

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

META PLATFORMS, INC., INSTAGRAM, INC., WHATSAPP LLC,
META PLATFORMS TECHNOLOGIES, LLC, AND GIPHY, INC.,
Petitioner,

v.

VL COLLECTIVE IP LLC,
Patent Owner.

IPR2023-00924
Patent 7,436,980 B2

Before KARL D. EASTHOM, JEFFREY S. SMITH, and
DAVID C. McKONE, *Administrative Patent Judges*.

SMITH, *Administrative Patent Judge*.

DECISION
Granting Institution of *Inter Partes* Review
35 U.S.C. § 314

I. INTRODUCTION

A. *Background and Summary*

Meta Platforms, Inc, Instagram, Inc., WhatsApp LLC, Meta Platforms Technologies, LLC, and GIPHY, Inc. (collectively, “Petitioner”) filed a Petition (Paper 3, “Pet.”) requesting *inter partes* review of claims 1–16 of U.S. Patent No. 7,436,980 B2 (Ex. 1001, “the ’980 patent”) pursuant to 35 U.S.C. § 311(a). VL Collective IP LLC (“Patent Owner”) filed a Preliminary Response (Paper 7, “Prelim. Resp.”) pursuant to 35 U.S.C. § 313.

Pursuant to 35 U.S.C. § 314(a), the Director may not authorize an *inter partes* review unless the information in the petition and preliminary response “shows that there is a reasonable likelihood that the petitioner would prevail with respect to at least 1 of the claims challenged in the petition.” For the reasons that follow, we institute an *inter partes* review as to claims 1–16 of the ’980 patent on the grounds of unpatentability asserted in the Petition.

B. *Real Parties In Interest*

Petitioner identifies “Meta Platforms, Inc., Instagram, Inc., WhatsApp LLC, Meta Platforms Technologies, LLC, and GIPHY, Inc.” as its real parties in interest. Pet. 2. Patent Owner identifies “VL Collective[,]
... VL IP Holdings LLC, which is a parent company of VL Collective, and VideoLabs, Inc., which is a parent company of VL IP Holdings LLC,” as its real parties in interest. Paper 8, 2.

C. Related Matters

Petitioner and Patent Owner state that the '980 patent is the subject of the following pending district court proceeding: *VideoLabs, Inc. v. Meta Platforms, Inc.*, No. 1-22-cv-00680 (D. Del.), filed May 24, 2022. Pet. 2; Paper 8, 2.

II. THE '980 PATENT (Ex. 1001)

The '980 patent “relates to image processing, and more particularly to automatic detection and tracking of objects in images.” Ex. 1001, 1:11–13. In describing the invention’s background, the patent states:

The problem of describing and recognizing categories of objects (e.g., faces, people, cars) is important to computer vision applications. It is common to represent objects as collections of features with distinctive appearance, spatial extent, and position. There is however a large variation in how many features are needed and how these features are detected and represented.

Therefore, a need exists for a system and method of detecting and tracking an object, implementing component detection and performing inference over space and time.

Id. at 1:15–24. The patent generalizes its solution as follows:

According to an embodiment of the present disclosure, a probabilistic framework for automatic component-based detection and tracking of objects in images and/or video combines object detection with tracking in a unified framework. Tracking makes use of object detection for initialization and re-initialization during transient failures for occlusions. Object detection considers the consistency of the detection over time. Modeling objects by an arrangement of image-base, and possibly overlapping, components facilitates detection of complex articulated objects as well as helps in handling partial object occlusions or local illumination changes.

Id. at 2:43–54.

Figures 2A and 2B are reproduced below.

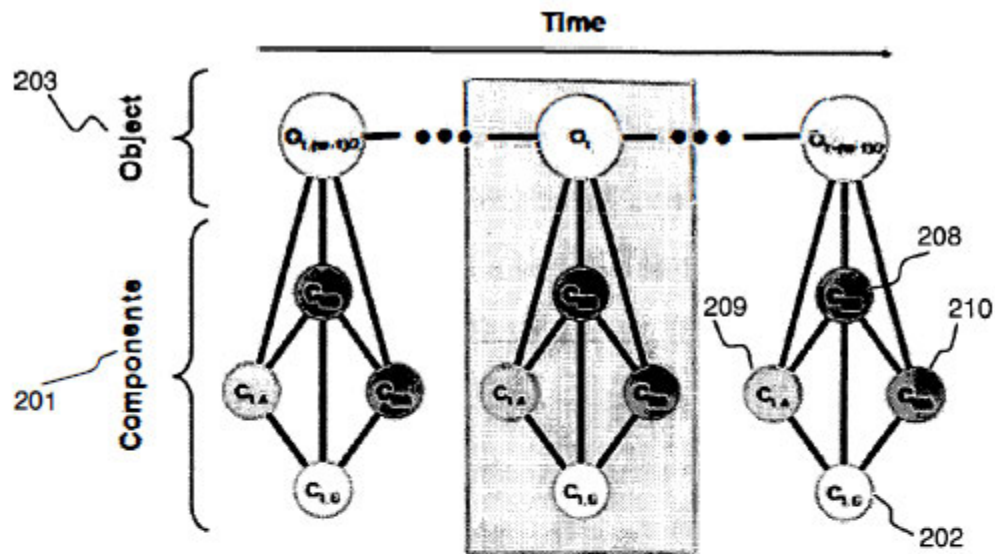


FIGURE 2A

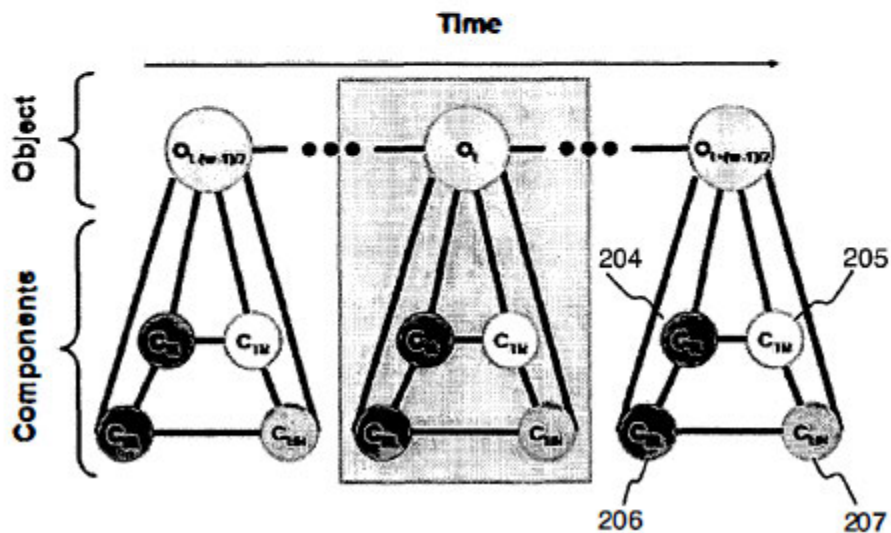


FIGURE 2B

Figures 2A and 2B above describe the building of a graphical model corresponding to an object (*id.* at 4:18–20), and respectively show “two-layer spatial graphical model[s] . . . for pedestrian and vehicle detection” (*id.* at 4:19, 4:23–26). “[A] fine, component, layer 201 includes a set of loosely connected parts, e.g., 202.” *Id.* at 4:20–21. “[A] course,

object, layer 203 corresponds to an entire appearance model of the object and is connected to all constituent components, e.g., 202.” *Id.* at 4:22–24. A model is shown over time for three sequential instances of time (i.e., three frames); the shaded region (center) o_t representing an instance of the graphical model of the object at time t . *Id.* at 4:18, 4:24–26, 4:34–37, 4:46–48. “[O]bject detection and tracking is formulated as an inference in [such] a two-layer graphical model in which a coarse layer node represents the whole object and fine layer nodes represent multiple components of the object.” *Id.* at 2:55–59.

Both sets of models/nodes (i.e., Figure 2A’s pedestrian and Figure 2B’s vehicle models) are “modeled using four overlapping image components.” *Id.* at 4:26–27. “For the vehicle the components are: top-left (TL) 204, top-right (TR) 205, bottom-right (BR) 206 and bottom-left (BL) 207 corners.” *Id.* at 4:27–31. “[For] the pedestrian, they are: head (HD) 208, left arm (LA) 209, right arm (RA) 210 and legs (LG) 202.” *Id.* at 4:31–33. “[The] two-layer graphical model allows the inference process to reason explicitly about the object as a whole, e.g., 203, and reduce the complexity of the graphical model by allowing the assumption of the conditional independence of components, e.g., 202 and 208–210, over time given the overall object appearance.” *Id.* at 4:37–42.

Figure 3 is reproduced below.

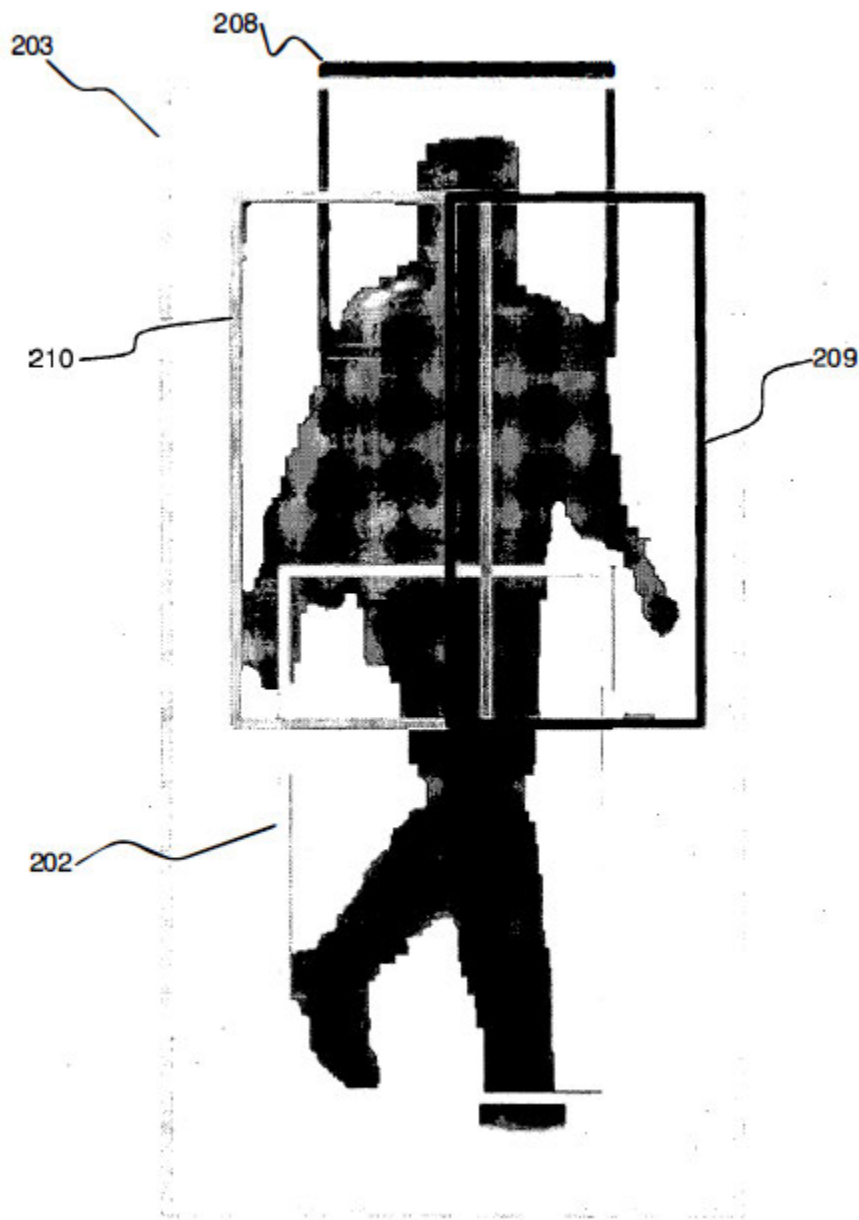


Figure 3 above shows pedestrian's components 202, 208–210, and “is an illustration of [the] pedestrian and identified components of the pedestrian” (*id.* at 2:31–32).

The patent discloses that “[w]hile it is possible to perform inference over the spatio-temporal model defined for the entire image sequence, there are many applications for which this is not an option due to the lengthy off-line processing needed.” Ex. 1001, 7:6–9. The patent discloses that a “w-

frame windowed smoothing algorithm is used where w is an odd integer ≥ 1 .” *Id.* at 7:9–10. The patent discloses that “with $w=1$, the algorithm resembles single frame component-based fusion.” *Id.* at 7:20–22.

The “method for object detection includes providing a spatio-temporal model, e.g., see FIGS. 2A and 2B, for an object 501[;] providing a video including a plurality of images including the object 502[;] measuring the object as a collection of components in each image of the video 503.” *Id.* at 7:45–50. The method further includes “determining a probability that the object is in each image 504 by using message passing between components represented as nodes of the spatio-temporal model, and detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object 505.” *Id.* at 7:50–55.

III. ILLUSTRATIVE CLAIM

Independent claim 1 of the ’980 patent is reproduced, below, with bracketed annotations inserting Petitioner’s identifiers of claim limitations (Pet. 81–83):

[1preamble] A computer implemented method for object detection comprising:

[1a] providing a spatio-temporal model for an object to be detected;

[1b] providing a video comprising a plurality of images including the object;

[1c] measuring the object as a collection of components in each image;

[1d] determining a probability that the object is in each image; and

[1e] detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.

IV. ASSERTED GROUNDS

Petitioner asserts that claims 1–16 of the '980 patent are unpatentable on the following grounds (Pet. 22).

Claim(s) Challenged	35 U.S.C. §¹	Reference(s)/Basis
1–16	103(a)	Zhao ²
1–16	103(a)	Steffens ³ and Zhao
1–16	103(a)	Ozer ⁴ and Zhao
1–16	103(a)	TLP ⁵

V. LEVEL OF ORDINARY SKILL

Petitioner identifies a person of ordinary skill in the art (“POSITA”) as someone with: “a bachelor’s degree in computer science, computer engineering, computer vision or visualization, or an equivalent field, and approximately two years of experience in software development for computer vision applications. EX1004, ¶16.” Pet. 11. Petitioner further asserts that “[a]dditional education might compensate for less experience and vice-versa.” *Id.*

¹ The Leahy-Smith America Invents Act, Pub. L. No. 112-29, 125 Stat. 284 (2011) (“AIA”), amended 35 U.S.C. § 103. Because the '980 patent has an effective filing date prior to the effective date of the applicable AIA amendment, we refer to the pre-AIA version of §103.

² Liang Zhao, *Dressed Human Modeling, Detection, and Parts Localization*, THE ROBOTICS INSTITUTE, 1–121 (2001) (Carnegie Mellon Univ.) (Ex. 1006).

³ U.S. Patent No. 6,301,370 B1; pub. Oct. 9, 2001 (Ex. 1007).

⁴ U.S. Patent No. 7,200,266 B2; pub. Apr. 3, 2007 (Ex. 1008).

⁵ Leonid Sigal, Sidharth Bhatia, Stefan Roth, Michael J. Black & Michael Isard, *Tracking Loose-limbed People*, IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION 1–8 (2004) (Dept. Comp. Sci., Brown Univ.) (Ex. 1009).

Patent Owner does not dispute the above level of ordinary skill. *See generally*, Prelim. Resp. Patent Owner rather asserts, “[f]or present purposes, [the] Preliminary Response establishes that institution should be denied even under Petitioners’ preferred definition of the level of ordinary skill.” Prelim. Resp. 16.

The level of ordinary skill in the art usually is evidenced by the references themselves. *See Okajima v. Bourdeau*, 261 F.3d 1350, 1355 (Fed. Cir. 2001); *In re GPAC Inc.*, 57 F.3d 1573, 1579 (Fed. Cir. 1995); *In re Oelrich*, 579 F.2d 86, 91 (CCPA 1978). As Petitioner’s description of a person of ordinary skill appears commensurate with the subject matter before us, we apply Petitioner’s definition for purposes of this Decision.

VI. CLAIM CONSTRUCTION

We interpret claim terms using “the same claim construction standard that would be used to construe the claim in a civil action under 35 U.S.C. 282(b).” 37 C.F.R. § 42.100(b) (2019). In this context, claim terms “are generally given their ordinary and customary meaning” as understood by a person of ordinary skill in the art in question at the time of the invention. *Phillips v. AWH Corp.*, 415 F.3d 1303, 1312–13 (Fed. Cir. 2005) (citations omitted) (en banc). “In determining the meaning of the disputed claim limitation, we look principally to the intrinsic evidence of record, examining the claim language itself, the written description, and the prosecution history, if in evidence.” *DePuy Spine, Inc. v. Medtronic Sofamor Danek, Inc.*, 469 F.3d 1005, 1014 (Fed. Cir. 2006) (citing *Phillips*, 415 F.3d at 1312–17). Extrinsic evidence is “less significant than the intrinsic record in determining ‘the legally operative meaning of claim language.’” *Phillips*, 415 F.3d at 1317 (citations omitted).

Petitioner contends that, “[f]or purposes of [the] petition only, Petitioners present no terms for construction.” Pet. 12. Petitioner accepts for purposes of institution, however, Patent Owner’s previously contended (in a motion to dismiss) construction for “spatio-temporal model” as follows: “The ’980 patent’s spatio-temporal model is a graphical model with nodes that correspond to each of the collection of components and to the object.” Pet. 12 (citing Ex. 1043, 6).

Patent Owner does not formally propose (e.g., no separate section directed to) a claim construction. *See generally* Prelim. Resp. Patent Owner does, however, address Petitioner’s above discussion of “spatio-temporal model” as follows:

Petitioners present additional arguments under an alternative interpretation of “spatio-temporal model”—“a graphical model comprising nodes corresponding to each of the collection of components and to the object”—that they neither advocate for nor support with a claim construction analysis or intrinsic evidence support. *See* Pet. 12, 18–21.

Prelim. Resp. 37–38. Aside from contending Petitioner neither “advocate[s] for nor support[s]” the “alternative interpretation,” Patent Owner responds that “Petitioners’ arguments fail under their alternative construction.”

Prelim. Resp. 38.

We construe only those claim terms that require analysis to determine whether to institute *inter partes* review. *See Vivid Techs., Inc. v. Am. Sci. & Eng’g, Inc.*, 200 F.3d 795, 803 (Fed. Cir. 1999) (holding that “only those terms need be construed that are in controversy, and only to the extent necessary to resolve the controversy”). At this stage, we do not need to construe any term.

VII. ANALYSIS

A. *Legal Standards*

“In an [*inter partes* review], the petitioner has the burden from the onset to show with particularity why the patent it challenges is unpatentable.” *Harmonic Inc. v. Avid Tech., Inc.*, 815 F.3d 1356, 1363 (Fed. Cir. 2016) (citing 35 U.S.C. § 312(a)(3) (requiring *inter partes* review petitions to identify “with particularity . . . the evidence that supports the grounds for the challenge to each claim”)); *see also* 37 C.F.R. § 42.104(b) (requiring a petition for *inter partes* review to identify how the challenged claim is to be construed and where each element of the claim is found in the prior art patents or printed publications relied upon).

A claim is unpatentable under 35 U.S.C. § 103(a) if “the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains.” *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 406 (2007). The question of obviousness is resolved on the basis of underlying factual determinations, including: (1) the scope and content of the prior art; (2) any differences between the claimed subject matter and the prior art; (3) the level of skill in the art; and (4) when in evidence, objective evidence of obviousness or nonobviousness, i.e., secondary considerations.⁶ *See Graham v. John Deere Co.*, 383 U.S. 1, 17–18 (1966). An obviousness analysis “need not seek out precise teachings directed to the specific subject matter of the challenged claim, for a court can take account of the inferences

⁶ The parties do not direct us to any objective evidence of non-obviousness at this stage of the proceeding.

and creative steps that a person of ordinary skill in the art would employ.” *KSR*, 550 U.S. at 418.

Additionally, the obviousness inquiry typically requires an analysis of “whether there was an apparent reason to combine the known elements in the fashion claimed by the patent at issue.” *KSR*, 550 U.S. at 418 (citing *In re Kahn*, 441 F.3d 977, 988 (Fed. Cir. 2016) (requiring “articulated reasoning with some rational underpinning to support the legal conclusion of obviousness”)). Furthermore, Petitioner does not satisfy its burden of proving obviousness by employing “mere conclusory statements,” but “must instead articulate specific reasoning, based on evidence of record, to support the legal conclusion of obviousness.” *In re Magnum Oil Tools Int’l, Ltd.*, 829 F.3d 1364, 1380 (Fed. Cir. 2016).

B. Claims 1–16 as Obvious over Zhao

1. Zhao – Exhibit 1006

a. Publication Date of Zhao

Petitioner contends that Zhao was publicly accessible at the Engineering and Science Library at Carnegie Mellon University by September 23, 2002, and was publicly accessible in the ProQuest repository by June 25, 2002. Pet. 5 (citing Ex. 1013, 3; Ex. 1014, 3). Patent Owner contends that the Petition improperly incorporates by reference Exhibits 1013 and 1014, because the Petition does not discuss particular evidence from these exhibits. Prelim. Resp. 28.

Exhibit 1013 is a declaration from Jessica Brenner, who has a Ph.D. and Masters in Library and Information Science. Ex. 1013 ¶ 4. Dr. Brenner testifies that the stamped cataloging data of Zhao is September 16, 2002, and that Zhao would have been publicly accessible within one week, by

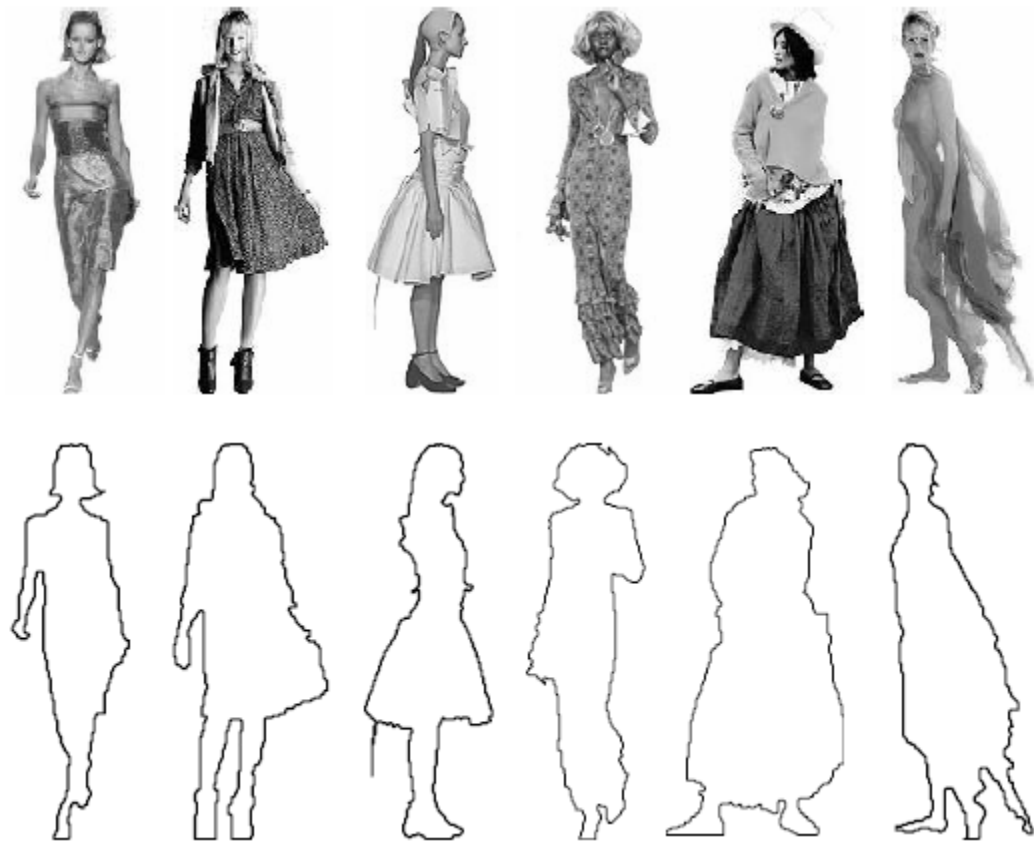
September 23, 2002. *Id.* ¶¶ 8–9. Exhibit 1014 is a declaration from Carl Mageski, who is a Technical Support Analyst employed by ProQuest. Ex. 1014 ¶ 2. Mr. Mageski testifies that he reviewed ProQuest’s records regarding Zhao and determined that the full text of Zhao was available on June 25, 2002. *Id.* ¶ 8. On this record, we are sufficiently persuaded that the testimonies of Dr. Brenner and Mr. Mageski show a reasonable likelihood that Zhao was publicly accessible on September 16, 2002.

b. Teachings of Zhao

Zhao is a “dissertation present[ing] an integrated human shape modeling, detection, and body part localization vision system.” Ex. 1006, 17. “It demonstrates that the system can (1) detect pedestrians in various shapes, sizes, postures, partial occlusion, and clothing from a moving vehicle using stereo cameras; and (2) locate the joints of a person automatically and accurately without employing any markers around the joints.” *Id.*

Zhao initially asks: “Why is it difficult to detect humans[?]” *Id.* at 20. Zhao answers: “The difficulties stem from the number of degrees of freedom in the human body, self-occlusion, appearance variation due to clothing, and the ambiguities in the projection of a 3D human shape onto the image plane.” *Id.*

Figures 1.3 and 1.6 are reproduced below.



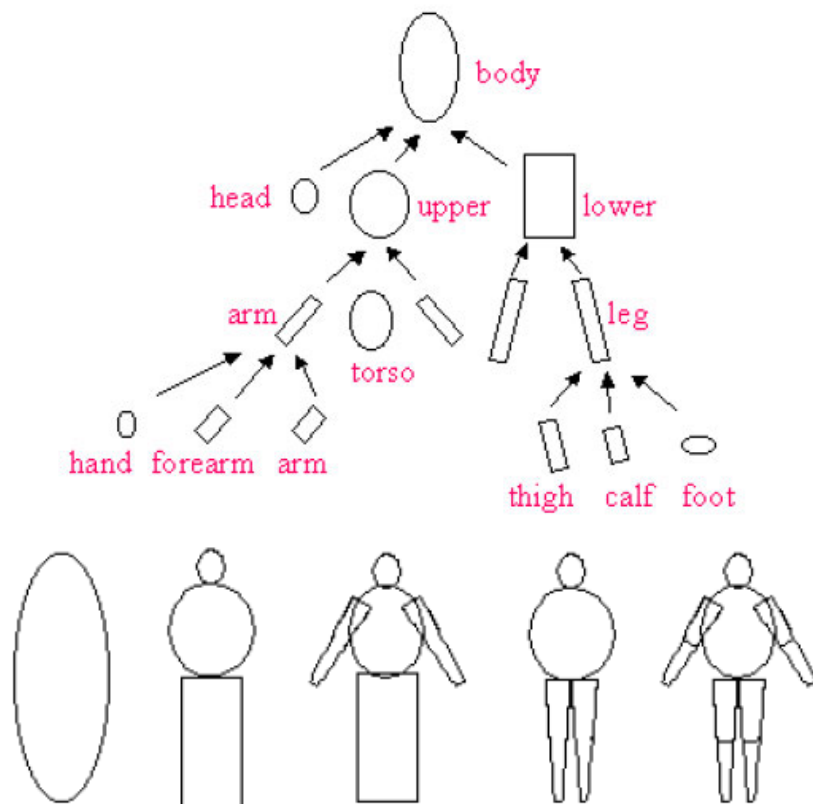
Figures 1.3 and 1.6 above respectively illustrate corresponding “[e]xamples of various appearances due to clothing” (*id.* at 21) and “[e]xamples of various shapes due to clothing (same as Fig. 1.3 but only contour)” (*id.* at 23).

The problem is generalized as follows:

Previous work usually does not model clothes but only the human body. However, clothes may drastically change the shape of a person (see Fig. 1.6). One of the effects of clothes is that they cover some body parts and merge them into a single component. This makes it difficult to distinguish the covered parts. The second effect is that the clothes may generate some spurious body parts along the silhouette that distract from the locations of the real body parts.

Id. at 24.

Figures 1.7 and 1.8 are reproduced below.



Figures 1.7 and 1.8 above respectively illustrate the solution of “merged body parts to model the merging of multiple body parts” and “[a]ssembling [of] human models” (*id.*).

The solution is generalized as follows:

To handle these effects, I first introduce merged body parts to model the merging of multiple body parts as shown in Fig. 1.7. Using the merged body parts, various shape configurations can be built as shown in Fig. 1.8, and the locations of the real body parts can be inferred from the merged parts covering them. The models containing the merged body parts are called dressed human models; they can represent the deformations caused by clothing, segmentation errors, or low image resolution. A dressed human model is dynamically assembled from the model parts in the body part identification procedure. An evaluation function is developed to select the appropriate model parts and

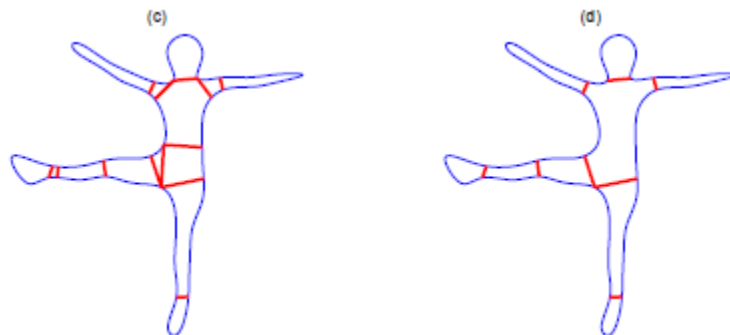
assembling scheme to label the decomposed contour segments. The identification of a part does not only depend on its own shape but also on contextual constraints from other parts. Thus, the labeling is globally optimal and the real body parts can be discriminated from the pseudo parts generated by clothes or other objects held by the person. . . .

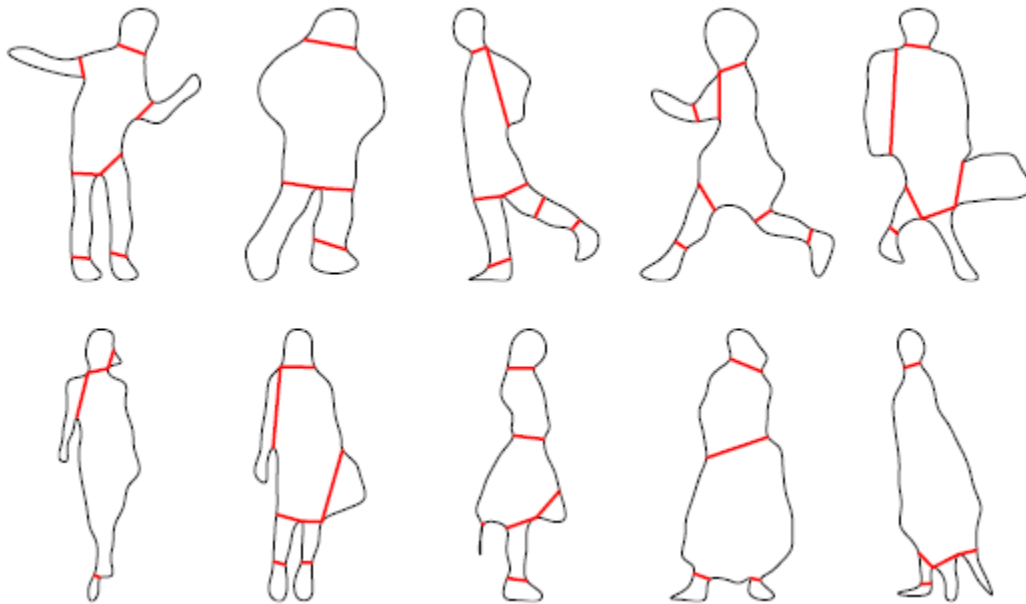
Second, a Bayesian similarity measure is derived from the human model that combines the local shape and global relationship constraints into a single equation to evaluate the degree of resemblance between a contour and the assembled human model. In contrast with previous work, the Bayesian similarity measure enables efficient shape matching and comparison robust to articulation, partial occlusion, and segmentation errors through coarse-to-fine human model assembling.

Third, a coarse-to-fine procedure is developed to locate the joints between body parts accurately: (1) match the extracted ribbons with the model body parts based on the derived similarity measure; (2) infer the locations of the missed body parts from the identified body parts; and (3) adjust the locations of the joints to achieve consistency with the modeled size and spatial relationships between the body parts.

Id. at 24–25.

Figures 2.3(c), 2.3(d), and 2.4 are reproduced below.





Figures 2.3(c), 2.3(d), and 2.4 above respectively illustrate a shape decomposition procedure “computing the cuts of the silhouette” (*id.* at 38), “grouping over-segmented parts” (*id.*), and “[e]xample results of natural shape decomposition” (*id.* at 40).

The solution implements “contextual information” as follows:

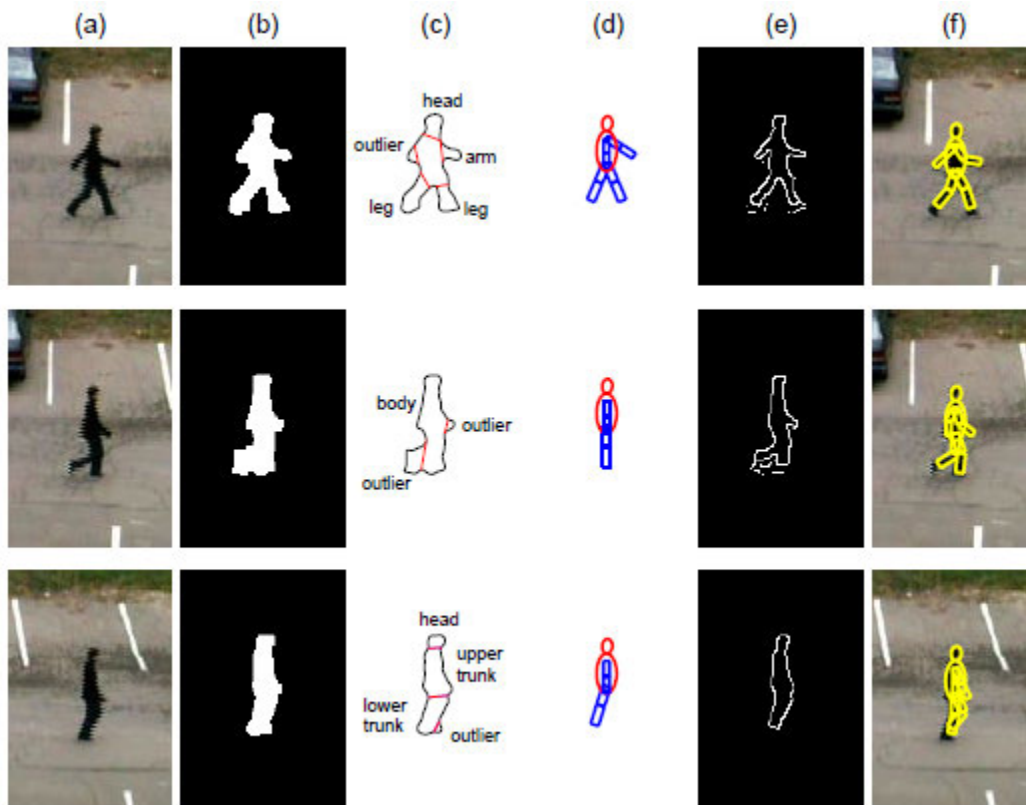
High performance object detection depends on reliable contour extraction, but contour extraction is an under-constrained problem without knowledge about the objects to be detected . .

This thesis proposes a recursive context reasoning (RCR) algorithm to solve the above dilemma. A TRS-invariant probabilistic model is designed to encode the shapes of the body parts and the context information — the size and spatial relationships between body parts. A Bayesian framework is developed to perform human detection and part identification under partial occlusion. A contour updating procedure is introduced to integrate the human model and the identified body parts to predict the shapes and locations of the parts missed by the contour detector; the refined contours are used to reevaluate

the Bayesian similarity measure and determine if the detected contour is a person or not. Therefore, contour extraction, body part localization, and human detection are improved by combining the context constraints from the identified body parts and the human model.

Id. at 69.

Figures 5.4(a) to (f) are reproduced below.



Figures 5.4(a) to (f) above show examples of locating the body parts of a person walking in a parking lot using the RCR algorithm. *Id.* at 91. Figures 5.4(a) to (f) collectively show “[b]ody part localization” (*id.* at 92), including “(a) images[,] (b) foreground object detected from background[,] subtraction[,] (c) identified body parts[,] (d) updated/predicted locations and outlines of the body parts (e) edge images[, and] (f) aligned body parts” (*id.*).

“The body parts of a person are located in a coarse-to-fine manner using the RCR algorithm.” *Id.* at 91. Specifically:

First, the person is segmented from the background (shown in Fig. 5.4(b)). This is done through background subtraction [71]. Second, the segmented region is decomposed into ribbons and these ribbons are matched with the modeled body parts including the extended parts (shown in Fig. 5.4(c)). The joints are initially located in the middle of a cut segment. Third, the locations of the joints are adjusted to achieve consistency with the modeled spatial and size relationships between the body parts. The locations and the sizes of the missed body parts are inferred from the extended body parts and the detected body parts (shown in Fig. 5.4(d)). Fourth, the predicted outlines of the body parts are aligned with the edge features in Fig. 5.4(e). The final results are shown in Fig. 5.4(f). . . . When the arms are overlapped with the torso, it is very hard to locate them, and they may be aligned with the outline of the torso by mistake. Another problem is that the left and right limbs tend to be confused in a side-view. Motion information obtained from previous frames can be used to predict the orientations of the limbs and to solve the ambiguity.

Id.

Figure 5.6, reproduced below, “presents a full cycle of a walking person.” *Id.*

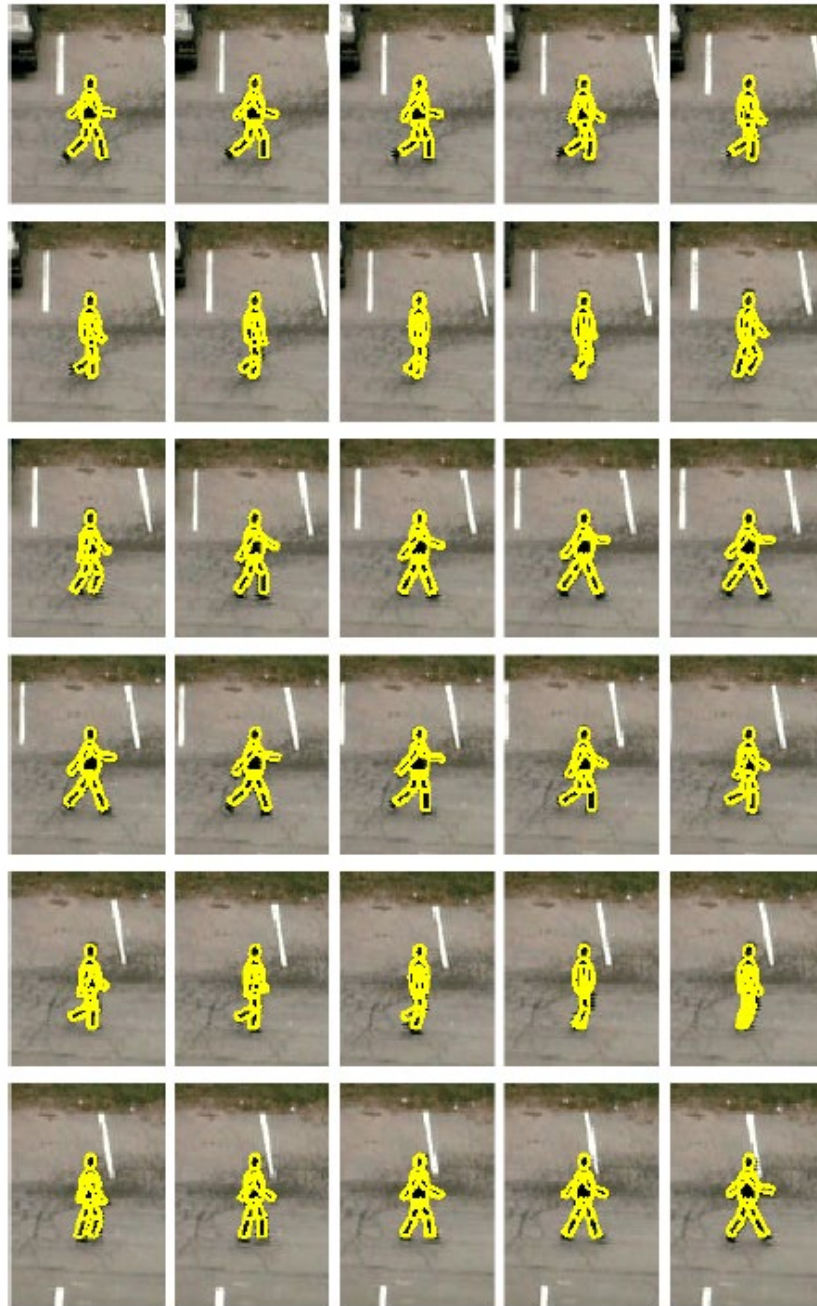


Figure 5.6 above is a series of pictures of a person, outlined in yellow, taken during a full cycle of walking. *Id.* As shown in Figure 5.6, “[i]n the first half cycle, no motion information is used to resolve the ambiguity with limbs’ orientations, while in the second half of the cycle, the prediction from

previous frames (constant angular velocity is assumed) is used to get better results of body part localization.” *Id.* at 91–92.

Figure 5.5, reproduced below, “illustrates how the left knee angle changes with time.” *Id.* at 92.

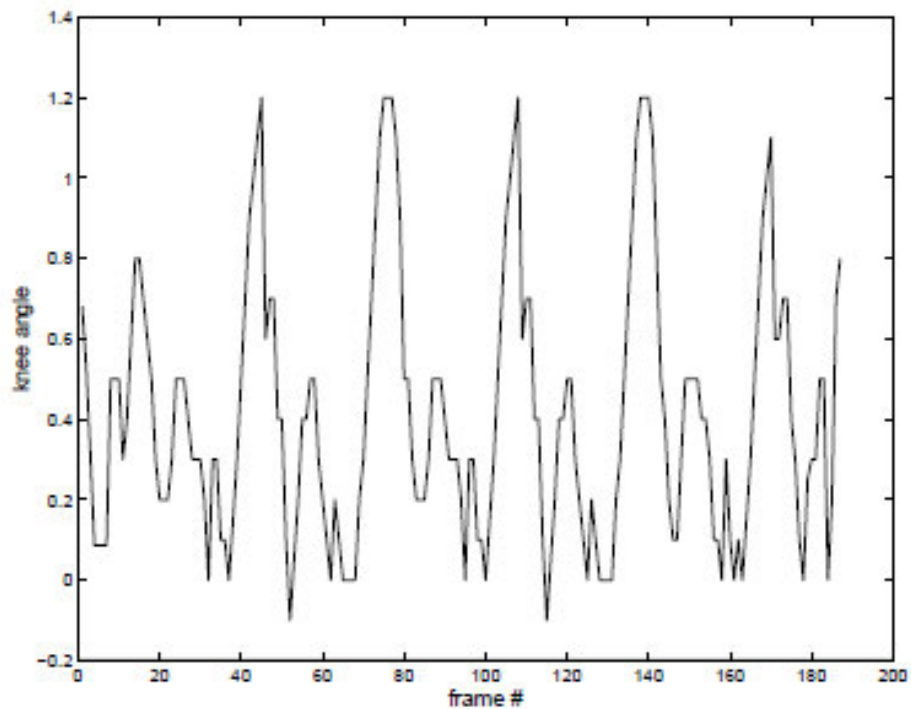


Figure 5.5 above is a graph with knee angle on the y-axis and frame# on the x-axis. As shown in Figure 5.5, “[t]he motion information is not used from frames 1 to 30, but is from frames 31 to 190.” *Id.* “Therefore, the left and right legs are switched sometimes during the first part.” *Id.* “From frames 31 to 190, an obvious pattern of walking cycle can be observed.” *Id.*

1. Reasons for Obviousness Based on Zhao

On a limitation-by-limitation basis, Petitioner’s contentions of “obviousness” are primarily assertions of how a POSITA would have understand Zhao’s disclosures. For example, with regard to claim 1, Petitioner repeatedly asserts how “a POSITA would have understood Zhao”

(Pet. 13, 18, 21, 27) and for each limitation contends that Zhao “disclosed or at least rendered obvious” (or a variation thereof) the limitation (Pet. 13–15, 17–18, 21, 23, 25–27). We will, accordingly, address these “obviousness” contentions when addressing the respective limitations.

2. *Independent Claim 1*

The preamble of claim 1 recites a “computer implemented method for object detection comprising.” Petitioner contends that Zhao teaches the preamble in disclosing a method for object detection in the field of computer vision. Pet. 13 (citing Ex. 1006, 17). Patent Owner does not contend otherwise at this stage of the proceeding. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that Zhao teaches the features recited in the preamble of claim 1.⁷

Claim 1 recites “providing a spatio-temporal model for an object to be detected.” Petitioner contends that Zhao teaches the spatial element of the model in disclosing that its model is designed to encode the size and spatial relationships between body parts. Pet. 14–15 (citing Ex. 1006, 69). Petitioner contends that Zhao teaches the claimed “spatio-temporal model” by incorporating temporal elements into its object model. Pet. 15. In particular, Petitioner contends that Zhao teaches using motion information obtained from previous frames to predict the orientation of limbs. *Id.* at 16 (citing Ex. 1006, 91). Petitioner contends that by incorporating temporal information regarding the change of angle at a subject’s knee with time into

⁷ Because Petitioner has shown that the features in the preamble are satisfied by the prior art, we need not determine whether the preamble is limiting at this time. *See Vivid Techs.*, 200 F.3d at 803.

the model, “an obvious pattern of walking can be observed.” *Id.* at 16–17 (citing Ex. 1006, 92–94, Figs. 5.5, 5.6).

Patent Owner contends that the Petition does not explain how the motion information of Zhao is incorporated into Zhao’s model, and that the motion information of Zhao is a kinematic input into Zhao’s spatial model. Prelim. Resp. 33–34 (citing Ex. 1006, 163). Prelim. Resp. 33. On this record, we disagree with Patent Owner, because the Petition relies on incorporating the motion information shown in Figure 5.5 of Zhao into the object model to predict the orientation of limbs and observe a pattern of walking. Pet. 15–17. Figure 5.5 of Zhao “illustrates how the left knee angle changes with time,” and this “motion information . . . is [used] from frames 31 to 190” such that “an obvious pattern of walking cycle can be observed.” Ex. 1006, 92. We are sufficiently persuaded that this disclosure of Zhao teaches incorporating temporal information into the model to detect the moving object throughout multiple frames. Further, even were we to accept Patent Owner’s contention that Zhao encodes kinematic information, we are sufficiently persuaded that the kinematic information of Zhao teaches motion information that provides a temporal element to the model of body parts that Zhao tracks across images. For example, Bregler⁸, teaches that the “motion of one body segment can be described as the motion of the previous segment in a kinematic chain and an angular motion around a body joint.” Ex. 1027, 1.

Patent Owner contends that Zhao teaches away from the use of motion-based detection techniques because Zhao states that there are several

⁸ C. Bregler & J. Malik, *Tracking people with twists and exponential maps*, CVPR, at 8–15 (1998).

limitations with these schemes. Prelim. Resp. 35. However, Zhao's discussion of other pedestrian systems on page 82 cited by Patent Owner does not disparage Zhao's system of incorporating motion into Zhao's model. Rather, Zhao appears to give an overview of the advantages and disadvantages of other detection systems. Ex. 1006, 82. Further, Patent Owner's contention is inconsistent with Zhao's teaching of incorporating the motion of the left knee angle over time into the model as discussed above. *Id.* at 91–94, Figs. 5.4(a)-(e), Fig. 5.5, Fig. 5.6.

Patent Owner contends that the Petition does not provide a reason to modify Zhao to arrive at the claimed invention. Prelim. Resp. 36. In particular, Patent Owner contends that the Petition's assertion that a person of ordinary skill would have incorporated temporal elements into an object model for an object to be detected by incorporating information measured from multiple image frames does not explain why a person of ordinary skill would have been motivated to do so. *Id.* at 36–37.

However, Petitioner contends that when detecting humans, overlapping body parts can create ambiguity and be confused with one another, and that motion information obtained from previous frames can be used to predict the orientations of the limbs and solve the ambiguity. Pet. 16 (citing Ex. 1006, 91, 99). Petitioner further contends that incorporating motion information of the knee into the object model allows a pattern of walking to be observed. *Id.* at 17 (citing Ex. 1006, 92–93). We are sufficiently persuaded that a person of ordinary skill would have incorporated temporal elements such as motion information into the object model of Zhao to yield the benefits of predicting the orientations of the limbs, solving ambiguity between limbs, and observing a pattern of walking.

Patent Owner contends that Petitioner's arguments under its alternative interpretation of "spatio-temporal model" also fail. Prelim. Resp. 37–41. However, both Patent Owner and Petitioner agree that no construction of any claim term is necessary as we discussed above in the claim construction section, and Petitioner provides a sufficient mapping of Zhao's disclosure to a "spatio-temporal model" under the plain meaning of this term. Therefore, at this stage, we do not need to address Petitioner's alternative construction and Patent Owner's responsive contentions. Further, we note that the claimed "spatio-temporal model" does not seem to be used in any subsequent step recited in claim 1. Rather, the claim performs the steps of providing, measuring, determining, and detecting without using or referring to the spatio-temporal model provided in the first step of the claim.

Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that Zhao teaches this limitation.

Claim 1 recites "providing a video comprising a plurality of images including the object." Petitioner contends that Zhao teaches this limitation in disclosing a pedestrian detection system which has been tested on the videos of urban areas obtained from a stereo system mounted on the top of a minivan. Pet. 21 (citing Ex. 1006, 84–85). Patent Owner does not contend otherwise at this stage of the proceeding. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that Zhao teaches this limitation.

Claim 1 recites "measuring the object as a collection of components in each image." Petitioner contends that Zhao teaches this limitation in

disclosing decomposing contours or silhouettes into their component parts, then performing part identification and human detection. Pet. 23–24 (citing Ex. 1006, 40, Fig. 2.4). Patent Owner does not contend otherwise at this stage of the proceeding. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that Zhao teaches this limitation.

Claim 1 recites “determining a probability that the object is in each image.” Petitioner contends that Zhao teaches this limitation in employing a Bayesian Similarity Measure, which is a part-based similarity measure that evaluates the resemblance between a contour and a model based on the best match between their body parts. Pet. 25–26 (citing Ex. 1006, 54–56, 64–65). Patent Owner does not contend otherwise at this stage of the proceeding. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that Zhao teaches this limitation.

Claim 1 recites “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.” Petitioner contends that Zhao teaches this limitation by determining the probability that the object is in each image, and comparing the probability to a threshold to detect the object. Pet. 26–27 (citing Ex. 1006, 64–65).

Patent Owner contends that the claimed “probabilities for each image” requires a plurality of probabilities, where each image has a probability, such that “comparing the probabilities for each image to a threshold” as claimed requires comparing multiple probabilities from multiple images to a threshold. Prelim. Resp. 42. Patent Owner contends that its construction is consistent with the specification of the ’980 patent, which discloses that a

“w-frame windowed smoothing algorithm” may be used “where w is an odd integer ≥ 1 .” *Id.* at 43 (citing Ex. 1001, 7:6–10). Patent Owner contends that Zhao only compares the probability of one image to a threshold. *Id.* at 44–45.

Even accepting Patent Owner’s construction that the scope of this term encompasses the specification’s disclosure that a “w-frame windowed smoothing algorithm” may be used instead of the entire image sequence, “where w is an odd integer ≥ 1 ,” we are sufficiently persuaded that Zhao teaches this limitation. The specification discloses that “with $w=1$, the algorithm resembles single frame component-based fusion,” such as the single frame threshold taught by Zhao. Ex. 1001, 7:20–22. Further, the specification discloses that Figure 5 shows “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object 505.” *Id.* at 7:53–55. However, Element 505 of Figure 5 discloses “detecting the object in any image upon comparing the probabilities for *the image* to a threshold for detecting the object.” *Id.* at Fig. 5 (emphasis added). On this record, Patent Owner’s contention that the scope of this term as read in light of the specification requires comparing multiple probabilities from multiple images to a threshold is inconsistent with the specification’s disclosure of this term.

Further, Patent Owner’s contention that Zhao only compares one image to a threshold and does not teach comparing multiple probabilities from multiple images to a threshold is inconsistent with the teachings of Zhao. Figures 5.4(a)–(e) of Zhao present examples of a person walking in a parking lot. Ex. 1006, 91. Zhao discloses that motion information of limbs of an object, such as the person walking in the parking lot, obtained from

other frames, can be used to predict the orientations of the limbs in a current frame. *Id.*; *see id.* at 93, Fig. 5.5; Pet. 15–17, 21–22; Ex. 1004 ¶¶ 79–82, 87. Thus, before comparing the current image to a threshold for detecting the object, Zhao has compared other images to a threshold, and uses the resulting information when comparing the current image to a threshold. On this record, we are sufficiently persuaded that Zhao teaches detecting the object of a person walking in a plurality of frames by comparing the probabilities of the current image and other images to a threshold.

Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that Zhao teaches this limitation.

Patent Owner further argues generally that the Petition should be denied because the Petition incorporates by reference Sections VI (giving an overview of the '980 patent) and VIII (giving an overview of the prior art) of Dr. Bajaj's declaration, as well as the declarations of Dr. Benner and Mr. Mageski to circumvent the word count limit. Prelim. Resp. 16–20 (citing Ex. 1004, Sections VI and VIII; Ex. 1013; Ex. 1014). The declarations of Dr. Brenner and Mr. Mageski are three pages each, with nine and ten paragraphs, respectively. We are sufficiently persuaded that Petitioner's citation to the declarations of Dr. Brenner and Mr. Mageski are proper. With respect to Sections VI and VIII of Dr. Bajaj's declaration, the grounds of the Petition do not rely on these sections to support the unpatentability contentions. Rather, the Petition merely identifies the sections of Dr. Bajaj's declaration where Dr. Bajaj summarizes the '980 patent and the prior art.

Based on the evidence and arguments currently of record, for purposes of institution, we are persuaded that Petitioner has demonstrated a reasonable likelihood of prevailing for this challenge to claim 1.

3. *Claims 2–16*

Petitioner challenges claims 2–16 as unpatentable over Zhao. Pet. 27–35. Patent Owner does not separately argue against Petitioner’s showing with respect to these claims. *See* Prelim. Resp. At this stage, because Petitioner demonstrates a reasonable likelihood of prevailing with respect to its challenge of independent claim 1 as unpatentable over Zhao, and because Patent Owner does not separately argue for the patentability of the other challenged claims of this ground, we do not see the need to further address the challenges to claims 2–16 on this ground. Further analysis is best left for trial after full development of the record.

C. *Claims 1–16 As Obvious Over Steffens and Zhao*

1. *Steffens – Exhibit 1007*

Steffens is a U.S. Patent titled “FACE RECOGNITION FROM VIDEO IMAGES.” *Id.* at code (54). Steffens’s “invention relates to vision-based object detection and tracking, and more particularly, to systems for detecting objects in video images, such as human faces, and tracking and identifying the objects in real time.” *Id.* at 1:14–16. The description of the invention’s background identifies the following problem in the art:

Recently developed object and face recognition techniques include the use of elastic bunch graph matching. The bunch graph recognition technique is highly effective for recognizing faces when the image being analyzed is segmented such that the face portion of the image occupies a substantial portion of the image. However, the elastic bunch graph technique may not reliably detect objects in a large scene where

the object of interest occupies only a small fraction of the scene. Moreover, for real-time use of the elastic bunch graph recognition technique, the process of segmenting the image must be computationally efficient or many of the performance advantages of the recognition technique are not obtained.

Accordingly, there exists a significant need for an image processing technique for detecting an object in video images and preparing the video image for further processing by an bunch graph matching process in a computationally efficient manner.

Id. at 1:20–37. The invention’s solution is generalized as follows:

In an embodiment of the invention, the object is detected and a portion of the image frame associated with the object is bounded by a bounding box. The bound portion of the image frame is transformed using a wavelet transformation to generate a transformed image. Nodes associated with distinguishing features of the object defined by wavelet jets of a bunch graph generated from a plurality of representative object images are located on the transformed image. The object is identified based on a similarity between wavelet jets associated with an object image in a gallery of object images and wavelet jets at the nodes on the transformed image.

. . . .

In an alternative embodiment of the invention, the object is in a sequence of images and the step of detecting an object further includes tracking the object between image frames based on a trajectory associated with the object. Also, the step of locating the nodes includes tracking the nodes between image frames and reinitializing a tracked node if the node’s position deviates beyond a predetermined position constraint between image frames.

Id. at 1:50–61, 2:6–13.

Figure 3 is partly reproduced below.



Figure 3 above “shows an acquired image” (*id.* at 4:23–24). “[T]he detected head [is] indicated by a bounding rectangle.” *Id.* at 4:24–25. “The head image is centered, resized, and provided to the landmark finding process.” *Id.* at 4:25–26. “The upper right image window shows the output of the landmark finding module with the facial image marked with nodes on the facial landmarks.” *Id.* at 4:26–29.

Figure 10 is reproduced below.

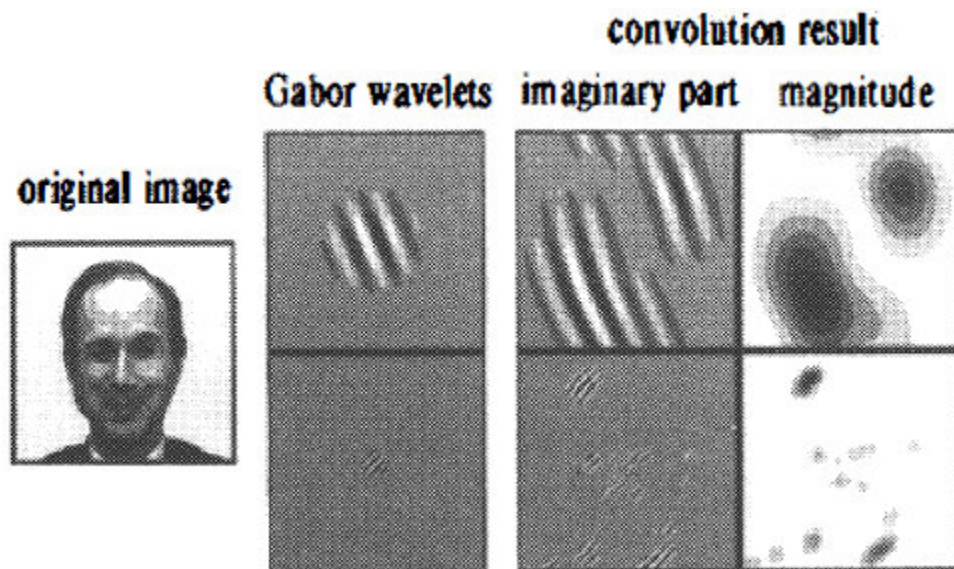


Figure 10 above describes a wavelet transform (*id.* at 8:14–15), presented as “a series of images showing processing of a facial image using

Gabor wavelets” (*id.* at 2:45–46). “A wavelet, centered at image position [x] is used to extract the wavelet component [J] from the image.” *Id.* at 8:35–36. The image is “sampled in a discrete hierarchy of 5 resolution levels . . . and 8 orientations” to “generat[e] 40 complex values for each sampled image point (the real and imaginary components[]).” *Id.* at 8:40–45. “[A]ll wavelet components centered in a single image point are considered as a vector which is called a jet 60. Each jet describes the local features of the area surrounding [x].” *Id.* at 8:48–50. “If sampled with sufficient density, the image may be reconstructed from jets[;] . . . each component of a jet [being] the filter response of a Gabor wavelet extracted at a point (x, y) of the image.” *Id.* at 8:51–55.

Figure 11, reproduced below, “is a series of graphs showing the construction of a jet, image graph, and bunch graph using the wavelet processing technique of FIG. 10.” *Id.* at 2:49–51.

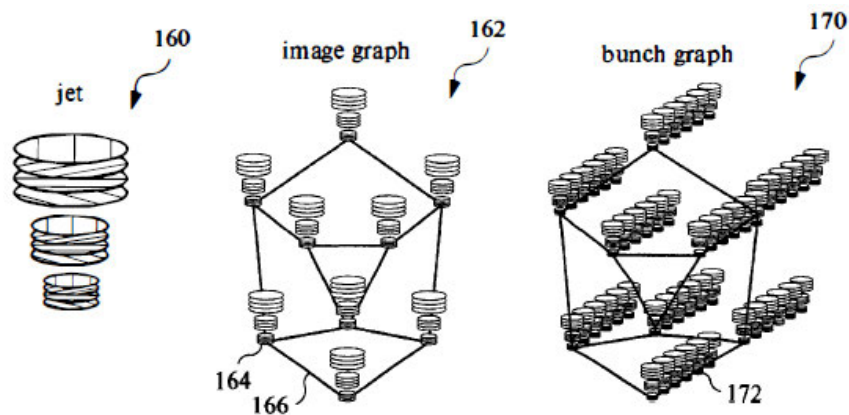


Figure 11 above shows a “labeled image graph 162 . . . used to describe the aspects of an object (in this context, a face).” *Id.* at 8:56–57. “The nodes 164 of the labeled graph refer to points on the object and are labeled by jets 160.” *Id.* at 8:58–59. “Edges 166 of the graph are labeled with distance vectors between the nodes.” *Id.* at 8:59–60. “Nodes and edges

define the graph topology[; g]raphs with equal geometry may be compared.” *Id.* at 8:60–61. “To compute the similarity between two graphs, the sum is taken over similarities of corresponding jets between the graphs.” *Id.* at 8:65–67.

Figure 12, reproduced below, “is a diagram of a[] model graph . . . for processing facial images.” *Id.* at 2:52–53.

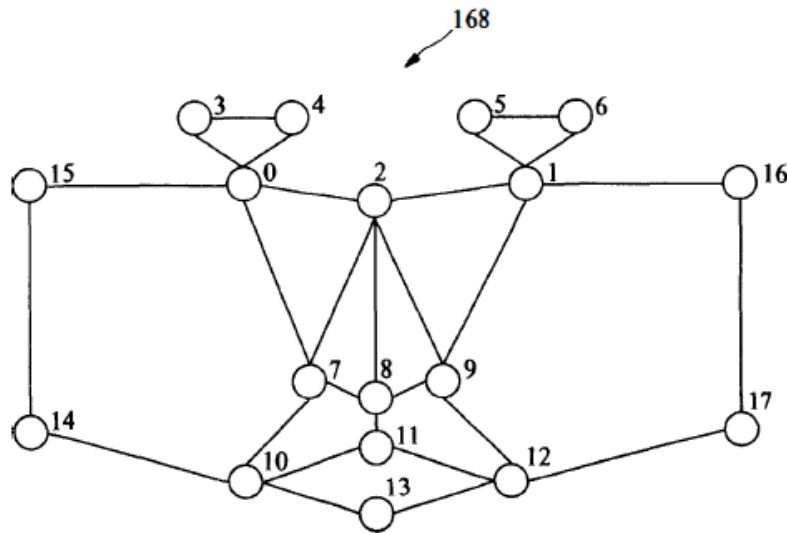


Fig 12 illustrates “[a] model graph 168” having a number of nodes 0–17. Ex. 1001, 9:1–21. A bunch graph 170 (not shown) is made by attaching “a whole bunch of jets 172 . . . to each node” in graph 168 rather than “only a single jet to each node.” *Id.* at 9:22–25. “Each jet is derived from a different facial image.” *Id.* at 9:25–26. “To form a bunch graph, a collection of facial images (the bunch graph gallery) is marked with node locations at defined positions of the head.” *Id.* at 9:26–29. “These defined positions are called landmarks.” *Id.* at 9:29. “When matching a bunch graph to an image, each jet extracted from the image is compared to all jets in the corresponding bunch attached to the bunch graph and the best-matching one is selected.” *Id.* at 9:29–33. “This matching process is

called elastic bunch graph matching.” *Id.* at 9:33–34. “When constructed using a judiciously selected gallery, a bunch graph covers a great variety of faces that may have significant different local properties.” *Id.* at 9:34–36.

2. *Reasons to Combine the Teachings
of Steffens and Zhao*

Petitioner contends that Steffens detects an object using a similarity measure, but does not explicitly state that the similarity measure calculates a probability that an object is in an image. Pet. 44 (citing Ex. 1004 ¶ 144). Petitioner contends that Zhao uses a similarity measure when detecting objects in an image, where the similarity measure includes a Bayesian Similarity Measure that determines the probability that an object is in each image of a video. Pet. 44 (citing Ex. 1004 ¶ 145). Petitioner contends that a person of ordinary skill would have replaced the similarity measure of Steffens with the Bayesian Similarity Measure of Zhao for the benefit of combining local and global relationship constraints into a single equation to evaluate the degree of resemblance as taught by Zhao. Pet. 36 (citing Ex. 1006, 25). Zhao discloses that the Bayesian Similarity Measure provides the benefit of “efficient shape matching and comparison robust to articulation, partial occlusion, and segmentation errors through coarse-to-fine human model assembling.” Ex. 1006, 25.

Patent Owner argues that Petitioner’s contention that a person of ordinary skill would have compared the probabilities to a threshold merely purports to demonstrate that the use of a threshold was known, but does not identify a reason to combine the elements in the manner claimed. Prelim. Resp. 54–55. On this record, we are persuaded that Zhao’s disclosure of comparing probabilities to a threshold for the benefit of deciding whether a

human is detected sufficiently shows a reason to compare probabilities to a threshold. Ex. 1006, 64.

Patent Owner argues that Petitioner's contention that a person of ordinary skill would have looked for a probabilistic framework such as Zhao's to combine with the teaching of Steffens does not explain how or why Steffens and Zhao would have been combined to arrive at the claimed invention. Prelim. Resp. 55. In particular, Patent Owner contends that the combination does not teach "detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object" as claimed. *Id.* Patent Owner makes a similar contention regarding the reasonable expectation of success. *Id.* at 56. We disagree with Patent Owner on this record. We are sufficiently persuaded that a person of ordinary skill, in order to identify an object based on similarity between an image and a corresponding model as taught by Steffens, would have used the Bayesian Similarity Measure of Zhao for the benefit of combining local and global relationship constraints into a single equation to evaluate the degree of resemblance as taught by Zhao. Ex. 1007, 5:12–13; Ex. 1006, 25, 54; Ex. 1004 ¶¶ 124–125. We are further persuaded that the combination teaches "detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object" as claimed for the reasons given in our analysis of this limitation below.

For purposes of this Decision, we are persuaded that Petitioner cites sufficient evidence to support its contention that a person of ordinary skill would have had reason to combine the teachings of Steffens and Zhao.

3. *Independent Claim 1*

The preamble of claim 1 recites a “computer implemented method for object detection comprising.” Petitioner contends that Steffens teaches the preamble in disclosing systems for detecting objects in video images. Pet. 37 (citing Ex. 1007, 1:14–17). Patent Owner does not contend otherwise at this stage of the proceeding. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that Steffens teaches the features recited in the preamble of claim 1.⁹

Claim 1 recites “providing a spatio-temporal model for an object to be detected.” Petitioner contends that Steffens teaches the spatial element of the spatio-temporal model in disclosing that nodes of a graph refer to points on an object, and edges of the graph are labeled with distance vectors between nodes. Pet. 38 (citing Ex. 1007, 8:58–60). Petitioner contends that Steffens teaches the temporal element in tracking the object between image frames based on a trajectory associated with the object. Pet. 39 (citing Ex. 1007, 2:6–13).

Patent Owner contends that Steffens does not model an object over any length of time, because Steffens distinguishes tracking from detecting landmarks. Prelim. Resp. 48–50 (citing Ex. 1007, 3:54–4:19, 7:51–53, 10:46–61). According to Patent Owner, the separate tracking process of Steffens may make use of position information generated from the landmark detection method, but the landmark tracking process does not provide a

⁹ Because Petitioner has shown that the features in the preamble are satisfied by the prior art, we need not determine whether the preamble is limiting at this time. *See Vivid Techs.*, 200 F.3d at 803.

model “for an object to be detected” as claimed. *Id.* We disagree with Patent Owner on this record. Steffens teaches a head detection and tracking process 14 (Ex. 1007, 3:54–4:19), and a preselector 16 that processes a series of face candidates that belong to the same trajectory as determined by the head tracking process 14 (*id.* at 7:9–11). On this record, we are sufficiently persuaded that detecting and tracking an object between frames based on a trajectory of the object as taught by Steffens teaches the “spatio-temporal model” as claimed.

Further, although claim 1 recites “providing a spatio-temporal model for an object to be detected,” the subsequent steps of the claim do not appear to use the spatio-temporal model to detect the object. To the extent Patent Owner is arguing that Steffens does not use the model in its alleged performance of the subsequent steps, that argument does not appear to be commensurate with the scope of the claim.

Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that Steffens teaches this limitation.

Claim 1 recites “providing a video comprising a plurality of images including the object.” Petitioner contends that Steffens teaches this limitation in disclosing detecting objects in video images. Pet. 42–43 (citing 1007, 1:14–16, 3:48–53). Patent Owner does not contend otherwise at this stage of the proceeding. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that Steffens teaches this limitation.

Claim 1 recites “measuring the object as a collection of components in each image.” Petitioner contends that Steffens teaches this limitation by

using nodes with an associated state vector (e.g., distance vectors between nodes) to measure the object as a collection of components. Pet. 43–44 (citing Ex. 1004 ¶ 142). Patent Owner does not contend otherwise at this stage of the proceeding. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that Steffens teaches this limitation.

Claim 1 recites “determining a probability that the object is in each image.” Petitioner contends that the combination of Steffens and Zhao teaches this limitation. Pet. 44–45. Petitioner contends that Steffens detects an object using a similarity measure, but does not explicitly state that the similarity measure calculates a probability that an object is in an image. Pet. 44 (citing Ex. 1004 ¶ 144). Petitioner contends that Zhao uses a similarity measure when detecting objects in an image, where the similarity measure includes a Bayesian Similarity Measure that determines the probability that an object is in each image of a video. Pet. 44 (citing Ex. 1004 ¶ 145). Petitioner contends that a person of ordinary skill would have replaced the similarity measure of Steffens with the Bayesian Similarity Measure of Zhao for the benefit of combining local and global relationship constraints into a single equation to evaluate the degree of resemblance as taught by Zhao. Pet. 36 (citing Ex. 1006, 25). Zhao discloses that the Bayesian Similarity Measure provides the benefit of “efficient shape matching and comparison robust to articulation, partial occlusion, and segmentation errors through coarse-to-fine human model assembling.” Ex. 1006, 25.

Patent Owner does not contend otherwise at this stage of the proceeding. Based on the evidence and arguments currently of record, for

purposes of institution, Petitioner has sufficiently demonstrated that the combination of Zhao and Steffens teaches this limitation.

Claim 1 recites “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.”

Petitioner contends that the combination of Zhao and Steffens teaches this limitation for the reasons given in Petitioner’s motivation to combine the teachings of Zhao and Steffens. Pet. 45 (citing Pet. 35–37). In particular, Petitioner contends that Zhao teaches comparing the probabilities for each image to a threshold for detecting an object, as discussed in Petitioner’s analysis of this limitation in ground 1. *Id.* (citing Pet. 26–27).

Patent Owner contends that the combination of Zhao and Steffens does not teach this limitation for the reasons given in Patent Owner’s analysis of this limitation in ground 1. Prelim. Resp. 52–53. We disagree with Patent Owner for the reasons given in our analysis of claim 1.

Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that the combination of Zhao and Steffens teaches this limitation.

Based on the evidence and arguments currently of record, for purposes of institution, we are persuaded that Petitioner has demonstrated a reasonable likelihood of prevailing for this challenge to claim 1.

4. *Claims 2–16*

Petitioner challenges claims 2–16 as unpatentable over Steffens and Zhao. Pet. 46–52. Patent Owner does not separately argue against Petitioner’s showing with respect to these claims. *See* Prelim. Resp. At this stage, because Patent Owner does not separately argue with respect to the other challenged claims of this ground, we do not see the need to further

address the challenges to claims 2–16 on this ground. Further analysis is best left for trial after full development of the record.

D. Claims 1–16 As Obvious Over Ozer and Zhao

1. Ozer – Exhibit 1008

Ozer is a U.S. Patent titled “METHOD AND APPARATUS FOR AUTOMATED VIDEO ACTIVITY ANALYSIS.” *Id.* at code (54). Ozer relates “to detect[ing] the presence of articulated objects, e.g. human body, and rigid objects and to identify[ing] their activities in compressed and uncompressed domains and in real-time.” *Id.* at 1:18–21. The description of the invention’s background includes a brief comparison of “[e]arly activity recognition systems [that] used beacons carried by the subjects” and, as Ozer’s invention’s implements, a “system that uses video . . . to recognize activities that can be used to command the operation of the environment.” *Id.* at 1:53–57.

Noting a problem that “most of the [early] activity recognition systems are suitable for a specific application type,” the invention’s solution is generalized as follows:

The invention described herein can detect a wide range of activities for different applications. For this reason, the scheme detects different object parts and their movement in order to combine them at a later stage that connects to high-level semantics. Each object part has its own freedom of motion and the activity recognition for each part is achieved by using several [Hidden Markov Models (HMMs)] in parallel.

Id. at 2:67–3:7. The solution is then specified as follows:

The system of the present invention can detect non-rigid (e.g. human body) and rigid object parts and recognize their activities in compressed and uncompressed domains. To achieve this, a method with two levels, namely low and high levels, is

used. The low-level part performs object detection and extracts parameters for the abstract graph representation of the image being processed. The high level part uses dynamic programming to determine the activities of the object parts, and uses a distance classifier to detect specific activities.

Low-level part performs object detection and extracts parameters for the abstract graph representation of the frame being processed in real time. Local consistency based on low level features and geometrical characteristics of the object regions is used to group object parts. Furthermore, higher order shape metrics is needed for the presentation of the complex objects. The object is decomposed for its presentation as a combination of component shapes. The result will be unaffected by a partial occlusion of the object.

The system is capable of managing the segmentation process by using object-based knowledge in order to group the regions according to a global consistency and introducing a new model-based segmentation algorithm by using a feedback from relational representation of the object. The major advantages of the model-based segmentation can be summarized as improving the object extraction by reducing the dependence on the low-level segmentation process and combining the boundary and region properties. Furthermore, the features used for segmentation are also attributes for object detection in relational graph representation.

Id. at 3:55–4:17. The solution further includes “graph matching” as follows:

Object detection is achieved by matching the relational graphs of objects with the reference model. . . .

After the detection of the object parts, the system is ready to recognize the activities of each object part and the overall activity of the object.

For example, if the object of interest is a human body, the system will first detect different object parts, e.g. hands, head, arms, legs, torso and compare these part attributes with the human model attributes via graph matching. If the object of interest is a rigid object the system will detect object parts and

compare the attributes of these parts with the object model via graph matching.

Id. at 4:24–44.

Figure 4 is reproduced below.

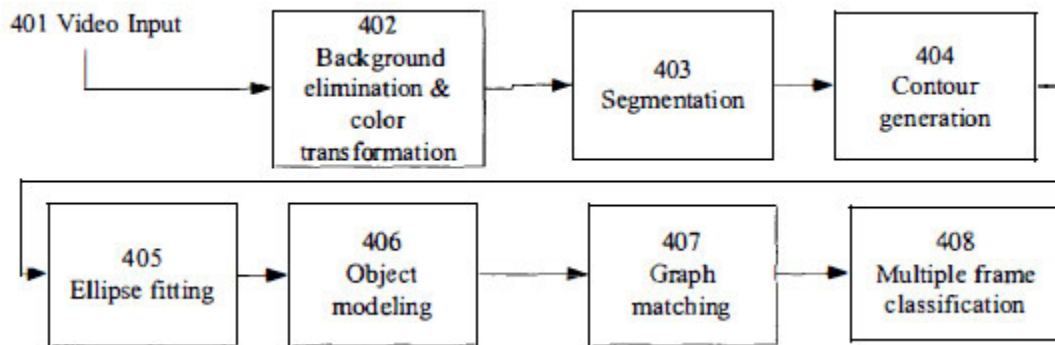


Figure 4 above is a block diagram of Ozer’s object detection and activity recognition system. *Id.* at 5:36–38. As shown in Figure 4, in “[b]ackground elimination and color transformation[, t]he first step (402) is the transformation of pixels into another color space regarding to the application.” *Id.* at 6:44–46. “Background elimination is performed by using these transformed pixel values for the current and background images.” *Id.* at 6:46–48.

In the next step (403) performing “[s]egmentation[,] . . . the foreground regions are extracted and the object of interest is segmented hierarchically into its smaller unique parts based on the combination of color components and statistical shape features after background elimination.” *Id.* at 6:52–56. “The meaningful adjacent segments are combined and used as the input of the following algorithm steps.” *Id.* at 6:56–58.

In “[c]ontour following[, c]ontour points of the segmented regions are extracted and stored [by step] (404).” *Id.* at 6:59–60. In “[e]llipse fitting[,] . . . step (405) fits ellipses to the contours.” “Even when object of interest is

not occluded by another object, due to the possible positions of non-rigid parts an object part can be occluded in different ways.” *Id.* at 6:64–67. “In this case, 2D approximation of parts by fitting ellipses with shape preserving deformations provides more satisfactory results.” *Id.* at 6:67–7:2. In “[o]bject modeling by invariant shape attributes[, f]or object detection, it is necessary to select part attributes which are invariant to two-dimensional transformations and are maximally discriminating between objects (406).” *Id.* at 7:4–7.

In “[g]raph matching[,] . . . step (407) . . . compare[s] the object model with a set of stored models.” *Id.* at 7:8–9. Specifically:

Each extracted region modeled with ellipses corresponds to a node in the graphical representation of the object of interest. Each object part and meaningful combinations represent a class w where the combination of binary and unary features are represented by a feature vector X and computed off-line. The combination of segments is controlled by the reference model and by the rule generator. If the graph-matching algorithm cannot find a meaningful correspondence of the combined segments in the reference model, the combination will be rejected and a new combination will be generated. For the purpose of determining the class of these feature vectors a piecewise quadratic Bayesian classifier with discriminant function $g(X)$ is used. The generality of the reference model attributes allows the detection of different kind of models for the same object type while the conditional rule generation decreases the rate of false alarms. The computations needed for each node matching are then a function of the feature size and the previously matched nodes of the branch under consideration. The marked regions are tracked by using ellipse parameters for the consecutive frames and graph-matching algorithm is applied for new objects appearing in the other regions.

Id. at 7:9–31.

“Output of the graph-matching algorithm is the classified object parts.” *Id.* at 7:34–35. “The movements of the object parts are described as a spatio-temporal sequence of feature vectors that consist of the direction of the object part movement.” *Id.* at 7:35–37. Lastly, in “Classifying Over Multiple Frames,” step (408) “checks direction of the movements of the object parts for a number of frames and calculates the probabilities of the activities with the known activities by using [HMMs] and chooses the pattern with the highest probability as the recognized activity in these frames.” *Id.* at 7:32, 38–42.

2. *Reasons to Combine the Teachings
of Ozer and Zhao*

Petitioner contends that Ozer teaches matching relational graphs with a reference model, but does not specify which matching technique to use to identify an object, only that the matching technique looks for meaningful correspondence between object parts and the reference model. Pet. 54 (citing Ex. 1008, 4:24–25, 7:8–31). Petitioner contends that a person of ordinary skill would have used the Bayesian Similarity Measure of Zhao to find meaningful correspondence in the method of Ozer for the benefit of combining local shape and global relationship constraints into a single equation to evaluate the degree of resemblance as taught by Zhao. *Id.* (citing Ex. 1006, 25).

Patent Owner contends that, even if combined, the combination would not teach “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object” as claimed. Prelim. Resp. 62. Patent Owner makes a similar argument regarding the expectation of success. *Id.* at 62–63. On this record, we

disagree with Patent Owner. We are sufficiently persuaded that a person of ordinary skill, in order to match relational graphs with a reference model as taught by Ozer, would have used the Bayesian Similarity Measure of Zhao for the benefit of combining local shape and global relationship constraints into a single equation to evaluate the degree of resemblance as taught by Zhao. We are further persuaded that the combination teaches “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object” as claimed for the reasons given in our analysis of this limitation below.

Accordingly, for purposes of this Decision, we are persuaded that Petitioner cites sufficient evidence to support its contention that a person of ordinary skill would have had reason to combine the teachings of Ozer and Zhao.

3. *Independent Claim 1*

The preamble of claim 1 recites a “computer implemented method for object detection comprising.” Petitioner contends that Ozer teaches the preamble in disclosing a method for object detection in video sequences. Pet. 55. Patent Owner does not contend otherwise at this stage of the proceeding. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that Ozer teaches the features recited in the preamble of claim 1.¹⁰

Claim 1 recites “providing a spatio-temporal model for an object to be detected.” Petitioner contends that Ozer teaches this limitation in disclosing

¹⁰ Because Petitioner has shown that the features in the preamble are satisfied by the prior art, we need not determine whether the preamble is limiting at this time. *See Vivid Techs.*, 200 F.3d at 803.

a spatio-temporal model for an object to be detected. Pet. 56–57 (citing Ex. 1008, 7:8–11, 7:33–45, 4:45–47).

Patent Owner contends that Ozer does not teach a spatio-temporal model because Ozer builds a separate model on each frame, and Ozer distinguishes tracking from detecting object parts. Prelim. Resp. 56–60. However, Ozer discloses using Hidden Markov Models, which classify activity patterns of body parts in space-time. Ex. 1008, 4:44–47. Ozer further describes the movements of the object parts “as a spatio-temporal sequence of feature vectors.” *Id.* at 7:35–37. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that Ozer teaches this limitation.

Claim 1 recites “providing a video comprising a plurality of images including the object.” Petitioner contends that Ozer teaches this limitation in disclosing a methodology able to decide on the presence of an object in video sequences. Pet. 58 (citing Ex. 1008, Abstr.). Patent Owner does not contend otherwise at this stage of the proceeding. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that Ozer teaches this limitation.

Claim 1 recites “measuring the object as a collection of components in each image.” Petitioner contends that Ozer teaches this limitation in disclosing a graph matching algorithm that looks for meaningful correspondence where each object part and meaningful combinations represent a class. Pet. 59–60 (citing Ex. 1008, 7:9–19, 7:25–28). Patent Owner does not contend otherwise at this stage of the proceeding. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that Steffens teaches this limitation.

Claim 1 recites “determining a probability that the object is in each image.” Petitioner contends that the combination of Ozer and Zhao teaches this limitation as discussed above in the motivation to combine the teachings of Ozer and Zhao. Pet. 60 (citing Pet. 53–55). Petitioner contends that Ozer teaches matching relational graphs with a reference model, but does not specify which matching technique to use to identify an object, only that the matching technique looks for meaningful correspondence between object parts and the reference model. Pet. 54 (citing Ex. 1008, 4:24–25, 7:8–31). Petitioner contends that a person of ordinary skill would have used the Bayesian Similarity Measure of Zhao to find meaningful correspondence in the method of Ozer for the benefit of combining local shape and global relationship constraints into a single equation to evaluate the degree of resemblance as taught by Zhao. *Id.* (citing Ex. 1006, 25). Patent Owner does not contend otherwise at this stage of the proceeding. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that the combination of Ozer and Zhao teaches this limitation.

Claim 1 recites “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.” Petitioner contends that Zhao teaches this limitation for the reasons given by Petitioner in its analysis of this limitation in ground 1. Pet. 60–61 (citing Pet. 26–27).

Patent Owner contends that Zhao does not teach this limitation for the reasons given in Patent Owner’s analysis of ground 1. Prelim. Resp. 61. We disagree with Patent Owner for the reasons given in our analysis of ground 1. Based on the evidence and arguments currently of record, for

purposes of institution, Petitioner has sufficiently demonstrated that Zhao teaches this limitation.

Based on the evidence and arguments currently of record, for purposes of institution, we are persuaded that Petitioner has demonstrated a reasonable likelihood of prevailing for this challenge to claim 1.

4. *Claims 2–16*

Petitioner challenges claims 2–16 as unpatentable over Ozer and Zhao. Pet. 61–65. Patent Owner does not separately argue against Petitioner’s showing with respect to these claims. *See* Prelim. Resp. At this stage, because Patent Owner does not separately argue with respect to the other challenged claims of this ground, we do not see the need to further address the challenges to claims 2–16 on this ground. Further analysis is best left for trial after full development of the record.

E. *Claims 1–16 As Obvious Over TLP*

1. *TLP – Exhibit 1009*

a. Publication Date of TLP

The parties agree that the effective filing date of the ’980 patent is May 27, 2004. Pet. 5; Prelim. Resp. 16. Petitioner contends that TLP was made publicly available for the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (“CVPR”) on April 9, 2004, and that the prosecution history of the ’980 patent shows that TLP was published in 2004. Pet. 6 (citing Ex. 1004 ¶¶ 65–66; Ex. 1010, 1; Ex. 1002, 37–48; Ex. 1001, 1; Ex. 1003, 12). Dr. Bajaj testifies that the deadline for full paper submission for the 2004 CVPR was November 19, 2003, after which, TLP was widely disseminated and made publicly available to academics for peer review. Ex. 1004 ¶ 66. Dr. Bajaj testifies that after being accepted for the

CVPR, the deadline for manuscript submission of TLP was April 9, 2004.
Id.

Patent Owner contends that submitting a paper to a peer review committee in connection with submitting the paper to a conference suggests an absence of actual public accessibility. Prelim. Resp. 25–27 (citing *SRI Int’l, Inc. v. Internet Sec. Sys.*, 511 F.3d 1186, 1197 (Fed. Cir. 2008)). Patent Owner further contends that the publication date of 2004 found in the prosecution history does not show that TLP was publicly available before May 27, 2004.

On this record, we agree with Patent Owner, that the Petition does not show a reasonable likelihood that TLP was published before May 27, 2004. However, as discussed above in our analyses of grounds 1–3, Petitioner has demonstrated a reasonable likelihood of prevailing with respect to at least one claim of the ’980 patent challenged in the Petition. Therefore, we institute on all grounds and all claims challenged in the Petition. *See PGS Geophysical AS v. Iancu*, 891 F.3d 1354, 1360 (Fed. Cir. 2018).

b. Teachings of TLP

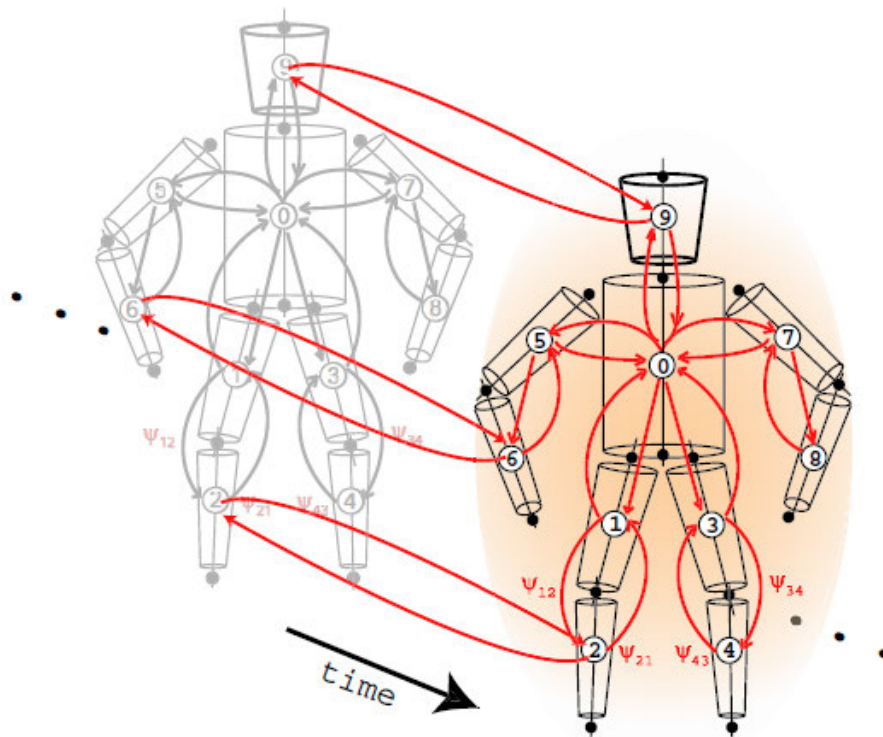
TLP is a “Brown University” and “Microsoft Research” publication titled “Tracking Loose-limbed People.” Ex. 1009, 1. TLP “pose[s] the problem of 3D human tracking as one of inference in a graphical model” and a solution of “learn[ing] probabilistic models for the temporal evolution of each limb (forward and backward in time).” *Id.* The background and improvement are generalized as follows:

In previous work[,] we presented the general loose-limbed body model and belief propagation algorithm but only addressed the problem of human pose estimation at a single time instant; here we extend the inference method over time to perform visual

tracking. [The previous work’s] potential functions [for] linking limbs were constructed manually[,] whereas here they are learned from training data and are also extended in time.

Id. at 2.

Figure 1, reproduced below, illustrates a “[g]raphical model for a person” whereby “[n]odes represent limbs and arrows represent conditional dependencies between limbs.” *Id.* at 1.



With reference to Figure 1 above, the “tracking problem” and “graphical model” are explained as follows:

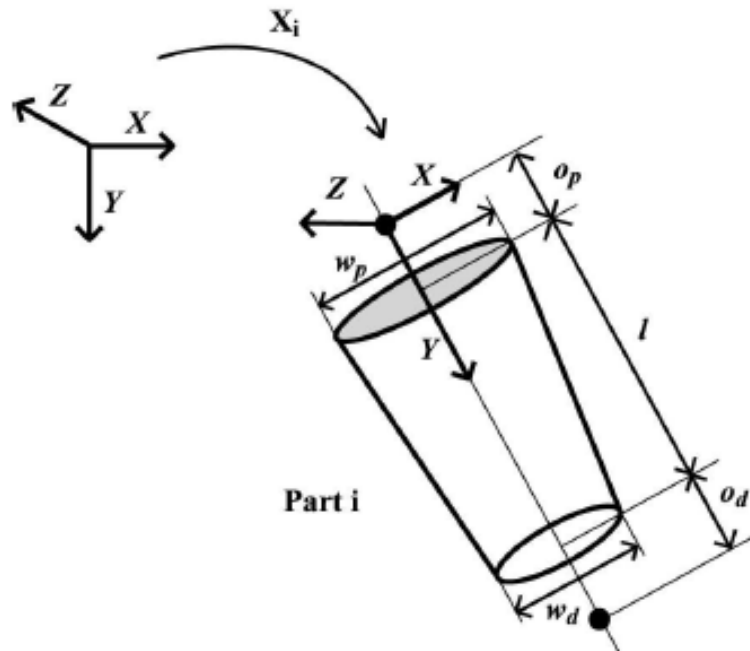
The tracking problem is formulated as one of inference in a graphical model and belief propagation is used to estimate the pose of the body at each time-step. Each node in the graphical model represents the 3D position and orientation of a limb (Figure 1). Directed edges between nodes represent statistical dependencies and these constraints between limbs are used to form messages that are sent to neighboring nodes in space and time. Additionally, each node has an associated likelihood

defined over a rich set of image cues using a learned Gibbs model[.]

Id. “[O]nly a few temporal dependencies have been shown: in [the] model each part i at time t is connected by an edge to part i at times $t - 1$ and $t + 1$. ”

Id.

Figure 2, reproduced below, illustrates a “[p]arameterization of part i . ” *Id.* at 3.



With reference to Figure 2 above, fixed parameters, estimated parameters, and potential functions are explained as follows:

In order to describe the body parts in a graphical model, we assume the variables in a node are conditionally independent of those in non-neighboring nodes given the values of the node’s neighbors. Each part/limb is modeled by a tapered cylinder having 5 fixed and 6 estimated parameters.

The fixed parameters . . . correspond respectively to the part length, width at the proximal and distal ends and the offset

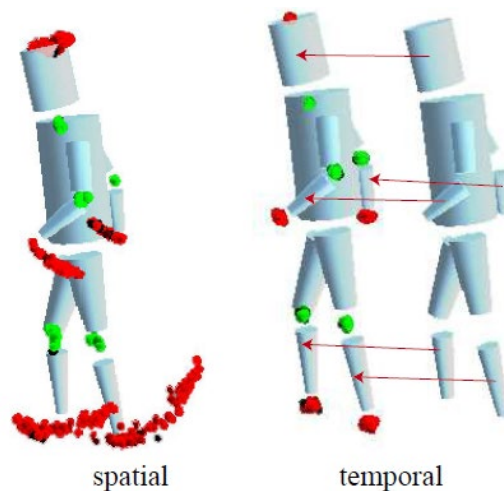
of the proximal and distal joints along the axis of the limb as shown in Figure 2.

The estimated parameters . . . represent the configuration of the part i in a global coordinate frame where [the parameters] . . . are the 3D position of the proximal joint and the angular orientation of the part respectively.

Each directed edge between parts i and j [(not shown)] has an associated potential function . . . that encodes the compatibility between pairs of part configurations and intuitively can be thought of as the probability of configuration X_j of part j conditioned on the X_i of part i .

Id. at 2–3 (paragraph breaks added).

Figure 3, reproduced below, illustrates “Learned Spatial and Temporal Potentials.” *Id.* at 3.



With reference to Figure 3 above, “spatial potentials show the distribution of limb positions and orientations conditioned on the neighboring limb,” and are explained as follows:

Spatial and temporal potentials are illustrated by sampling from them. The potentials for the lower limbs (arms and legs) and head are shown. The spatial potentials show the distribution of limb positions and orientations conditioned on the neighboring limb. Green spheres indicate the joint position of a sample while

the red spheres indicate the distal end of the limb for each sample. The spread of these samples illustrates the learned distribution encoded by the potentials. The temporal potentials shown are for the forward time direction.

Id. at 3.

1. *Reasons for Obviousness in view of TLP*

On a limitation-basis, Petitioner’s contentions of “obviousness” are mostly, rather, assertions of how a POSITA would have understood TLP’s disclosures. For example, with regard to claim 1, Petitioner repeatedly asserts, for each limitation, how “a POSITA would have understood” a disclosure of TLP (Pet. 66, 68, 71) meets the limitation and states that TLP “disclosed or at least rendered obvious” (or a near *verbatim* variation thereof) the limitation and/or an included feature (Pet. 64, 720). We will, accordingly, address these “obviousness” contentions when addressing the respective limitations.

2. *Independent Claim 1*

The preamble of claim 1 recites a “computer implemented method for object detection comprising.” Petitioner contends that TLP teaches the preamble in disclosing a probabilistic method for fully automatic 3D human detection and tracking. Pet. 65 (citing Ex. 1009, 6). Patent Owner does not contend otherwise at this stage of the proceeding. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that TLP teaches the features recited in the preamble of claim 1.¹¹

¹¹ Because Petitioner has shown that the features in the preamble are satisfied by the prior art, we need not determine whether the preamble is limiting at this time. *See Vivid Techs.*, 200 F.3d at 803.

Claim 1 recites “providing a spatio-temporal model for an object to be detected.” Petitioner contends that TLP teaches this limitation in disclosing a model for a human to be detected, where each node corresponds to a body part and edges correspond to position and angle relationships between adjacent body parts in space and time. Pet. 66–67 (citing Ex. 1009, 1–2, Fig. 1).

Patent Owner contends that each node of TLP corresponds to a body part, but Petitioner does not identify a node that corresponds to the full body, or “object” as claimed. Prelim. Resp. 63–64. However, claim 1 does not require nodes for body parts and a separate node for the body. Rather, claim 1 recites “a spatio-temporal model for an object.” TLP discloses “our model of the body is a collection of loosely-connected limbs.” Ex. 1009, 1. We are sufficiently persuaded that the model of the body disclosed by TLP teaches the “spatio-temporal model for an object” as claimed. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that TLP teaches this limitation.

Claim 1 recites “providing a video comprising a plurality of images including the object.” Petitioner contends that TLP teaches this limitation in disclosing a method for object detection and tracking for video. Pet. 67–68 (citing Ex. 1009, 1). Patent Owner does not contend otherwise at this stage of the proceeding. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that TLP teaches this limitation.

Claim 1 recites “measuring the object as a collection of components in each image.” Petitioner contends that TLP teaches this limitation in showing an object being measured as a collection of components. Pet. 68–

70 (citing Ex. 1009, Fig. 6(c), 1–4, 7). Patent Owner does not contend otherwise at this stage of the proceeding. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that TLP teaches this limitation.

Claim 1 recites “determining a probability that the object is in each image.” Petitioner contends that TLP teaches this limitation in disclosing an inference algorithm that combines the body model with a probabilistic image likelihood model to define a likelihood of observing the image measurements. Pet. 70–71 (citing Ex. 1009, 4). Patent Owner does not contend otherwise at this stage of the proceeding. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that TLP teaches this limitation.

Claim 1 recites “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.” Petitioner contends that TLP teaches this limitation in determining a likelihood that an object is in each image and estimating a belief in the graphical model. Pet. 71 (citing Ex. 1009, 4; Ex. 1004 ¶ 214). Dr. Bajaj testifies that to infer the belief in a model, a person of ordinary skill would have known to compare a determined probability against a threshold. Ex. 1004 ¶ 214 (citing Ex. 1038, 15; Ex. 1041, 3).

Patent Owner contends that Petitioner has not shown a reason to compare the probability against a threshold, has not shown that comparing probabilities against a threshold was inherent to estimating a belief, and has not shown that comparing the probability to a threshold would have been common sense. Prelim. Resp. 65–67.

At this stage of the proceedings, we rely on Dr. Bajaj's testimony and supporting evidence in determining that the Petition shows a reasonable likelihood that a person of ordinary skill would have known to compare the probabilities to a threshold in order to detect the object. Based on the evidence and arguments currently of record, for purposes of institution, Petitioner has sufficiently demonstrated that TLP teaches this limitation.

Based on the evidence and arguments currently of record, for purposes of institution, we are persuaded that Petitioner has demonstrated a reasonable likelihood of showing how TLP teaches or suggests each limitation of claim 1. However, as noted above, at this stage of the proceeding we find that Petitioner has not shown a reasonable likelihood of showing that TLP is prior art to the '980 patent. Accordingly, at this stage of the proceeding, Petitioner has not shown a reasonable likelihood of prevailing in showing claim 1 is obvious over TLP.

3. *Claims 2–16*

Petitioner challenges claims 2–16 as unpatentable over TLP. Pet. 72–76. Patent Owner does not separately argue against Petitioner's showing with respect to these claims. *See* Prelim. Resp. At this stage, because Patent Owner does not separately argue with respect to the other challenged claims of this ground, we do not see the need to further address the challenges to claims 2–16 on this ground. Further analysis is best left for trial after full development of the record.

VIII. CONCLUSION

Based on the arguments presented in the Petition, we conclude that Petitioner has demonstrated a reasonable likelihood of prevailing with respect to claim 1 of the '980 patent challenged in the Petition. Accordingly,

we institute a trial on all claims and all grounds asserted in the Petition. The Board has not made a final determination under 35 U.S.C. § 318(a) with respect to the patentability of any challenged claim. Any final determination will be based on the record developed during trial. We place Patent Owner on express notice that any argument not asserted in a timely-filed Response to the Petition, or in another manner permitted during trial, may be deemed waived, even if that argument was presented in the Preliminary Response.

IX. ORDER

In consideration of the foregoing, it is hereby

ORDERED that, pursuant to 35 U.S.C. § 314(a), an *inter partes* review of claims 1–16 of the '980 patent is instituted with respect to the grounds set forth in the Petition; and

FURTHER ORDERED that, pursuant to 35 U.S.C. § 314(c) and 37 C.F.R. § 42.4(b), *inter partes* review of the '980 patent shall commence on the entry date of this Decision, and notice is hereby given of the institution of a trial.

IPR2023-00924
Patent 7,436,980 B2

FOR PETITIONER:

W. Todd Baker
Ellisen Shelton Turner
Jonathan D. Brit
KIRKLAND & ELLIS LLP
todd.baker@kirkland.com
ellisen.turner@kirkland.com
jonathan.brit@kirkland.com

FOR PATENT OWNER:

Christine E. Lehman
Michael Matulewicz-Crowley
Philip J. Eklem
REICHMAN JORGENSEN LEHMAN
& FELDBERG LLP
clehman@reichmanjorgensen.com
mmatulewicz-crowley@reichmanjorgensen.com
peklem@reichmanjorgensen.com