data comprise a mapping of facial features.
Claim 18	
18	The method according to claim 10, further comprising displaying an image on a screen of the computing device while capturing the at least one first image and/or the at least one second image, and processing the at least one first image and/or the at least one second image to detect a reflection of the displayed image off of the user's face.
Claim 19	
19	The method according to claim 10, wherein the user's face is held steady and the camera moves from the first location to the second location.
Claim 20	
20	The method according to claim 10, wherein the first data and the second data are maintained on the computing device.

’606 PATENT - LISTING OF CHALLENGED CLAIMS

Reference	Claim Limitations
Claim 1	
1[pre]	A method for verifying three-dimensionality of a user's face using images of the user's face captured using a camera equipped computing device, the method comprising:
1[a]	capturing at least one first image of the user taken with the camera of the computing device at a first distance from the user;
1[b]	processing the at least one first image to obtain first biometric data from the at least one first image;
1[c]	capturing at least one second image of the user taken with the camera of the computing device at a second distance from the user, the second distance being different than the first distance;
1[d]	processing the at least one second image to obtain second biometric data based on the at least one second image;
1[e]	comparing the first biometric data with the second biometric data to determine whether the first biometric data matches the second biometric data;
1[f]	comparing the first biometric data to second biometric data to determine whether differences between the at least one first image and the at least one second image match expected differences resulting from movement of the camera or the user which changed the distance between the user and camera from the first distance to the second distance;
1[g]	determining that the user's face is three-dimensional when:
1[h]	the first biometric data does not match the second biometric data; and
1[i]	the second biometric data has the expected differences as compared to the first biometric data resulting from the change in distance between the user and the camera when capturing the at least one first image and the at least one second image.

Claim 2	
2[a]	The method according to claim 1, further comprising: interpolating the first biometric data and the second biometric data to obtain estimated intermediate biometric data;
2[b]	capturing at least one third image of the user taken with the camera of the computing device at a third distance from the user, the third distance being between the first distance and the second distance;
2[c]	processing the at least one third image to obtain third biometric data based on the at least one third image; and
2[d]	comparing the estimated intermediate biometric data with the third biometric data to determine whether the third biometric data matches the estimated intermediate biometric data.
Claim 3	
3	The method according to claim 1, further comprising verifying the presence of the user's ears in the at least one first image, and verifying the absence or reduced visibility of the user's ears in the at least one second image, wherein the first distance is larger than the second distance.
Claim 4	
4	The method according to claim 1, wherein the computing device is configured to display one or more prompts on a screen of the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.
Claim 5	
5	The method according to claim 4, wherein the one or more prompts are ovals sized on the screen within which the face of the user is placed to capture the at least one first image and the at least one second image at the first and second distances.

Claim 6	
6	The method according to claim 4, wherein the computing device is a hand-held device, and the user holds the computing device at the first distance and the second distance to capture the at least one first image and the at least one second image.
Claim 7	
7	The method according to claim 6, wherein the computing device comprises a laptop or desktop computer and, with the computing device stationary, the user moves from the first distance to the second distance to capture the at least one first image and the at least one second image.
Claim 8	
8	The method according to claim 1, further comprising displaying an image on a screen of the computing device while capturing the at least one first and/or the at least one second image.
Claim 9	
9	The method according to claim 1, wherein the first biometric data and the second biometric data are transmitted over a network to a server.
Claim 10	
10[pre]	A system for verifying three-dimensionality of a user's face using images of the user's face captured using a camera equipped computing device, the system comprising:
10[a]	a computing device having a camera, screen, processor, and memory configured with non-transitory machine readable code that is executable by the processor, the machine readable code configured to:
10[b]	capture at least one first image of the user taken with the camera of the computing device when the camera is located a first distance from the user;

10[c]	process the at least one first image to obtain first biometric data from the at least one first image;
10[d]	capture at least one second image of the user taken with the camera of the computing device when the camera is located a second distance from the user, the second distance being different than the first distance;
10[e]	process the at least one second image to obtain second biometric data based on the at least one second image;
10[f]	compare the first biometric data to the second biometric data to determine whether the first biometric data matches the second biometric data;
10[g]	compare the first biometric data to second biometric data to determine whether differences between the at least one first image and the at least one second image match expected differences resulting from movement of the camera or the user which changed the distance between the user and camera from the first distance to the second distance;
10[h]	determine that the user's face is three-dimensional when:
10[h1]	the first biometric data does not match the second biometric data; and
10[h2]	the second biometric data has the expected differences as compared to the first biometric data resulting from the change in distance between the user and the camera.
Claim 11	
11[a]	The method according to claim 10, further comprising: interpolating the first biometric data and the second biometric data to obtain estimated intermediate biometric data;
11[b]	capturing at least one third image of the user taken with the camera of the computing device at a third distance from the user, the third distance being between the first distance and the second distance;
11[c]	processing the at least one third image to obtain third biometric data based on the at least one third image; and

11[d]	comparing the estimated intermediate biometric data with the third biometric data to determine whether the third biometric data matches the estimated intermediate biometric data.
Claim 12	
12	The method according to claim 10, further comprising verifying the presence of the user's ears in the at least one first image, and verifying the absence or reduced visibility of the user's ears in the at least one second image, wherein the first distance is larger than the second distance.
Claim 13	
13	The method according to claim 10, wherein the computing device is configured to display one or more prompts on a screen of the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.
Claim 14	
14	The method according to claim 10, wherein comparing the first biometric data to the second biometric data and the determining that the user's face is three-dimensional occurs at a server that is remote from the camera equipped computing device.
Claim 15	
15	The method according to claim 13, wherein the computing device is a hand-held device, and the user holds the computing device at the first distance and the second distance to capture the at least one first image and the at least one second image.
Claim 16	
16	The method according to claim 10, wherein the computing device comprises a laptop or desktop computer and, with the computing device stationary, the user moves from the first distance to the second distance.
Claim 17	

17	The method according to claim 10, further comprising displaying an image on a screen of the computing device while capturing the at least one first image and the at least one second image.
Claim 18	
18	The method according to claim 10, wherein the first biometric data and the second biometric data are maintained on the computing device.
Claim 19	
19[pre]	A method for verifying three-dimensionality of a user's face using images of the user's face captured using a camera equipped computing device, the method comprising:
19[a]	receiving first biometric data generated from at least one first image of the user taken with the camera of the computing device located at a first distance from the user;
19[b]	receiving second biometric data generated from at least one second image of the user taken with the camera of the computing device located at a second distance from the user, the second distance being different than the first distance;
19[c]	comparing the first biometric data with the second biometric data to determine whether the first biometric data matches the second biometric data;
19[d]	comparing the first biometric data with second biometric data to determine whether differences between the at least one first image and the at least one second image match expected differences resulting from movement of the camera or the user which changes the distance between the user and camera for capture of the one or more first images at the first distance and capture of the one or more second images at the second distance;
19[e]	determining that the user's face is three-dimensional when:
19[e1]	the first biometric data is not identical to the second biometric data; and

19[e2]	the second biometric data has expected differences as compared to the first biometric data, the expected differences resulting from the change in distance between the user and the camera when the at least one first image was captured at the first distance and the at least one second image was captured at the second distance.
Claim 20	
20	The method of claim 19, wherein the receiving of the first biometric data and the second biometric data occurs at a server and the first biometric data and the second biometric data are received over one or more of a LAN, WAN, or Internet type network.

’938 PATENT - LISTING OF CHALLENGED CLAIMS

Reference	Claim Limitations
Claim 1	
1[pre]	A non-transient computer readable medium containing non-transitory machine executable code configured to determine if the three-dimensional shape is consistent with that of a human face, the non-transitory machine executable code configured to:
1[a]	receive or derive first biometric data from at least one first image of a user taken with a computing device camera located at a first distance from the user;
1[b]	receive or derive second biometric data from at least one second image of the user taken with the computing device camera located at a second distance from the user, the second distance being different than the first distance;
1[c]	compare the first biometric data with second biometric data for expected differences that result from characteristics of a human face and the at least one first image and the at least one second image being captured at different distances from the user;
1[d]	determine that the three-dimensional shape is not exhibited when the second biometric data does not have expected differences compared to the first biometric data, the expected differences comprising at least differences due to the change in the relative distance between the user's facial features and the camera when the at least one first image was captured at the first distance and the at least one second image was captured at the second distance, wherein the expected differences result from fish-eye type distortion in at least one of the at least one first image and the at least one second image and due to the three-dimensional nature of the human face and the change in distance between the camera and the user.
Claim 2	
2	The non-transient computer readable medium of claim 1 wherein the expected differences appear as changes in the relative size and shape of facial features of the user.

Claim 3	
3	The non-transient computer readable medium of claim 1 wherein determining that three-dimensionality is not exhibited happens during an authentication session.
Claim 4	
4	The non-transient computer readable medium of claim 1 wherein the computing device camera is part of a computing device and the machine executable code is configured to display an interface on the computing device's screen to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.
Claim 5	
5[a]	The non-transient computer readable medium of claim 1 wherein the machine executable code is further configured to compare at least portions of the first data, second data, or both to enrollment data derived from an enrollment image, the enrollment image captured and stored prior to an authentication session; and
5[b]	determining the user is not authenticated when the first data, the second data, or both do not sufficiently correspond to the enrollment data.
Claim 6	
6	The non-transient computer readable medium of claim 1 wherein the computing device camera is part of a computing device and the computing device is a hand-held device, and the user holds the device at the first distance to capture the at least one first image and then holds the computing device at the second distance to capture the at least one second image.
Claim 7	
7	The non-transient computer readable medium of claim 1 wherein the first biometric data and the second biometric data comprise image data of facial features.

Claim 8	
8[pre]	A method for determining when a user, based on images of the user's face, does not exhibit three-dimensionality, the method comprising:
8[a]	capturing at least one first image of the user's face taken with a camera located at a first distance from the user's face, the camera associated with a computing device;
8[b]	processing the at least one first image or a portion thereof to create first data;
8[c]	moving the camera to a second distance from the user's face, where the second distance is different from the first distance;
8[d]	capturing, at the second distance, at least one second image of the user's face taken with the camera associated with the computing device;
8[e]	processing the at least one second image or a portion thereof to create second data;
8[f]	examining the first data and the second data to determine whether differences between the first data and the second data indicate an expected type of distorting change in at least one image that is consistent with a real person being imaged and which is indicative of three-dimensionality;
8[g]	determining the user's face is not three-dimensional when the first data and the second data do not have expected differences indicating the user exhibits three-dimensionality.
Claim 9	
9[a]	The method of claim 8 further comprising: capturing one or more additional images at distances from the user's face that are between the first distance and the second distance;
9[b]	for at least one of the one or more additional images, generating additional data;
9[c]	examining the additional data, the first data, and the second data, or portions thereof, to determine whether expected differences therebetween indicate the user's face exhibits three-dimensionality.

Claim 10	
10	The method of claim 8 further comprising displaying one or more prompts on a screen associated with the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.
Claim 11	
11	The method of claim 10 wherein the one or more prompts are an on the screen shape within which an image of a face of the user is aligned during capture the at least one first image and the at least one second image.
Claim 12	
12	The method of claim 8 wherein the computing device is a hand-held device, and the user holds the computing device at the first distance from the user's face when capturing at least one first image and holds the computing device at the second distance from the user's face when capturing the at least one second image.
Claim 13	
13	The method of claim 8 wherein the first data and the second data comprise at least in part biometric data.
Claim 14	
14	The method of claim 8 wherein moving the camera comprises moving the camera linearly toward or away from the user's face.
Claim 15	
15	The method of claim 8 further comprising illuminate a screen of the computing device while capturing the at least one first image and/or the at least one second image to improve quality of an image being captured.
Claim 16	
16	The method of claim 8 wherein a face of the user is held steady when capturing the at least one first image and the at least one second image and the camera moves from the first location to the second location.

Claim 17	
17[pre]	A method, performed using a computing device, for providing authentication of a person during an authentication session, the method comprising:
17[a]	capturing a first image of a head of the person with a camera at a first distance from the person, the camera associated with the computing device;
17[b]	changing a distance between the person and the camera to a second distance, which is different from the first distance;
17[c]	capturing a second image of the head of the person with the camera at the second distance from the person;
17[d]	comparing one or more aspects of the head from the first image or first biometric data derived from the first image to one or more aspects of the head from the second image or second biometric data derived from the second image to determine whether expected differences are not present, wherein the expected differences:
17[e]	would be present when the first image and second images of the head of the person being captured at different distances has three-dimensional characteristics but not if the head did not have three-dimensional characteristics; and
17[f]	the expected differences result from differences in relative dimensions of a person's face appearing different when capturing images is done close to the person's face and far from the persons face; and
17[g]	if the expected differences are not present, denying authentication of the person and providing notice thereof to one or more of the person, a third party, or a software application, wherein the authentication is authentication of liveness, three-dimensionality, or both.
Claim 18	
18	The method of claim 17 wherein the steps of comparing, denying authentication, and providing notice are performed by a server that is remote from the computing device.

Claim 19	
19	The method of claim 17 wherein the authentication is authentication of three-dimensionality.
Claim 20	
20[pre]	A method for determining whether a user exhibits three-dimensionality, the method comprising:
20[a]	capturing at least one first image of a user's face taken with a camera located a first distance from the user, the camera associated with a computing device;
20[b]	processing the at least one first image or a portion thereof to create first data, the first data derived from the user's face;
20[c]	intentionally moving the camera from the first location to a second location, the second location being a second distance from the user, or the user moving to change a distance between the user and the camera from the first distance to the second distance;
20[d]	capturing at least one second image of the user's face taken with the camera located a second distance from the user, the second distance being different than the first distance;
20[e]	processing the at least one second image or a portion thereof to create second data, the second data derived from the user's face;
20[f]	analyzing the first data to determine at least if the first data exhibits first characteristics that indicate the first data was derived from an image of the user captured at the first distance;
20[g]	analyzing the second data to determine at least if the second data exhibits second characteristics that indicate the second data was derived from an image the user captured at the second distance, wherein the first characteristics or the second characteristics include at least distortion within the at least one first image or the at least one second image;
20[h]	determining the user does not exhibit the expected degree of three-dimensionality when either or both of the following occur:
20[h1]	the step of analyzing the first data determines the first data does not exhibit first characteristics that indicate the first data was derived from an image of the user captured at the first distance; or

20[h2]	the step of analyzing the second data determines the second data does not exhibit second characteristics that indicate the second data was derived from an image of the user captured at the second distance.
Claim 21	
21	The method of claim 20 further comprising displaying one or more prompts on a screen associated with the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.
Claim 22	
22	The method of claim 20 wherein the at least one first image and the at least one second image are captured with a hand-held computing device, and the user holds the computing device at the first distance when capturing at least one first image and at the second distances when capturing the at least one second image.
Claim 23	
23	The method of claim 20 wherein the first data and the second data comprise at least in part biometric data.
Claim 24	
24	The method of claim 20 wherein the first data and the second data comprise at least in part image data of facial features.

’910 PATENT - LISTING OF CHALLENGED CLAIMS

Reference	Claim Limitations
Claim 1	
1[pre]	A computing device for verifying three-dimensionality of a user via a user's camera equipped computing device, the computing device comprising:
1[a]	a processor configured to execute machine executable code;
1[b]	a screen configured to provide a user interface to the user;
1[c]	a camera configured to capture images;
1[d]	one or more memories configured to store machine readable instructions that are stored on the memory of the authentication server which when executed by the processor, cause the computing device to:
1[d1]	capturing at least one first image of the user taken with the camera of the computing device at a first location which is a first distance from the user;
1[d2]	processing the at least one first image or a portion thereof to create first data;
1[d3]	capturing at least one second image of the user taken with the camera of the computing device is at a second distance from the user, the second distance being different than the first distance, the capturing at least one second image of the user occurring after movement of the camera or the user to establish the camera at the second distance from the user;
1[d4]	processing the at least one second image or a portion thereof to create second data;
1[d5]	comparing the first data to the second data to determine whether expected differences exist between the first data and the second data which indicates three-dimensionality of the user;
1[d6]	verifying the images of the user exhibit three dimensional traits when the expected differences exist between the first data and the second data as a result of capturing the at least one first image and the at least one second image at different distances from the user.

Claim 2	
2[a]	The system according to claim 1, further comprising: interpolating the first data and the second data to obtain estimated intermediate data;
2[b]	capturing at least one third image of the user taken with the camera of the computing device at a third distance from the user, the third distance being between the first distance and the second distances;
2[c]	processing the at least one third image or a portion thereof to obtain third data; and
2[d]	comparing the estimated intermediate data with the third data to determine whether the third data matches the estimated intermediate data.
Claim 3	
3	The system according to claim 1, further comprising verifying the presence of one or more features on a side of a user's head in the at least one first image, and verifying the absence or reduced visibility of the one or more features on the side of the user's head in the at least one second image due to image capture at different distances from the user's head, wherein the first distance is larger than the second distance.
Claim 4	
4	The system according to claim 1, wherein the machine readable instructions is configured to display one or more prompts on the screen of the computing device to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.
Claim 5	
5[a]	The system according to claim 1, further comprising comparing the first data, second data, or both to enrollment data derived from an enrollment image, the enrollment image captured and stored prior to an authenticating; and
5[b]	only authenticating the user when the first data, the second data, or both match the enrollment data within a predetermined threshold.

Claim 6	
6	The system according to claim 1, wherein the computing device is a hand-held device, and the user holds the device at the first and second distance to capture the at least one first image and the at least one second image.
Claim 7	
7	The system according to claim 1, wherein the first data and the second data comprise biometric data.
Claim 8	
8	The system according to claim 1, wherein the first data and the second data comprise a mapping of facial features.
Claim 9	
9	The method according to claim 1, wherein the first image and the second image is of the user's face and the user's head and facial features are held steady and without movement during capture of the first image and the second image.
Claim 10	
10[pre]	A method for evaluating three-dimensionality of a user, the method comprising:
10[a]	capturing at least one first image of the user taken with a camera at a first location which is a first distance from the user;
10[b]	processing the at least one first image or a portion thereof to create first data;
10[c]	moving the camera from the first location to a second location, the second location being a second distance from the user, or the user moving to change the distance between the user and the camera from the first distance to the second distance;
10[d]	capturing at least one second image of the user taken with the camera when the camera is the second distance from the user, the second distance being different than the first distance;
10[e]	processing the at least one second image or a portion thereof to create second data;

10[f]	comparing the first data to the second data to determine whether expected differences exist between the first data and the second data which indicate three-dimensionality of the user;
10[g]	verifying the images of the user exhibit three-dimensional traits when the first data and the second data have expected differences resulting from the at least one first image being captured with the camera at a different distance from the user than when the at least one second image is captured.
Claim 11	
11[a]	The method according to claim 10, further comprising: interpolating the first data and the second data to obtain estimated intermediate data;
11[b]	capturing at least one third image of the user taken with the camera at a third distance from the user, the third distance being between the first distance and the second distances;
11[c]	processing the at least one third image or a portion thereof to obtain third data; and
11[d]	comparing the estimated intermediate data with the third data to determine whether the third data matches the estimated intermediate data.
Claim 12	
12	The method according to claim 10, further comprising verifying the presence of ears of the user in the at least one first image, and verifying the absence or reduced visibility of the ears in the at least one second image, wherein the first distance is larger than the second distance.
Claim 13	
13	The method according to claim 10, further comprising one or more prompts on a screen to guide the user to capture the at least one first image at the first distance and the at least on second image at the second distance.
Claim 14	
14	The method according to claim 13, wherein the one or more prompts are an oval shape guide on the screen within which an

	image of a face of the user is aligned to capture the at least one first image and the at least one second image.
Claim 15	
15	The method according to claim 10, wherein the camera is part of a computing device is a hand-held device, and the user holds the computing device at the first distance when capturing at least one first image and at the second distances when capturing the at least one second image.
Claim 16	
16	The method according to claim 10, wherein the first data and the second data comprise biometric data.
Claim 17	
17	The method according to claim 10, wherein the first data and the second data comprise a map of facial features.
Claim 18	
18	The method according to claim 10, further comprising illuminate a screen of a computing device while capturing the at least one first image and/or the at least one second image, and processing the at least one first image and/or the at least one second image to detect a reflection of the illumination from a face of the user.
Claim 19	
19	The method according to claim 10, wherein a face of the user is held steady when capturing the at least one first image and the at least one second image and the camera moves from the first location to the second location.
Claim 20	
20	The method according to claim 10, wherein the first data and the second data are maintained on a computing device.
Claim 21	
21	The method of claim 10 wherein the camera is part of is one of a smartphone, tablet, laptop, or desktop computer.

Claim 22	
22[pre]	A method, performed by a user using a user's computer device, for verifying three-dimensionality of the user, the method comprising:
22[a]	capturing a first image of the user's head with a camera at a first distance from the user, the camera associated with the user's computing device;
22[b]	changing a distance between the user and the camera to a second distance by the user moving the camera, or the user moving relative to the camera, or both;
22[c]	capturing a second image of the user's head with the camera when the camera is at the second distance from the user, the second distance being different than the first distance;
22[d]	comparing one or more aspects of the user's head from the first image to one or more aspects of the user's head from the second image to determine whether expected differences, between the first image and the second image, exist which indicates three-dimensionality of the user, such that the expected differences between the first image and the second image result from the first image being captured when the camera is at a different distance from the user than when the second image is captured; and
22[e]	responsive to the comparing determining that expected differences between the first image and the second image exist, providing notice to the user, a third party, or both that the three-dimensionality of the user is verified.
Claim 23	
23	The method of claim 22 wherein the one or more aspects of the user's head from the first image is first data resulting from processing the first image and the one or more aspects of the user's head from the second image is second data resulting from processing the second image.
Claim 24	
24	The method of claim 22 wherein the user's head is the user's face.