Trials@uspto.gov
571-272-7822

Paper 54
Date: December 5, 2024

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.*

Opinion for the Board filed by *Administrative Patent Judge* SMITH.

Opinion Dissenting filed by *Administrative Patent Judge* McKONE.

SMITH, *Administrative Patent Judge*.

DECISION
Final Written Decision
Determining Some Challenged Claims Unpatentable
35 U.S.C. § 318(a)

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). We issued an Institution Decision (Paper 11, "Inst. Dec.") instituting the petitioned review. Patent Owner filed a Request for Rehearing. Paper 13. We denied the Request for Rehearing. Paper 19 ("Reh'g. Dec.").

VL Collective IP LLC ("Patent Owner") filed a Patent Owner's Response (Paper 26, "PO Resp.") pursuant to 35 U.S.C. § 313. Petitioner filed a Reply (Paper 32) and Patent Owner filed a Sur-Reply (Paper 44, "PO Sur-Reply"). We also issued an Order granting Petitioner's Motion to Submit Supplemental Information (Paper 28) and an Order granting Patent Owner's Motion to Submit Supplemental Information (Paper 41). We held a hearing on September 10, 2024, and entered a transcript into the record. Paper 53 ("Tr.").

We have jurisdiction under 35 U.S.C. § 6(b)(4) and § 318(a). This Decision is a final written decision under 35 U.S.C. § 318(a) and 37 C.F.R. § 42.73 as to the patentability of claims 1–16 of the '980 patent. We determine Petitioner has shown by a preponderance of the evidence that claims 1–3, 5–7, 9–11, and 13–15 are unpatentable, and Petitioner has not shown by a preponderance of evidence that claims 4, 8, 12, and 16 are unpatentable.

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 of the ’980 patent are reproduced below.

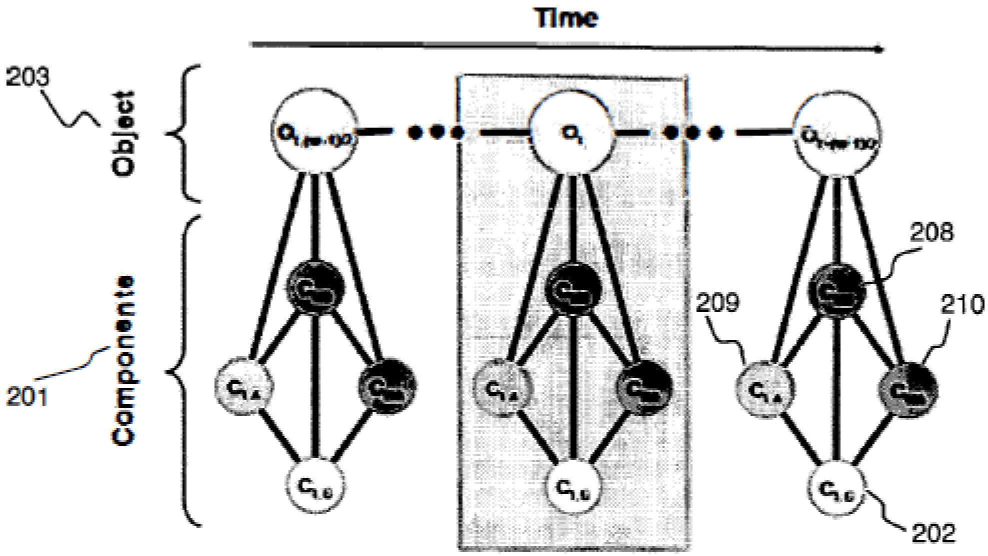


FIGURE 2A

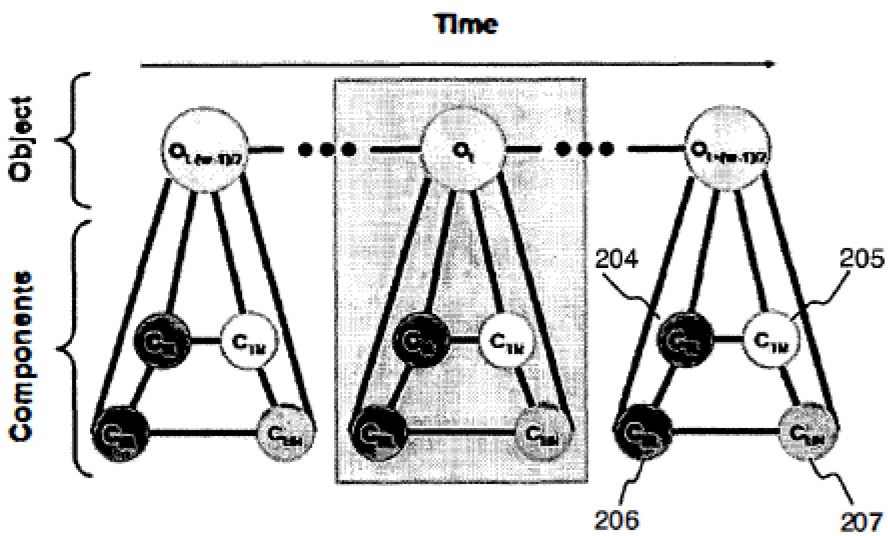


FIGURE 2B

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.

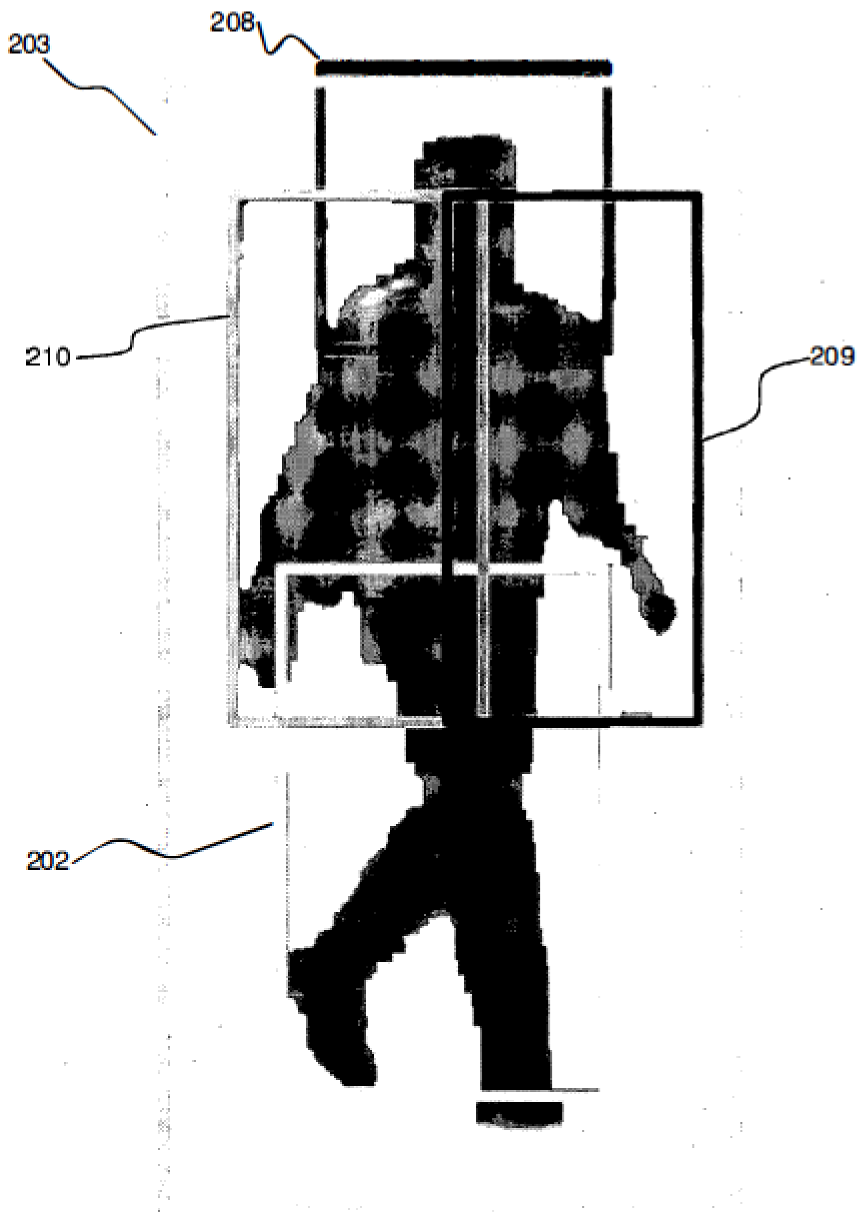


Figure 3 above shows a pedestrian's components 202, 208–210, and "is an illustration of [the] pedestrian and identified components of the pedestrian" (*id.* at 2:31–32).

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." *Id.* at 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):

> [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. 6).

Claim(s) Challenged	35 U.S.C. §[1]	Reference(s)/Basis
1–16	103(a)	Zhao[2]
1–16	103(a)	Steffens,[3] Zhao
1–16	103(a)	Ozer,[4] 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; iss. Oct. 9, 2001 (Ex. 1007).
[4] U.S. Patent No. 7,200,266 B2; iss. 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. PO Resp. 11.

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.

Petitioner contends that, "[f]or purposes of [the] petition only, Petitioners present no terms for construction." Pet. 12.

<u>"Spatio-temporal model"</u>

Claim 1 recites "providing a spatio-temporal model for an object to be detected." Patent Owner, in its Sur-Reply, contends that the claimed "spatio-temporal model" should be construed as a spatial model where an object is detected in an image based on the probabilities that the object is in preceding and/or subsequent images. PO Sur-Reply 15. Patent Owner contends that its proposed construction is in response to Petitioner's alleged construction of this term as a spatial model where an image's analysis takes into account information from preceding and/or subsequent images. *Id.* at n.2 (citing Reply 8–11; Ex. 1052 ¶¶ 6–22, 23–35).

However, pages 8–11 of Petitioner's Reply contend that Zhao teaches the last limitation of claim 1, namely, "detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object." Petitioner's Reply does not provide a construction of the claimed "spatio-temporal model," and in particular, does not provide the construction alleged by Patent Owner in Patent Owner's Sur-Reply. Patent Owner's general citation to Petitioner's Reply at pages 8–11 and subsequent conclusion that the Reply proposes a claim construction of "spatio-temporal model" does not sufficiently show that Petitioner in fact proposed a construction of this term, let alone the specific construction alleged by Patent Owner. In fact, Petitioner, in Reply, stated that Patent Owner's characterization of the claimed "spatio-temporal model" as a two-layer graphical model impermissibly reads limitations from the Specification into the claims, impermissibly excludes other spatio-temporal models from the scope of the claim, and effectively reads dependent claim 3 into claim 1. Reply 4–5.

Our rules state that generally, "a reply or sur-reply may only respond to arguments raised in the preceding brief." Consolidated Trial Practice Guide 2019, 74 (quoting 37 C.F.R. § 42.23). A "sur-reply that raises a new issue . . . may not be considered." *Id.* The Federal Circuit has held that "when the Board adopts a new claim construction following institution, an IPR petitioner must have adequate notice and opportunity to respond under the new construction. In particular, the petitioner must be afforded a reasonable opportunity in reply to present argument and evidence under that new construction." *Axionics v. Medtronic, Inc.*, 75 F.4th 1374, 1383 (Fed. Circ. 2023). In this proceeding, Petitioner was not afforded that opportunity. Therefore, we do not consider Patent Owner's untimely construction of "spatio-temporal model."

Even were we to consider Patent Owner's untimely construction of "spatio-temporal model," we find this proposed construction improperly renders subsequent steps of the claim void, meaningless, or superfluous. For example, Patent Owner, relying on the phrase "for an object to be detected" in the claim term "providing a spatio-temporal model for an object to be detected," contends that the claim and the rest of the Specification indicates that "the function of the spatio-temporal model is to detect an object, and to do so based on the probabilities that the object is in earlier or later images." PO Sur-Reply 16–17. However, the Specification, including claim 1 itself, distinguishes between "providing a spatio-temporal model" and "detecting the object." Reading the scope of "providing a spatio-temporal model for an object" to encompass detecting the object renders the claimed step of "detecting the object" "void, meaningless, or superfluous," which is "counter to an important principle of interpretation." *Intel Corp. v.*

Qualcomm, Inc., 21 F.4th 801, 809–10 (Fed. Cir. 2021). Further, as we noted in our Institution Decision, the claimed "spatio-temporal model" is not used in any subsequent step recited in claim 1. Inst. Dec. 25. Rather, the claim performs the subsequent steps of "providing a video," "measuring the object," "determining a probability," and "detecting the object" without using or referring to the "spatio-temporal model for [the] object" "provid[ed]" in the first step of claim 1. *Id.*

Patent Owner's proposed construction of "spatio-temporal model" also improperly reads limitations from the Specification into the claim. For example, Patent Owner contends that the "spatio-temporal model is a spatial model that is 'extended over time'" and "has probabilistic connectivity, *e.g.*, 'edges,' 'constraints,' or 'compatibilities' connecting the object in multiple frames . . . . The use of object probabilities from multiple frames in the spatio-temporal model is further shown in the joint probability distribution for the spatio-temporal model." PO Sur-Reply 17 (citing Ex. 1001, 2:59–60, 3:47–4:8, 4:34–45, 4:49–52, Figs. 2A and 2B).

However, the cited sections of the Specification describe "a two-layer graphical model" (2:55–60), "a spatio-temporal directed graphical model" (3:47–49), "a two-layer graphical model" (4:37–38), and "a single object layer model . . . built with bi-directional temporal constraints" (4:43–45). In contrast, columns 7 and 8 repeatedly refer to a "spatio-temporal model," not a spatio-temporal graphical model. Ex. 1001, 7:6–8:3. Contrary to Patent Owner's contention, the Specification does not express "a clear indication . . . that the patentee intended the claims to be so limited" to the graphical model described in columns 2 through 4. *See Dealertrack, Inc. v. Huber*, 674 F.3d 1315, 1327 (Fed. Cir. 2012); *accord Phillips v. AWH Corp.*, 415

F.3d at 1323 ("[C]laims may embrace 'different subject matter than is illustrated in the specific embodiments in the specification.'"). Further, claim 3, not claim 1, recites that the "spatio-temporal model is a graphical model." We decline to limit the "spatio-temporal model" recited in claim 1 to the features of the graphical model described in columns 2 through 4 of the Specification. We agree with Petitioner that Patent Owner's proposed construction impermissibly reads limitations from the Specification into the claims, impermissibly excludes other spatio-temporal models from the scope of the claim, and effectively reads dependent claim 3 into claim 1. We agree with Petitioner that the scope of the "spatio-temporal model" recited in claim 1 is not limited to a graphical model, nor to any particular feature of the graphical model, described in the Specification, and encompasses other spatio-temporal models. Reply 4–5 (citing Ex. 1051, 59:6–62:13; Ex. 1052 ¶¶ 6–22).

In our Decision to Institute, we stated that "incorporating temporal information" such as motion information "into the model to detect the moving object" falls within the scope of the claimed "spatio-temporal model." Inst. Dec. 23. We also stated that "kinematic information . . . teaches motion information that provides a temporal element to the model of body parts," relying on Bregler's teaching 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." *Id.* (quoting Ex. 1027, 1). We construe the plain meaning of "a spatio-temporal model for an object to be detected" to encompass at least a model that encodes spatial relationships between parts of the object and incorporates temporal information used to detect the moving object, such as motion information including, but not limited to, Bregler's kinematic information that describes the motion of at least one segment of the object and Zhao's motion information obtained from previous frames. *See id.* at 22–23, 25 (Discussing the plain meaning of this claim term). Thus, a model for an object that encodes spatial relationships between parts of the object and incorporates temporal information of at least one segment of the object falls within the scope of "a spatio-temporal model for an object to be detected" as claimed.

"Detecting the object in any image"

Claim 1 recites "detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object." The parties disagree over the proper construction of this term. PO Resp. 18–25; Reply 1–4. However, we do not need to resolve the disagreement, because we find that the prior art teaches this limitation under the proposed constructions of either party.

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 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

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

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). 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. We agree with Dr. Brenner and Mr. Mageski and find 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.

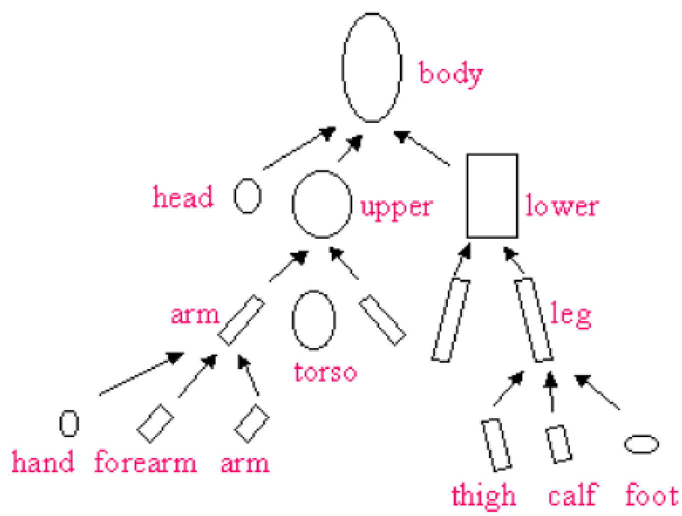
Figures 1.3 and 1.6 above respectively illustrate corresponding "[e]xamples of various appearances due to clothing" (*id.* at 21) and "[e]xamples of various shapes due to clothing (same as Fig. 1.3 but only contour)" (*id.* at 23).

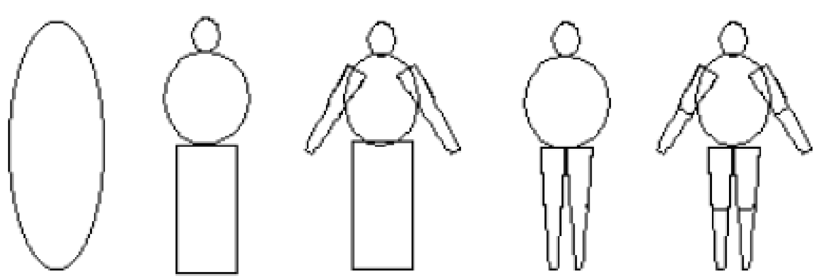
The problem is generalized as follows:

> 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.





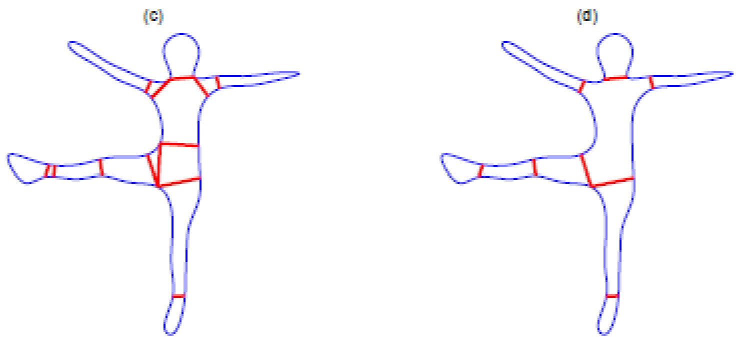
Figures 1.7 and 1.8 above respectively illustrate the solution of "merged body parts to model the merging of multiple body parts" and "[a]ssembling [of] human models" (*id.*).

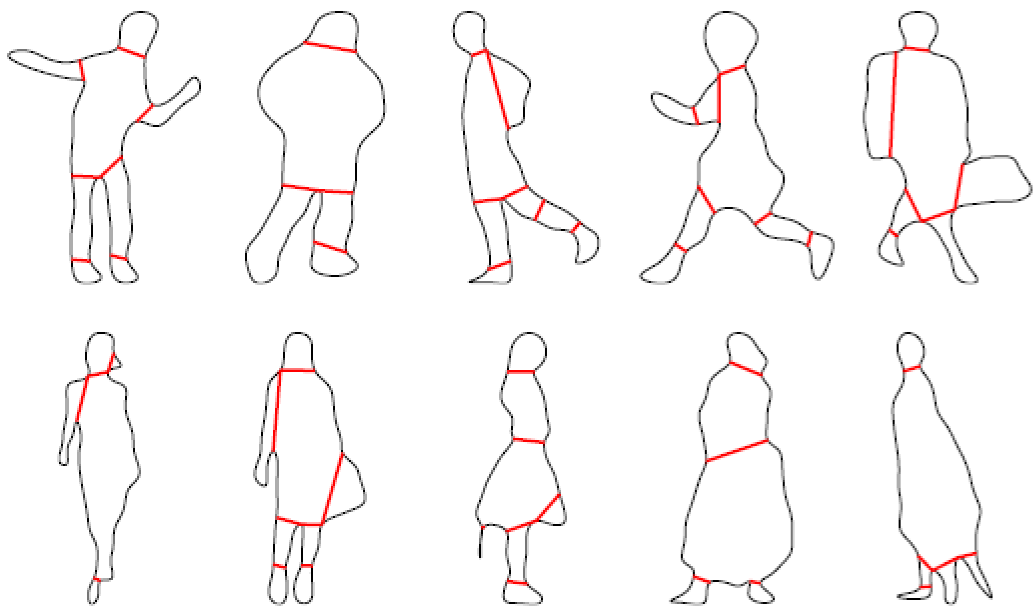
The solution is generalized as follows:

> 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).

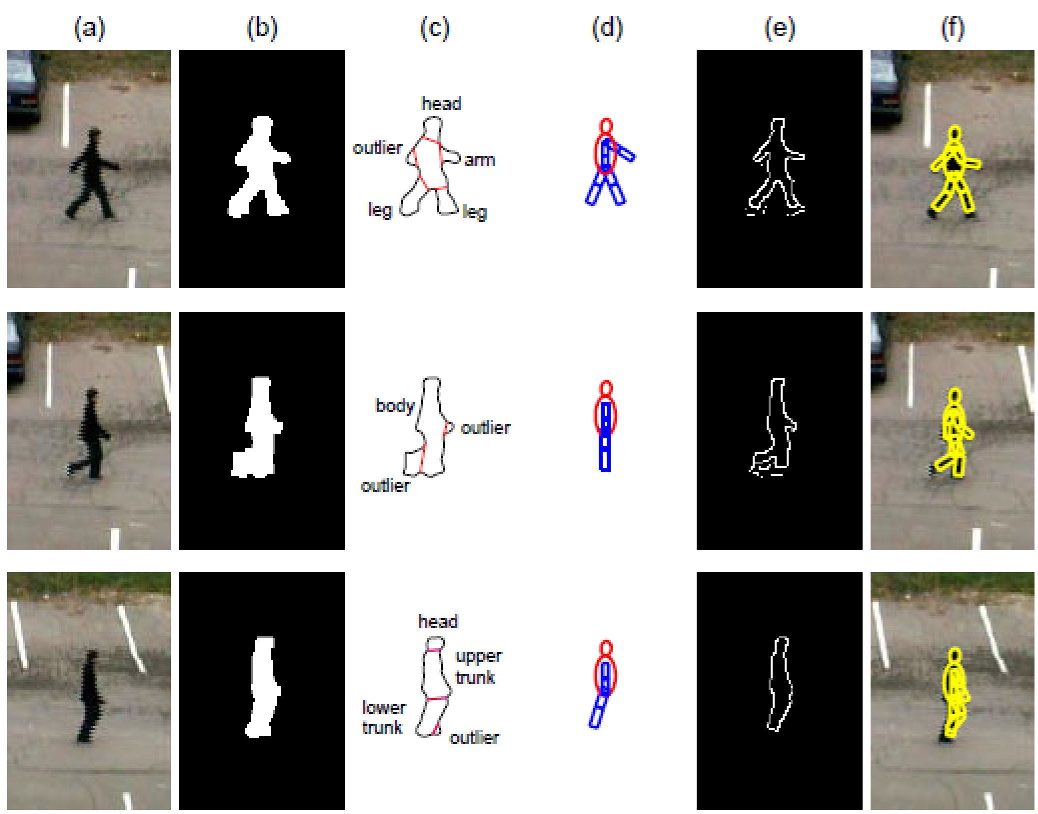
The solution implements "contextual information" as follows:

> 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.*

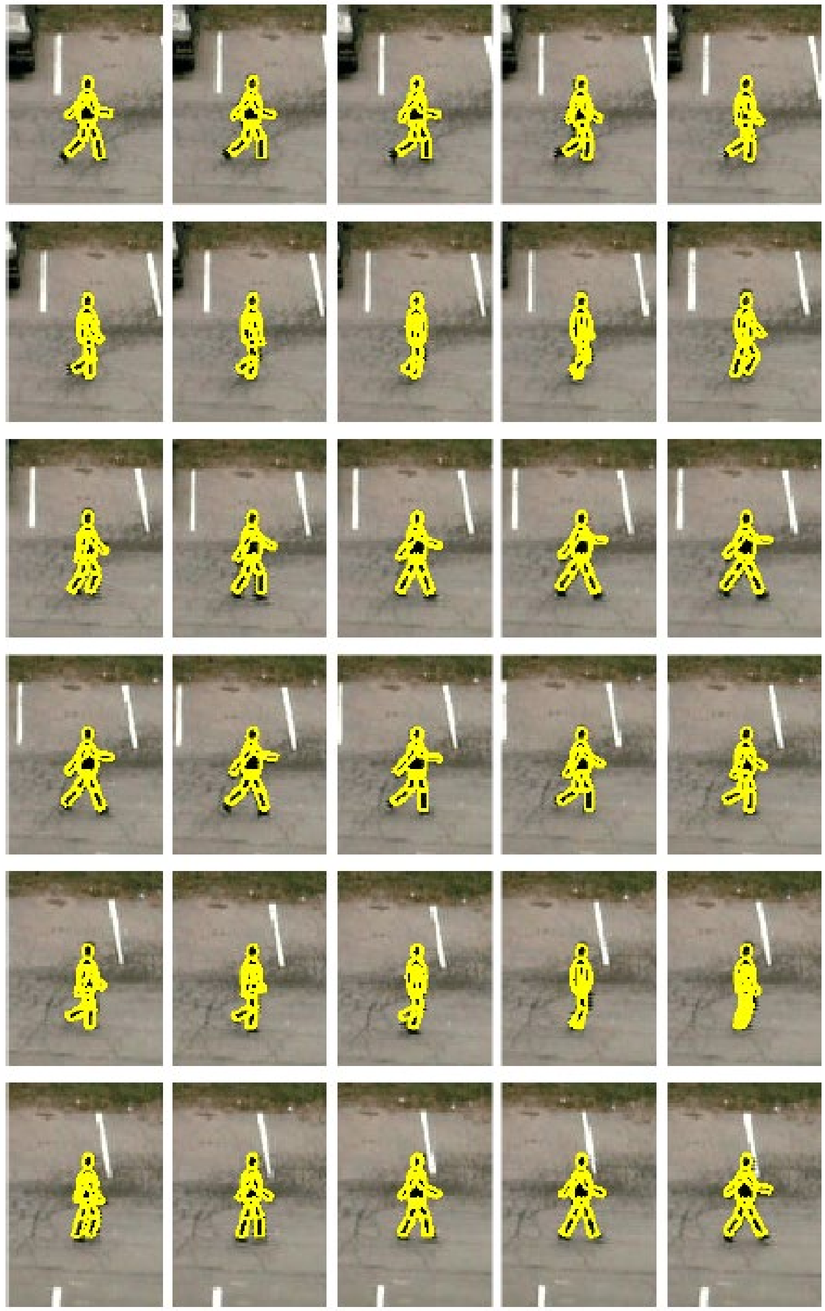


Figure 5.6 above is a series of pictures of a person, outlined in yellow, taken during a full cycle of walking. *Id.* at 94. As shown in Figure 5.6, "[i]n the first half cycle, no motion information is used to resolve the ambiguity with limbs' orientations, while in the second half of the cycle, the 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.

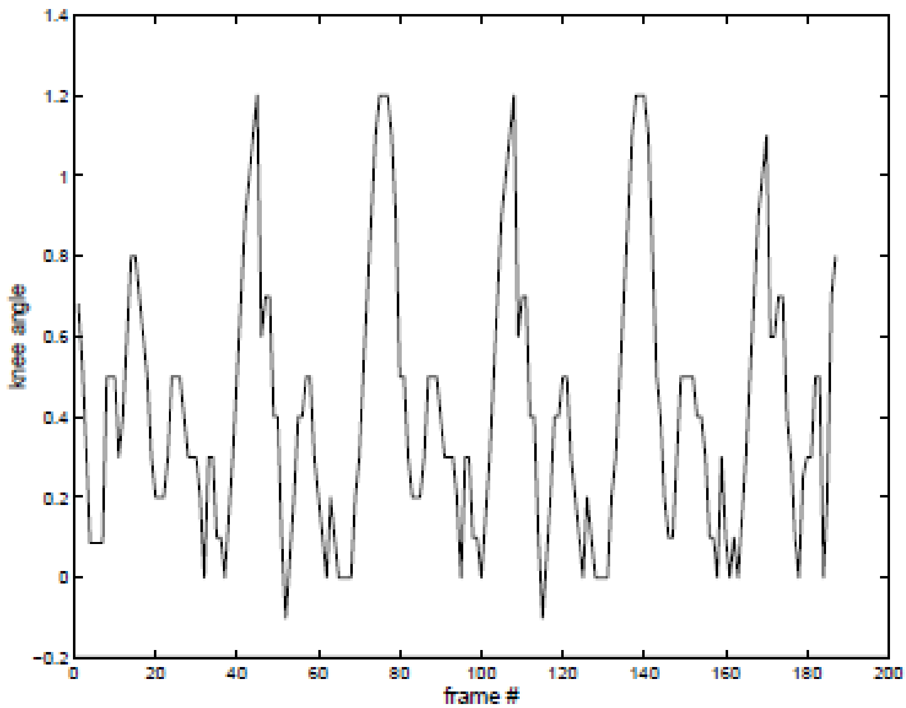


Figure 5.5 above is a graph with knee angle on the y-axis and frame # on the x-axis. As shown in Figure 5.5, "[t]he motion information is not used from frames 1 to 30, but is from frames 31 to 190." *Id.* "Therefore, the left and right legs are switched sometimes during the first part." *Id.* "From frames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

2. 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 understood 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 (*id.* at 13–15, 17–18, 21, 23, 25–27). We will, accordingly, address these "obviousness" contentions when addressing the respective limitations.

3. *Independent claims 1 and 9*

<u>[1preamble]</u>

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. We find that Petitioner has shown that Zhao teaches the preamble of claim 1.[7]

<u>[1a] "Providing a spatio-temporal model for an object to be detected"</u>

Limitation 1a of claim 1 recites "providing a spatio-temporal model for an object to be detected." Petitioner contends that Zhao teaches the spatial feature of the model in disclosing that its model is designed to encode the size and spatial relationships between body parts. Pet. 13–15 (citing Ex. 1006, 69; Ex. 1004 ¶¶ 74–86). Petitioner contends that Zhao teaches the temporal feature of the claimed "spatio-temporal model" by incorporating temporal elements into its object model. *Id.* at 15–17 (citing Ex. 1006, 91–94, Figs. 5.5, 5.6). In particular, Petitioner contends that Zhao teaches using motion information obtained from previous frames to predict the orientation

[7] Because Petitioner has shown that the features in the preamble are taught or suggested by the prior art, we need not determine whether the preamble is limiting. *See Vivid Techs., Inc. v. Am. Sci. & Eng.*, 200 F.3d 795, 803 (Fed. Cir. 1999).

of limbs. *Id.* at 16 (citing Ex. 1006, 91). For example, 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. *Id.* (citing Ex. 1006, 91, 99). Petitioner contends that by incorporating temporal information regarding the change of angle at a subject's knee with time into the model, "an obvious pattern of walking can be observed." *Id.* at 16–17 (citing Ex. 1006, 92–94, Figs. 5.5, 5.6).

Dr. Bajaj testifies that Zhao discloses the claimed "spatio-temporal model" by incorporating temporal elements into its object model. Ex. 1004 ¶ 78. Dr. Bajaj testifies that Zhao's "RCR algorithm is useful for vision tasks such as object tracking . . . by incorporating motion information into the model." *Id.* ¶ 80 (quoting Ex. 1006, 99). Dr. Bajaj testifies that Zhao discloses that "[m]otion information obtained from previous frames can be used to predict the orientations of the limbs and to solve the ambiguity." *Id.* (alteration in original). Dr. Bajaj testifies that Figure 5.5 of Zhao discloses the impact of incorporating temporal elements such as motion information from previous frames. *Id.* ¶ 81. Dr. Bajaj testifies that, by incorporating motion information into the model, Zhao's process can observe an obvious pattern of walking cycle. *Id.* ¶ 82. Dr. Bajaj testifies that Zhao teaches the claimed "spatio-temporal model for an object to be detected" because Zhao's model incorporates temporal elements for an object to be detected. *Id.*

Patent Owner, in its Response, does not present arguments or evidence to the contrary. In its Sur-Reply, Patent Owner contends that Zhao does not teach the claimed "spatio-temporal model" because Zhao's "body model has no temporal components at all." PO Sur-Reply 28. However, Patent Owner did not raise this issue previously in its Response, therefore, this "sur-reply that raises a new issue . . . may not be considered." Consolidated Trial Practice Guide 2019, 74.

Even considering Patent Owner's untimely argument, we disagree with Patent Owner's contention that Zhao's body model has no temporal components. As discussed above, the plain meaning of "a spatio-temporal model for an object to be detected" encompasses at least a model that encodes spatial relationships between parts of the object and incorporates temporal information used to detect the moving object, such as the motion information of Zhao. Ex. 1006, 91. Therefore, we agree with Petitioner and Dr. Bajaj and find that Zhao's model (a) encodes spatial relationships between parts of the object and (b) incorporates motion information obtained from previous frames to detect the object. Pet. 13–17; Ex. 1004 ¶¶ 74–82; Ex. 1006, 91–95. We find that Zhao's model teaches "a spatio-temporal model for an object to be detected" as claimed. We find that Petitioner has shown that Zhao teaches limitation 1a.

[1b] "Providing a video comprising a plurality of images"

Limitation 1b of 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. We find that Petitioner has shown that Zhao teaches limitation 1b.

<u>[1c] "Measuring the object as a collection of components"</u>

Limitation 1c of 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. We find that Petitioner has shown that Zhao teaches limitation 1c.

<u>[1d] "Determining a probability that the object is in each image"</u>

Limitation 1d of 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. We find that Petitioner has shown that Zhao teaches limitation 1d.

<u>[1e] "Detecting the object in any image"</u>

Limitation 1e of 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 Petitioner does not explain how Zhao teaches limitation 1e. PO Resp. 26–30. Patent Owner contends that Zhao does not disclose a way of incorporating prior or later knowledge of the identification of the object into a detection framework. *Id.* at 30. Patent Owner contends that Zhao's Recursive Context Reasoning (RCR) algorithm does not detect a human in an image under consideration based on the probability of humans being detected in other images. *Id.* at 31. Patent Owner contends that the RCR algorithm does not involve determining or considering the probability of the object (the person) being present in another image or comparing that probability to a threshold. *Id.* at 32–33. Patent Owner contends that although Zhao uses motion information from previous frames to predict the orientations of limbs and to solve ambiguity, a person of ordinary skill in the art would have understood that this motion information does not include, nor is it based on, the probability of the object being present in another image or comparing that probability to a detection threshold. *Id.* at 33.

Petitioner contends that Patent Owner's argument is based on a new proposed claim construction by Patent Owner, that the claimed "detecting the object" requires multi-image analysis. Reply 8. Petitioner contends that Patent Owner ignores Zhao's disclosure that in a "cycle of a person walking, the first half of the cycle involves no motion information, while the second half involves '***the prediction from previous frames*** (constant angular velocity is assumed)' in order to get better results of body part localization." *Id.* at 10 (quoting Ex. 1006, 92) (emphasis in original). Petitioner contends that a person of ordinary skill in the art "would understand that if the cycle of a person walking involves prediction from previous frames, not only is there a temporal component, the RCR algorithm contour updating procedure and prediction procedure would contain comparison of that probability in the previous frame to a detection threshold." *Id.* at 10–11.

Patent Owner in its Sur-Reply contends that Zhao does not teach this limitation because Zhao's body model has no temporal probabilistic connectivity. PO Sur-Reply 28. Patent Owner contends that Zhao's body model does not combine probabilities from multiple frames and has no temporal components at all. *Id.* According to Patent Owner, while Zhao's RCR process can use some limited information from previous frames in the form of motion information, the motion information does not include probabilities for objects. *Id.* Patent Owner contends that "Zhao therefore does not disclose the claimed spatio-temporal model." *Id.*

Analysis

We agree with Petitioner that Patent Owner's contentions are based on Patent Owner's new claim construction of "detecting the object in any image" as requiring comparing multiple probabilities from multiple images to a threshold. Reply 8; *see* PO Resp. 29–30 (Patent Owner contending that the Petition does not show how Zhao teaches "'detecting the object *in any image*' *i.e.*, in any given image, 'upon comparing the probabilities *for each image*,' *i.e.*, two or more images, 'to a threshold for detecting the object.'"). (emphasis in original). This new construction is in contrast to Patent Owner's originally proposed claim construction in the Preliminary Response, where Patent Owner construed this limitation to encompass the specification's disclosure that a w-frame windowed smoothing algorithm may be used where w is an odd integer ≥ 1. Prelim. Resp. (Paper 7) 43 (quoting Ex. 1001, 7:56–8:7). In our Institution Decision we applied Patent Owner's originally proposed construction that the scope of this limitation encompasses w=1 in deciding that Petitioner sufficiently showed that the prior art taught this limitation. Inst. Dec. 26–28; Reh'g. Dec. 3. Patent

Owner's new construction proposes excluding the disclosed embodiment of w=1 from the scope of the claim. The Federal Circuit has held that "when the Board adopts a new claim construction following institution, an IPR petitioner must have adequate notice and an opportunity to respond under the new construction." *Axionics*, 75 F.4th at 1383. As discussed above, we do not need to resolve whether Patent Owner's new proposed construction of this claim term is correct, because even excluding the disclosed embodiment of w=1, such that this limitation requires comparing probabilities "[of] at least two" images (PO Resp. 19; *see id.* at 18–25), we find that Zhao's RCR algorithm teaches "detecting the object in any image" as claimed even under this construction. Because our patentability analysis assumes without deciding that Patent Owner's new proposed construction is correct, we consider Petitioner's arguments in Reply so that Petitioner has "an opportunity to respond under the new construction."

According to Patent Owner, the scope of "detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object" encompasses "compar[ing] the likelihood of object presence in each image to a threshold and send[ing] 'messages' based on that likelihood (e.g., when it scores above the threshold) to the object in the image under consideration." PO Resp. 25; *see* .PO Sur-Reply at 21–22 (citing Ex. 1003, 8–9) (contending that "send[ing] messages . . . between frames" as disclosed in the provisional application provides written description support for this limitation). We agree with Petitioner that Zhao teaches that the RCR algorithm uses motion information messages sent "from previous frames" when detecting a person in a cycle of walking. Reply 10 (quoting Ex. 1006, 92). We agree with Petitioner that a person of ordinary skill in the art would have understood "that if the cycle of a person walking involves prediction from previous frames, not only is there a temporal component, the RCR algorithm contour updating procedure and prediction procedure would contain a comparison of that probability in the previous frame to a detection threshold." *Id.* at 10–11 (citing Ex. 1006, 56, 91–95).

In particular, we find that the motion information of Zhao is based on comparing the likelihood of object presence in previous images to a threshold. Ex. 1006, 56, 64–65, 91–95, Fig. 5.6; *see id.* at 92 ("In the first half [of the] cycle [of Figure 5.6], 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 . . . is used to get better results of body part localization."). When the likelihood of object presence in previous images scored above the threshold, such as in frames 1 to 30 as described with respect to Figure 5.5, the object was detected and the motion information was determined from the detected object. *Id.* at 65 ("[T]he similarity measure . . . can be used to perform human detection: the contour C corresponds to a person if BSM(C) ≥ threshold."), 91 ("Motion information obtained from previous frames can be used."); *see id*. at Fig. 5.6. When the likelihood scored below the threshold in a frame, the object was not detected, therefore, no motion information for the object was determined from that frame. *See id.* The motion information includes information about the likely location of the object in the current image. *Id.* at 91–93. We find that even under Patent Owner's new proposed construction, the motion information of Zhao teaches "compar[ing] the likelihood of object presence in each image to a threshold and send[ing] [motion information] 'messages' based on that likelihood (e.g., when it scores above the threshold) to the object under consideration." PO Resp. 25.

We further find that Zhao teaches "detecting the object in any image upon comparing the probabilities" even under Patent Owner's new proposed construction. Zhao describes using the motion information in the process of detecting an object in a current frame, such as frame 31 as described with respect to Figure 5.5, using the RCR algorithm. Ex. 1006, 26, 91–94. Figures 5.4(a)–(f) show results produced at various stages of the RCR algorithm and are reproduced below. *Id.* at 92; *see id.* at 91.

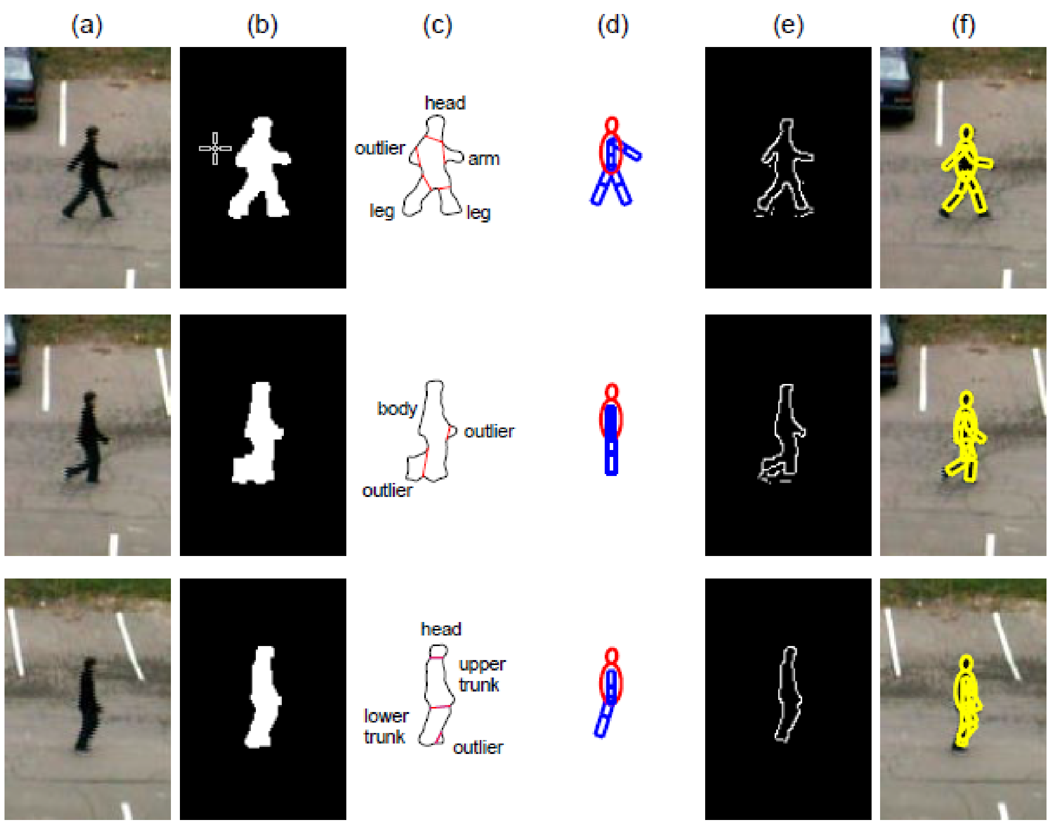


Figure 5.4: Body part localization: (a) images (b) foreground object detected from background subtraction (c) identified body parts (d) updated/predicted locations and outlines of the body parts (e) edge images (f) aligned body parts.

Figures 5.4(a)–(f) above show examples of locating a person in a coarse-to-fine manner using the RCR algorithm. *Id.* at 91–92; *see id.* at 26, Fig. 1.9. As shown in Figure 5.4(a), an image is input into the RCR algorithm. *Id.* at 91. The person is segmented from the background as shown in Figure 5.4(b). *Id.* The segmented region is decomposed into ribbons and these ribbons are matched with the modeled body parts as shown in Figure 5(c). *Id.* The locations and the sizes of missed body parts are inferred from the detected body parts as shown in Figure 5(d). *Id.* The predicted outlines of the body parts are aligned with the edge features as shown in Figure 5(e). *Id.* In this example, which shows a side-view of a person, the arms may be aligned with the outline of the torso when the arms are overlapped with the torso, or the left and right limbs may be confused. *Id.* Motion information obtained from previous frames is used to predict the orientations of the arms and to solve the ambiguity. *Id.* After aligning the outline using motion information, the object is detected as shown in Figure 5.4(f). *Id.*

A flow chart of the RCR algorithm that detects the object is shown in Figure 1.9 of Zhao and is reproduced below.

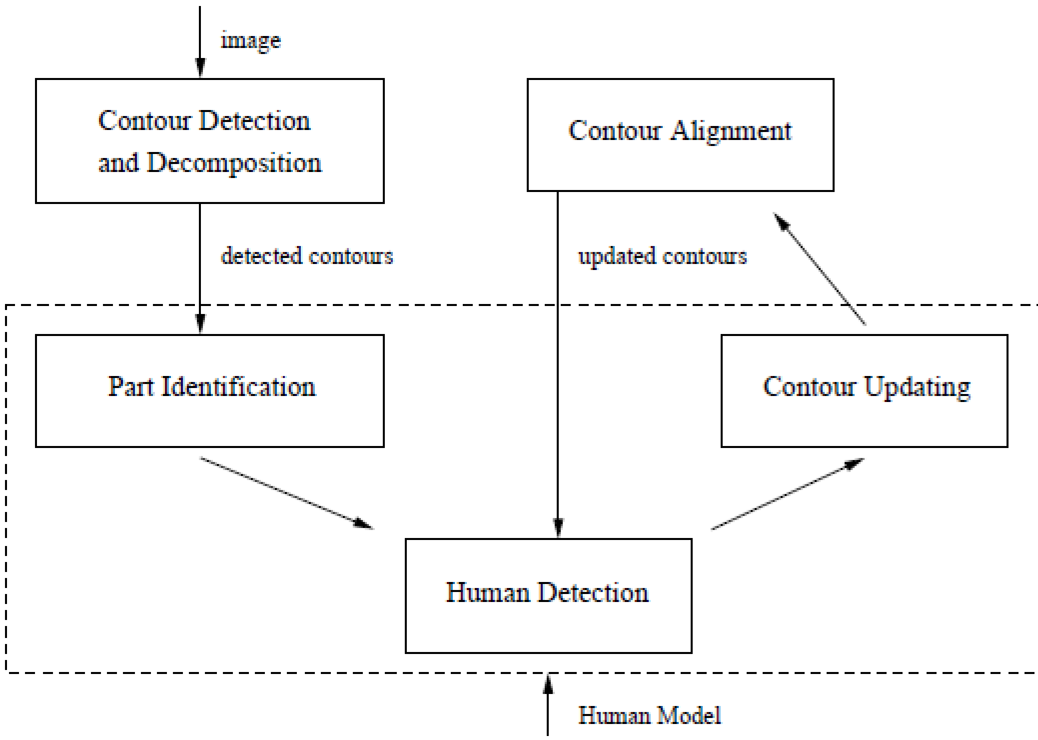


Figure 1.9: Flow chart of the RCR algorithm

Figure 1.9 above shows a flow chart of the RCR algorithm. *Id.* at 26. As shown in Figure 1.9, the algorithm receives an image, detects contours, identifies parts of the object, and detects the object. *Id.* at 26, 65, 91, Figs. 5.4(a), 5.4(b), and 5.4(c). Then, the algorithm updates and aligns the contours. *Id.* at 26, 91, Figs. 5.4(d) and 5.4(e). As discussed in the previous paragraph, aligning the contours includes using motion information obtained from previous images. *Id.* at 91. Then, the refined contours are used to reevaluate the likelihood of a person being present in the image as shown in Figure 1.9. *Id.* at 26, 65, 91, Fig. 5.4(f). That is, after initially detecting the object by comparing the probability that the object is in the current image to a threshold, the recursive algorithm updates and aligns the contours using motion information (which is "based on comparing the probabilities to a threshold" in other images under Patent Owner's construction (PO Resp. 24)), then repeats the step of detecting the object by comparing the probability of the current image to the threshold as shown in Figure 1.9. *Id.* at 26, 65, 91–92; *see id.* at 56 (Figure 3.1 showing model body parts identified in a coarse-to-fine manner, where "[t]he more body parts being identified, the more likely the extracted contour is a person."); 32 ("The hierarchical organization of the body parts allows efficient object recognition . . . in a coarse-to-fine manner.").

Patent Owner, at the hearing, contended that the motion information of Zhao has an "attenuated relationship between the computation of the motion vector and the detection of the object in a prior image" that is excluded by Patent Owner's new proposed construction that "upon" as claimed encompasses "based on," because, according to Patent Owner, "based on" means that "[i]t's got to be more direct than that. You've got to actually be using it." Tr. 55:11–18, 58:11–14; *see id.* at 60:1–5. We disagree with Patent Owner's contention presented at the hearing. Under Patent Owner's new proposed construction, "detecting the object in any image '*upon*' comparing the probabilities for each image to a detection threshold means detecting the object '*based on*' comparing the probabilities to a threshold." PO Resp. 24. According to Patent Owner, "'object detection' . . . includes instances of the object at a 'current time' and 'earlier' or 'later' times." *Id.* at 24–25. Patent Owner continues explaining in its Response that a person of ordinary skill in the art "would understand the claims as claiming a process of detecting an object in an image is *based on* the probabilities of the object being in other images." *Id.* at 25. Patent Owner's new proposed construction does not include "direct," and we do not read "direct" into the claim. Even under Patent Owner's new proposed construction of "upon" as meaning "based on," Zhao's process of detecting an object in an image is based on the probabilities of the object being in other images.

Patent Owner further argued at the hearing that "Zhao says almost nothing about the motion vector. How it's calculated or where it comes from, or even how it's used other than vaguely." Tr. 55:4–6. However, under Patent Owner's new proposed construction, messages are sent to the object in the image under consideration based on the likelihood that the object was detected in an earlier or later image. PO Resp. 25. Patent Owner's new proposed construction says almost nothing about the message sent to the object in the image under consideration, how the message is calculated or where the message comes from, or even how the message is used, other than the message is based on the probability of the object being in another image. *See id.* at 24–25. Before comparing the current image to a threshold for detecting the object, Zhao has compared other images to a threshold, and sends the resulting motion information, which is "based on comparing the probabilities to a threshold," to the current image in order to detect the object in the current image. Thus, Zhao's "[m]otion information obtained from previous frames" (Ex. 1006, 91) teaches messages based on comparing the probabilities in previous images to a threshold under Patent Owner's new proposed construction.

Contrary to Patent Owner's contentions, we find (a) the motion information of Zhao incorporates prior knowledge of the identification of the object into the detection framework, (b) the motion information of Zhao is "based on comparing the probabilities" of a person being present in other images to "the threshold" under Patent Owner's proposed construction, and (c) the RCR algorithm detects a person in the image under consideration using motion information messages, sent to the object in the image under consideration for detecting the object. Thus, we find that Zhao discloses that the RCR algorithm uses motion information of an object detected in previous images when detecting an object in the image under consideration by sending motion information "messages" based on the probabilities of object presence in previous images scoring above a threshold to the object in the image under consideration. Ex. 1006, 26, 56, 64–65, 91–95; *see* PO Resp. 24–25. We find that even under Patent Owner's new proposed construction, the motion information of Zhao teaches "compar[ing] the likelihood of object presence in each image to a threshold and send[ing] 'messages' based on that likelihood (e.g., when it scores above the threshold) to the object under consideration" in the current image. PO Resp. 25. We find that that Zhao's RCR algorithm teaches a "process of detecting an object . . . based on the probabilities of the object being in other images" as well as the probability of the object being in the current image. *Id.*

We find that Petitioner has shown that Zhao teaches this limitation, and that Petitioner has shown, by a preponderance of the evidence, that Zhao would have rendered claim 1 obvious. With respect to independent claim 9, Patent Owner relies on the arguments presented for claim 1. We disagree for the reasons given in our analysis of claim 1 and find Petitioner has shown, by a preponderance of the evidence, that Zhao also would have rendered claim 9 obvious.

4. Claims 2, 3, 5–7, 10, 11, and 13–15

Patent owner does not separately argue dependent claims 2, 3, 5–7, 10, 11, and 13–15. Therefore, the dependent claims fall together with the independent claims. *Incept LLC v. Palette Life Sciences, Inc.*, 77 F.4th 1366, 1375 (Fed. Cir. 2023). Having considered the complete record before us, we find that Petitioner has shown, by a preponderance of the evidence, that Zhao would have rendered claims 2, 3, 5–7, 10, 11, and 13–15 obvious.

5. Claims 4 and 12

Claim 4 depends from claim 1 and recites "detecting the object in a current image according to measurements of the object as a collection of components determined from a prior image and a later image relative to the current image." The Petition contends that "while Zhao explicitly describes this principle by detecting an object according to measurements of the object as a collection of components determined from a 'prior image,' a [person of ordinary skill in the art] would have further understood that doing so with a 'later image' too would have been obvious." Pet. 29 (citing Ex. 1004 ¶¶ 106–110).

Patent Owner contends that Petitioner and Dr. Bajaj do not identify any disclosure in Zhao that teaches the use of a later image as claimed. PO Resp. 35–36. Patent Owner contends that Zhao's use of motion information obtained from previous frames does not teach using a later image. *Id.* at 37. Patent Owner contends that Petitioner and Dr. Bajaj have not provided a rationale explaining why a person of ordinary skill in the art would have modified Zhao to arrive at the specific limitations of claim 4. *Id.*

Petitioner contends that Patent Owner ignores Zhao's disclosure analyzing data in image frames from a video using measurements from adjacent frames necessarily includes previous and later frames. Reply 11. Petitioner contends that Dr. Bajaj clarifies that it would have been obvious to extrapolate from later frames. *Id.* (citing Ex. 1052 ¶¶ 36–39). Petitioner contends that a person of ordinary skill in the art would have understood that adjacent features would involve looking at prior and later images in order to predict positions associated with said features. *Id.* at 12.

Dr. Bajaj, in his original Declaration, cites to several references in asserting that analyzing data from video by using measurements from previous and later frames was widespread. Ex. 1004 ¶ 109 (citing Ex. 1011, 1, 9; Ex. 1037; Ex. 1024, 339–57). In his supplemental Declaration, Dr. Bajaj cites to Sigal's disclosure that "forward-backward smoothing, either over a time window or an entire sequence, is straight forward."). Ex. 1052 ¶ 37 (quoting Ex. 1015, 7). Dr. Bajaj also cites to Baumberg's disclosure of approximating a spatio-temporal shape model walking sequence given an input video image sequence through measurement estimations of shape displacements in each image frame and estimations of differences of velocity and acceleration between 3 adjacent image frames. *Id.* ¶¶ 38–39 (citing Ex. 1011, 1, 3, 9, Fig. 5).

Analysis

We agree with Patent Owner that Zhao discloses using motion information obtained from previous frames, but does not disclose using motion information obtained from future frames. Ex. 1006, 91. Petitioner has not identified any reason why a person of ordinary skill in the art would have modified Zhao's disclosure of motion information to obtain motion information from both previous and future frames in order to yield "measurements of the object as a collection of components determined from a prior image and a later image relative to the current image" as recited in claim 4.

We are not persuaded by Petitioner's reliance on Dr. Bajaj's testimony. In his original Declaration, Dr. Bajaj does not persuasively explain how any of Exhibits 1011, 1024, and 1037, alone or in combination with Zhao, teach "measurements of the object as a collection of components determined from a prior image and a later image relative to the current image" as claimed. In his supplemental Declaration, Dr. Bajaj's testimony that "forward-backward smoothing forward-backward smoothing, either over a time window or an entire sequence, is straight forward" (Ex. 1052 ¶ 37 (quoting Ex. 1015, 7)) does not explain how or why a person of ordinary skill would have either modified or replaced Zhao's motion information with Sigal's forward-backward smoothing and/or with Baumberg's measurement estimates of shape displacements and differences of velocity and acceleration to yield "measurements of the object as a collection of components determined from a prior image and a later image relative to the current image" as recited in claim 4.

We find that Petitioner has not shown, by a preponderance of the evidence, that Zhao teaches "detecting the object in a current image according to measurements of the object as a collection of components determined from a prior image and a later image relative to the current image" as recited in claim 4. Claim 12 contains a similar limitation. The Petition relies on the contentions presented for claim 4 in contending that claim 12 would have been obvious. Pet. 29. We disagree and find that Petitioner has not shown, by a preponderance of the evidence, that Zhao teaches the "detecting" limitation of claim 12.

6. Claims 8 and 16

Claim 8 depends from claim 1 and recites "a joint probability distribution for the spatio-temporal model with N components is:

$$P\left(X_0^O, X_0^{C_0}, X_0^{C_1}, \ldots, X_0^{C_N}, \ldots, X_T^O, X_T^{C_0}, X_T^{C_1}, \ldots, X_T^{C_N}\right) = \frac{1}{Z}\prod_{ij}\psi_{ij}(X_i^O, X_j^O)$$

$$\prod_{ik}\psi_{ik}\left(X_i^O, X_i^{C_k}\right)\prod_{ikl}\psi_{kl}\left(X_i^{C_k}, X_i^{C_l}\right)\prod_{i}\phi_i(X_i^O)\prod_{ik}\phi_i\left(X_i^{C_k}\right).$$

Petitioner contends that the joint probability used in Zhao would have been known to a person of ordinary skill in the art as a pairwise Markov field representation of a spatio-temporal graphical model. Pet. 34–35 (citing Ex. 1012, Section 2.1; Ex. 1004 ¶ 121). Patent Owner contends that neither Petitioner nor Dr. Bajaj provides an explanation as to how it would have been obvious to modify Zhao to arrive at the particular probability distribution recited in claim 8. PO Resp. 38. Petitioner contends that Zhao's teachings disclose a joint probability represented by the claimed equation, and any differences in the representation of the information and relationship would have been obvious. Reply 12 (citing Ex. 1052 ¶¶ 65–85). Petitioner contends that "the lack of a certain specific node or having extra nodes does not affect the formula of the joint probability distribution" because "Sudderth teaches the same joint probabilistic distribution formulas . . . using the same belief propagation approach as the '980 Patent." *Id.* at 13. Patent Owner contends that Petitioner's attempt to distill claim 8 to a generic formula reads out any distinction between the claimed “object” and “components.” PO Sur-Reply 30.

We find that Petitioner has not persuasively explained how a person of ordinary skill in the art would have modified the teachings of Zhao to arrive at the claimed “joint probability distribution.” We agree with Dr. Saber that the formulas cited by Dr. Bajaj “would have to be rewritten in order to be like that of claim 8.” Ex. 2007 ¶ 45. We agree with Dr. Saber that a person of ordinary skill in the art would have understood that the formulas in the references cited by Dr. Bajaj could not simply be rewritten or expanded to yield claim 8, “because they do not include representations of each piece of information in claim 8.” *Id.*

We find that Petitioner has not shown, by a preponderance of the evidence, that Zhao would have rendered claim 8 obvious. Because claim 16 recites a similar limitation, we find that Petitioner has not shown, by a preponderance of the evidence, that Zhao would have rendered claim 16 obvious.

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.” Ex. 1007, 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 1:14–17. 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 [a] 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.

The portion of 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.

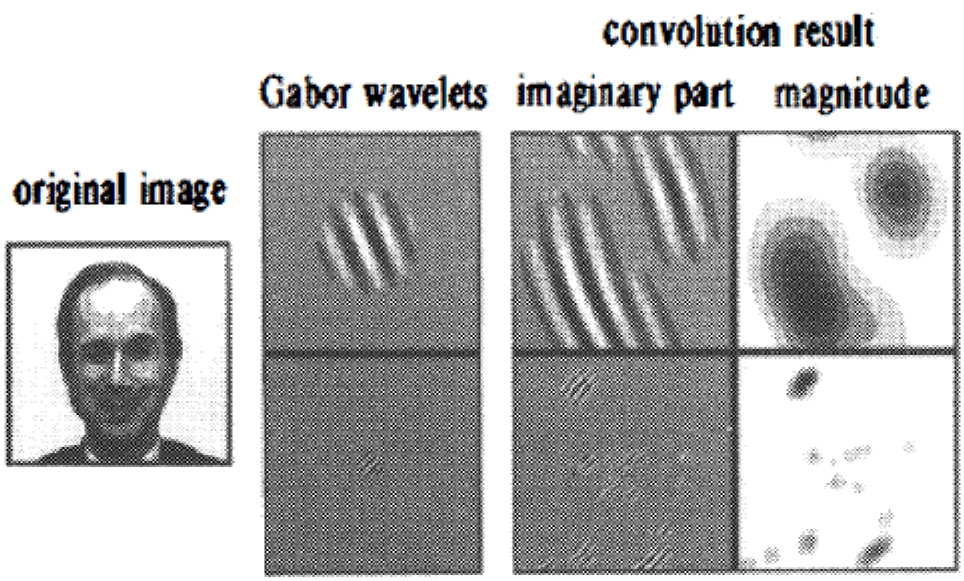


Figure 10 above describes a wavelet transform (*id.* at 8:14–15), presented as "a series of images showing processing of a facial image using Gabor wavelets" (*id.* at 2:46–47). "A wavelet, centered at image position [x] is used to extract the wavelet component [J] from the image." *Id.* at 8:35–36. 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.

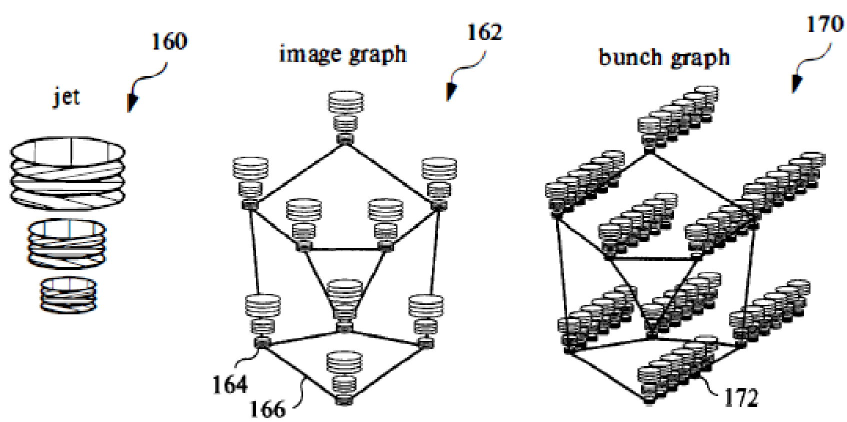


Figure 11 above shows a "labeled image graph 162 . . . used to describe the aspects of an object (in this context, a face)." *Id.* at 8:56–57. "The nodes 164 of the labeled graph refer to points on the object and are labeled by jets 160." *Id.* at 8:58–59. "Edges 166 of the graph are labeled with distance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

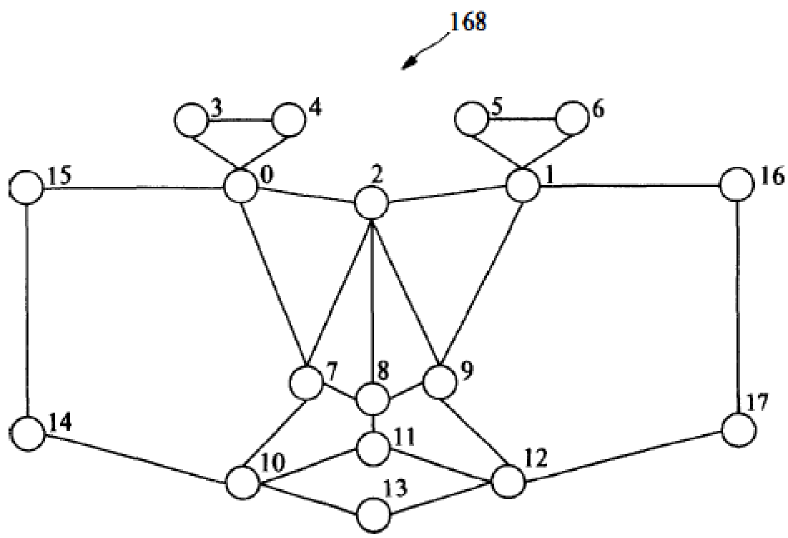


Fig 12 illustrates "[a] model graph 168" having a number of nodes 0–17. *Id.* at 9:1–21. A bunch graph 170 (not shown) is made by attaching "a whole bunch of jets 172 . . . to each node" in graph 168 rather than "only a single jet to each node." *Id.* at 9:22–25. "Each jet is derived from a different 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. Pet. 44 (citing Ex. 1004 ¶ 144). Petitioner contends that although Steffens does not explicitly state that the similarity measure calculates a probability that an object is in an image, doing so would have been obvious to a person of ordinary skill in the art. *Id.* In particular, Petitioner contends that a person of ordinary skill would have known that similarity measures may include probabilistic frameworks for identifying objects. *Id.* at 36 (citing Ex. 1004 ¶ 125).

Petitioner contends that one such similarity measure is described in Zhao as a Bayesian Similarity Measure. *Id.* (citing Ex. 1006, 25). Petitioner contends that Zhao uses the Bayesian Similarity Measure when detecting objects in an image. *Id.* at 44 (citing Ex. 1004 ¶ 145). Petitioner contends that a person of ordinary skill would have combined the Bayesian Similarity Measure of Zhao with the similarity measure of Steffens for the benefit of combining local and global relationship constraints into a single equation to evaluate the degree of resemblance as taught by Zhao. *Id.* at 36 (citing Ex. 1006, 25). Petitioner contends that a person of ordinary skill in the art would have had a reasonable expectation of success, because the similarity measure of Zhao elaborates on the similarity measure of Steffens. *Id.*

Patent Owner contends that a person of ordinary skill in the art would have understood that Steffens's wavelet based approach is fundamentally different and incompatible with Zhao's contour based approach. PO Resp. 42 (citing Ex. 2004 ¶ 129). According to Patent Owner, Zhao's Bayesian Similarity Measure for human detection is fundamentally based on contour extraction, which a person of ordinary skill in the art would have viewed as incompatible with Steffens's wavelet based approach. *Id.* Patent Owner also contends that Zhao teaches away from combining Steffens's wavelet based bunch graph approach with Zhao. *Id.* at 42–43 (citing Ex. 2004 ¶ 130).

Petitioner contends that probabilistic frameworks were widely known in the prior art and were frequently used as part of tracking objects in image video sequences, such as the sequences of Steffens and Zhao. Reply 14–16. Petitioner contends that neither reference teaches away from the other. *Id.* at 16. Petitioner contends that a person of ordinary skill in the art would have known the benefit of Steffens's wavelet approach, and why such approach would work specifically in recognizing facial features. *Id.* (citing Ex. 1052 ¶¶ 40–57).

Patent Owner contends that Steffens's approach is not probabilistic. PO Sur-Reply 31 (citing Ex. 2007 ¶ 48). Patent Owner contends that Steffens includes no discussion of a probabilistic similarity measure. *Id.* Patent Owner contends that Steffens does not simply use different terminology than Zhao or the '980 patent—it employs a fundamentally different framework that is incompatible with Zhao's. *Id.*

Analysis

We disagree with Patent Owner's contention that Steffens's approach is not probabilistic. We agree with Dr. Bajaj that although the similarity measure described by Steffens is not explicitly formulated as calculating a probability, a person of ordinary skill in the art would have found doing so obvious. Ex. 1004 ¶ 144; *see id.* ¶ 125 (citing Ex. 1007, 7:63–67, 5:12–13). We also agree with Dr. Bajaj that a person of ordinary skill in the art would have understood that Steffens's disclosure about the similarity measure (also called a confidence value (Ex. 1007, 13:20–36)), which ranges from 0 to 1, teaches a probability, because "the more similar two vectors are, . . . the closer to 1 the scalar product would be." Ex. 1052 ¶ 42; *see id.* ¶¶ 52–53; Ex. 1007, 13:20–36. Although Dr. Saber testifies that Steffens's similarity measure is not a probability (Ex. 2007 ¶ 48), Dr. Saber does not persuasively explain or rely on persuasive evidence showing why a person of ordinary skill in the art, when recognizing that Steffens's similarity measure is higher for vectors that are more similar and lower for vectors that are less similar, would not have considered the similarity measure as a probability. Therefore, we find the similarity measure of Steffens teaches the claimed "probability."

Further, we find that, contrary to Patent Owner's contention, Petitioner is not replacing Steffens's wavelet based approach with Zhao's contour based approach. Rather, Petitioner and Dr. Bajaj rely on Zhao to explicitly show what is already implicitly described in Steffens, that a person of ordinary skill in the art would have understood that a similarity measure has a probability as taught by Zhao's Bayesian Similarity Measure, and that, in order to detect an object, would have compared the probability to a threshold. Pet. 36, 44–45; Reply 14; Ex. 1004 ¶¶ 125, 144; Ex. 1052 ¶¶ 42, 52–53; Ex. 1006, 64–65; Ex. 1007, 6:43–45 ("If the confidence value falls below a predetermined threshold, the trajectory is deleted."), 13:20–23, 13:35–36. We agree with Petitioner and Dr. Bajaj and find that Zhao shows that a person of ordinary skill in the art would have understood that in disclosing a similarity measure, Steffens discloses a probability, as well as comparing the probability to a threshold. Pet. 36, 44; Ex. 1004 ¶¶ 125, 144; Ex. 1052 ¶ 42. The teachings of Zhao relied on in the Petition, namely, the Bayesian Similarity Measure, are cumulative to the teachings of Steffens. Given that we find that Steffens alone teaches the claimed "probability," we are not persuaded by Patent Owner's contention that Steffens could not be combined with Zhao.

Further, even were we to agree with Patent Owner that Steffens alone does not teach the claimed "probability," we agree with Petitioner and Dr. Bajaj that Steffens discloses a similarity measure where "confidence values . . . indicate the reliability of face detection" and that a person of ordinary skill in the art would have implemented Steffens's similarity measure using the particular similarity measure of Zhao for the benefit of "evaluat[ing] the degree of resemblance" (Ex. 1004 ¶ 125 (citing Ex. 1007, 5:12–13; Ex. 1006, 25)), as well as the benefits of "giv[ing] large similarity measurements within the class while giving small ones between classes," "support[ing] articulation and occlusion," being "robust to noise, deformation, and blur," and being "efficient to compute" (Ex. 1006, 49–50).

We disagree with Patent Owner's contention that Zhao's Bayesian Similarity Measure is incompatible with Steffens's similarity measure because Zhao's process is contour based and Steffens's process is wavelet based. Zhao provides an obvious example of a Bayesian Similarity Measure used in detecting an object. The idea that a designer wanting to implement a similarity measure in Steffens would ignore Zhao because Zhao compares the similarities of contours instead of the similarities of wavelets "makes little sense. A person of ordinary skill is also a person of ordinary creativity, not an automaton." *KSR*, 550 U.S. at 420–21. The obviousness "analysis need not seek our precise teachings" because "a court can take account of the inferences and creative steps that a person of ordinary skill in the art would employ." *Id.* at 418. As discussed above, Petitioner simply contends that using Zhao's Bayesian Similarity Measure to perform the function of Steffens's similarity measure would have been obvious. Petitioner and Dr. Bajaj provided convincing evidence that using a Bayesian Similarity Measure as the similarity measure in Steffens was within the level of ordinary skill in the art. Ex. 1052 ¶¶ 52–57. Thus, to the extent that "Steffens includes no discussion of a probabilistic similarity measure" as alleged by Patent Owner (PO Sur-Reply 31), we find that "substituting of one element [(the similarity measure of Steffens)] for another known in the field [(the Bayesian Similarity Measure of Zhao)]" does no more than yield the predictable result of evaluating the degree of resemblance as taught by Zhao. *KSR*, 550 U.S. at 416.

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*

[1preamble]

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. We find that Petitioner has shown that Steffens teaches the features recited in the preamble of claim 1.[8]

[1a] "Providing a spatio-temporal model for an object to be detected"

Limitation 1a of 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. *Id.* at 39 (citing Ex. 1007, 2:6–13). 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 is not commensurate with the scope of the claim. Patent Owner does not present arguments to the contrary.

---

[8] 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. *See Vivid Techs.*, 200 F.3d at 803.

We find that Petitioner has shown that Steffens teaches limitation 1a.

<u>1[b] “Providing a video comprising a plurality of images”</u>

Limitation 1b of 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 Ex. 1007, 1:14–16, 3:48–53). Patent Owner does not present arguments to the contrary.

We find that Petitioner has shown that Steffens teaches limitation 1b.

<u>[1c] “Measuring the object as a collection of components”</u>

Limitation 1c of 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 present arguments to the contrary.

We find that Petitioner has shown that Steffens teaches limitation 1c.

<u>[1d] “Determining a probability that the object is in each image”</u>

Limitation 1d of 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. *Id.* at 44 (citing Ex. 1004 ¶ 144). In particular, Petitioner contends that a person of ordinary skill would have known that similarity measures may include probabilistic frameworks for identifying objects. *Id.* at 36 (citing Ex. 1004 ¶ 125).

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. *Id.* at 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. *Id.* at 36 (citing Ex. 1006, 25).

Patent Owner presents arguments to the contrary, which we addressed above in our analysis of the reasons to combine the teachings of Zhao and Steffens. PO Resp. 41–43; PO Sur-Reply 31. We disagree with Patent Owner for the reasons given above in our analysis of the reasons to combine.

We find that Steffens alone teaches "determining a probability that the object is in each image" as recited in claim 1 for the reasons given in above. We also find that the combination of Steffens and Zhao teaches "determining a probability that the object is in each image" for the reasons given above.

<u>[1e] "Detecting the object in any image"</u>

Limitation 1e of 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 a person of ordinary skill in the art would have detected an object in an image by comparing probabilities for each image to a threshold as taught by Steffens and Zhao. Pet. 45. Petitioner contends that a person of ordinary skill in the art, in order to detect an object in Steffens's system for detecting objects in video images, would have implemented the similarity measure of Steffens using a known probabilistic measure such as Zhao's Bayesian Similarity Measure, to determine the probability that an object is in each image and to compare the probabilities for each image to a threshold for detecting the object as taught by both Steffens and Zhao. *Id.* (citing Ex. 1004 ¶¶ 146–149).

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. PO Resp. 43. We disagree with Patent Owner for the reasons given in our analysis of claim 1.

Patent Owner contends that Petitioner has not shown that the combination of Steffens and Zhao teaches this limitation under Patent Owner's construction requiring comparing probabilities for each image of the plurality of images to a threshold for detecting the object. *Id.* at 43–45; *see id.* at 18. However, even under Patent Owner's construction, we find that Steffens teaches this limitation. Steffens discloses "the step of detecting an object further includes tracking the object between image frames based on a trajectory associated with the object." Ex. 1007, 2:6–9. Steffens discloses that the "head tracking process . . . generates head position information that may be used to generate head trajectory tracking." *Id.* at 6:25–27. "For every position estimate found for the frame acquired at time t, the algorithm looks . . . for the closest head position estimate that was determined for the previous frame at time t-1 and connects it." *Id.* at 6:34–37. "Every trajectory is assigned a confidence . . . . If the confidence value falls below a predetermined threshold, the trajectory is deleted." *Id.* at 6:42–44. "As shown in Fig. 8, the preselector 16 processes a series of face candidates that belong to the same trajectory as determined by the head tracking process." *Id.* at 7:9–11. Thus, contrary to Patent Owner's contentions, Steffens discloses that the head position information generated from the previous frame at time t-1, which included comparing the probability of the head in the previous frame to a threshold, is used to detect the head in the current frame.

We also agree with Petitioner and Dr. Bajaj that a person of ordinary skill in the art, in order to detect an object in Steffens's system for detecting objects in video images, would have implemented the similarity measure of Steffens using a known probabilistic measure such as Zhao's Bayesian Similarity Measure, to determine the probability that an object is in each image and to compare the probabilities for each image to a threshold for detecting the object in the image as taught by both Steffens and Zhao for the reasons given above in our analysis of the reasons to combine Steffens and Zhao. *See* Pet. 45 (citing Ex. 1004 ¶¶ 146–149). In detecting the object in a current frame, we also find that a person of ordinary skill in the art would have used motion information based on detecting the object in a previous frame as taught by both Steffens and Zhao. Ex. 1006, 91; Ex. 1007, 6:25–48.

We find that Petitioner has shown that Steffens teaches this limitation under Patent Owner's proposed construction. We also find that Petitioner has shown that the combination of Steffens and Zhao teaches this limitation under Patent Owner's proposed construction.

We find that Petitioner has shown, by a preponderance of the evidence, that Steffens renders claim 1 obvious, either alone or in combination with Zhao.

4. Claims 2, 3, 5–7, 10, 11, and 13–15

Patent Owner does not separately argue dependent claims 2, 3, 5–7, 10, 11, and 13–15. Therefore, the dependent claims fall together with the independent claims. *Incept*, 77 F.4th at 1375. On the full record before us, we find that Petitioner has shown, by a preponderance of the evidence, that Zhao renders claims 2, 3, 5–7, 10, 11, and 13–15 obvious.

5. Claims 4 and 12

Claim 4 depends from claim 1 and recites "detecting the object in a current image according to measurements of the object as a collection of components determined from a prior image and a later image relative to the current image." The Petition contends that "[b]y tracking the object and/or its associated nodes 'between image frames' as part of its process for recognizing objects, a [person of ordinary skill in the art] would have known to detect the object according to measurements of the object . . . from a 'prior image' and a 'later image' relative to the current image" and that a person of ordinary skill "would have known that tracking an object . . . 'between frames' would include doing so according to a 'prior image' and a 'later image.'" Pet. 48–49 (citing Ex. 1004 ¶¶ 157–158); *see* Reply 17–18 (citing Ex. 1052 ¶¶ 36–39). Patent Owner contends that "Steffens confirms to [a person of ordinary skill in the art] that its tracking technique operates in only the forward direction." PO Resp. 48 (citing Ex. 2004 ¶ 139); PO Sur-Reply 32 (citing Ex. 2007 ¶ 53).

We agree with Patent Owner for the reasons given by Patent Owner. Neither Petitioner nor Dr. Bajaj has persuasively shown how Steffens's tracking process operates on a current frame, a later frame, and a prior frame to teach "measurements of the object as a collection of components determined from a prior image and a later image relative to the current image" as recited in claim 4. The Petition relies on the contentions presented for claim 4 in contending that claim 12 would have been obvious. Pet. 48–49. We disagree and find that Petitioner has not persuasively shown that Steffens teaches the "detecting" limitation of claim 12.

We find that Petitioner has not shown, by a preponderance of the evidence, that the combination of Steffens and Zhao would have rendered claims 4 and 12 obvious.

6. Claims 8 and 16

Claim 8 depends from claim 1 and recites "a joint probability distribution for the spatio-temporal model with N components is:

$$P\left(X_0^O, X_0^{C_0}, X_0^{C_1}, \ldots, X_0^{C_N}, \ldots, X_T^O, X_T^{C_0}, X_T^{C_1}, \ldots, X_T^{C_N}\right) = \frac{1}{Z}\prod_{ij}\psi_{ij}(X_i^O, X_j^O) \prod_{ik}\psi_{ik}\left(X_i^O, X_i^{C_k}\right)\prod_{ikl}\psi_{kl}\left(X_i^{C_k}, X_i^{C_l}\right)\prod_{i}\phi_i(X_i^O)\prod_{ik}\phi_i\left(X_i^{C_k}\right).$$

Petitioner contends that Zhao teaches this limitation for the reasons given in Petitioner's analysis of the ground based on obviousness over Zhao. Pet. 52. We disagree with Petitioner as discussed in our analysis of the ground based on obviousness over Zhao.

We find that Petitioner has not shown, by a preponderance of the evidence, that the combination of Steffens and Zhao would have rendered claim 8 obvious. Because claim 16 recites a similar limitation, we find that

Petitioner has not shown, by a preponderance of the evidence, that the combination of Steffens and Zhao would have rendered claim 16 obvious.

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." Ex. 1008, 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 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 described 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.

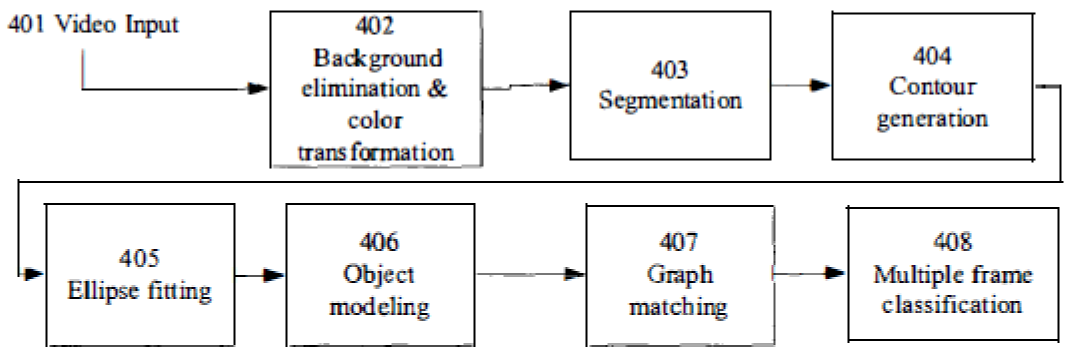


Figure 4 above is a block diagram of Ozer's object detection and activity recognition system. *Id.* at 5:36–38. As shown in Figure 4, in "[b]ackground elimination and color transformation[, t]he first step (402) is the transformation of pixels into another color space regarding to the application." *Id.* 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, 7:38–42.

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 does not present arguments to the contrary.

We find that Petitioner has shown that a person of ordinary skill in the art would have had a reason to combine Ozer and Zhao.

3. Independent Claim 1

[1preamble]

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.

Considering the full record before us, we find that Petitioner has shown that Ozer teaches the features recited in the preamble of claim 1.[9]

[1a] "Providing a spatio-temporal model for an object to be detected"

Limitation 1a of claim 1 recites "providing a spatio-temporal model for an object to be detected." Petitioner contends that Ozer teaches this limitation in disclosing 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 does not present arguments to the contrary.

Considering the full record before us, we find that Petitioner has shown that Ozer teaches limitation 1a.

[1b] "Providing a video comprising a plurality of images"

Limitation 1b of 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 present arguments to the contrary.

---

[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. *See Vivid Techs.*, 200 F.3d at 803.

Considering the full record before us, we find that Petitioner has shown that Ozer teaches limitation 1b.

[1c] "Measuring the object as a collection of components"

Limitation 1c of 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 present arguments to the contrary.

Considering the full record before us, we find that Petitioner has shown that Ozer teaches limitation 1c.

[1d] "Determining a probability that the object is in each image"

Limitation 1d of 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. *Id.* at 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 present arguments to the contrary.

Considering the full record before us, we find that Petitioner has shown that the combination of Ozer and Zhao teaches limitation 1d.

[1e] "Detecting the object in any image"

Limitation 1e of 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 the obviousness ground based on Zhao. 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 the obviousness ground based on Zhao. PO Resp. 50. We disagree with Patent Owner for the reasons given in our analysis of the obviousness ground based on Zhao. We find that Petitioner has shown that the combination of Ozer and Zhao teaches limitation 1e for the reasons given by Petitioner and Dr. Bajaj. Pet. 60–61; Reply 18; Ex. 1004 ¶ 187; Ex. 1052 ¶¶ 58–64.

Considering the full record before us, we find that Petitioner has shown, by a preponderance of the evidence, that the combination of Ozer and Zhao renders claim 1 obvious.

4. Claims 2, 3, 5–7, 10, 11, and 13–15

Patent owner does not separately argue dependent claims 2, 3, 5–7, 10, 11, and 13–15. Therefore, the dependent claims fall together with the independent claims. *Incept*, 77 F.4th at 1375. Having considered the full record before us, we find that Petitioner has shown, by a preponderance of the evidence, that Zhao renders claims 2, 3, 5–7, 10, 11, and 13–15 obvious.

5. *Claims 4 and 12*

Claim 4 depends from claim 1 and recites “detecting the object in a current image according to measurements of the object as a collection of components determined from a prior image and a later image relative to the current image.” The Petition contends that Ozer teaches that movements of object parts are described as a spatio-temporal sequence of feature vectors that include the object’s direction of movement. Pet. 62–63. Petitioner contends that it would have been known to a person of ordinary skill to detect an object according to measurements of the object as a collection of components determined from a prior image and a later image. *Id.* (citing Ex. 1004 ¶¶ 193–194); *see* Reply 19–20 (citing Ex. 1052 ¶¶ 58–64).

Ozer discloses that “marked regions are tracked by using ellipse parameters for the consecutive frames.” Ex. 1008, 7:28–29. Neither Petitioner nor Dr. Bajaj has persuasively shown how Ozer’s tracking process operates on a current frame, a later frame, and a prior frame to teach “measurements of the object as a collection of components determined from a prior image and a later image relative to the current image” as recited in claim 4. The Petition relies on the contentions presented for claim 4 in contending that claim 12 would have been obvious. Pet. 62–63. We disagree and find that Petitioner has not persuasively shown that Zhao teaches the “detecting” limitation of claim 12.

We find that Petitioner has not persuasively shown that the combination of Ozer and Zhao teaches “detecting the object in a current image according to measurements of the object as a collection of components determined from a prior image and a later image relative to the current image” as recited in claim 4. Claim 12 contains a similar limitation.

The Petition relies on the contentions presented for claim 4 in contending that claim 12 would have been obvious. Pet. 29. We disagree and find that Petitioner has not persuasively shown that the combination of Ozer and Zhao teaches the "detecting" limitation of claim 12.

6. Claims 8 and 16

Claim 8 depends from claim 1 and recites "a joint probability distribution for the spatio-temporal model with N components is:

$$P\left(X_0^O, X_0^{C_0}, X_0^{C_1}, \ldots, X_0^{C_N}, \ldots, X_T^O, X_T^{C_0}, X_T^{C_1}, \ldots, X_T^{C_N}\right) = \frac{1}{Z}\prod_{ij} \psi_{ij}(X_i^O, X_j^O)$$

$$\prod_{ik} \psi_{ik}\left(X_i^O, X_i^{C_k}\right)\prod_{ikl} \psi_{kl}\left(X_i^{C_k}, X_i^{C_l}\right)\prod_{i} \phi_i(X_i^O)\prod_{ik} \phi_i\left(X_i^{C_k}\right).$$

Petitioner contends that this limitation would have been obvious for the reasons given in Petitioner's analysis of the obviousness ground based on Zhao. Pet. 65. We disagree with Petitioner for the reasons given in our analysis of the obviousness ground based on Zhao.

We find that Petitioner has not shown, by a preponderance of the evidence, that the combination of Ozer and Zhao would have rendered claim 8 obvious. Because claim 16 recites a similar limitation, we find that Petitioner has not shown, by a preponderance of the evidence, that the combination of Ozer and Zhao would have rendered claim 16 obvious.

E. Claims 1–16 As Obvious Over TLP

The dispositive issue for this Ground is whether TLP qualifies as prior art. The parties agree that the filing date of the provisional application of the

’980 patent is May 27, 2004. Pet. 5; PO Resp. 7. The parties dispute, however, whether the ’980 patent is entitled to the benefit of the filing date of the provisional application. PO Resp. 7; Reply 5–6. The parties also disagree about the date that TLP became publicly accessible. Reply 7; PO Sur-Reply 22–27.

1. Effective Filing Date of the ’980 Patent

Patent Owner contends that the ’980 patent is entitled to the benefit of the filing date of its provisional application. PO Resp. 7. Patent Owner contends that written description support for “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object” recited in claim 1 is found in the provisional application in the Abstract, Section 1 “Introduction,” Section 2.3 “Non-parametric BP (Belief Propagation),” Section 2.4 “AdaBoost Image Likelihoods,” and Section 2.5 “Proposal Process.” *Id.* at 8 (citing Ex. 2004 ¶ 36; Ex. 1003).

Petitioner contends that, under Patent Owner’s proposed construction of this claim limitation, the provisional application does not provide written description support for this claim limitation. Reply 5–6. Petitioner contends that the only discussion about the claimed “threshold” in the provisional application is with respect to AdaBoost, but without further disclosure of “detecting the object in an image upon comparing the probabilities for the image to a threshold for detecting the object” as claimed. *Id.* at 6 (citing Ex. 1003, 9–10).

Patent Owner, relying on the testimony of Dr. Saber, contends that the provisional application provides written description support for “determining a probability that the object is in each image" as claimed in describing that it is customary to consider the object present if

$$h_k(I) \geq \frac{1}{2}\sum_{k=1}^{K} \alpha_k.$$

PO Sur-Reply 21 (citing Ex. 1003, 10; Ex. 2007 ¶ 32). Patent Owner contends that the provisional application provides written description support for "detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object" as claimed in describing sending messages between the set of neighbors of node i as well as between frames based on the local evidence or likelihood associated with the node i and the potential designating the compatibility between the states of node i and j which can be in different frames. *Id.* at 21–22 (citing Ex. 1003, 8–9; Ex. 2007 ¶ 33). In other words, according to Patent Owner, the object is detected based on object probabilities from multiple frames. *Id.* at 22.

We agree with Patent Owner. The equation cited by Patent Owner provides written description support for "determining a probability that the object is in each image" and "comparing the probabilit[y] for [the] image to a threshold." The message sent to the current frame from another frame, which results from comparing the probability that the object is in the other frame to a threshold, provides written description support for "comparing the probabilit[y] for [another] image to a threshold." We find that the '980 patent is entitled to the benefit of the filing date of the provisional application, which is May 27, 2004. Therefore, to qualify as prior art, Petitioner must show that TLP was publicly available before May 27, 2004.

2. *Date of Public Accessibility of TLP*

Petitioner contends that evidence of record shows that TLP was publicly available on April 9, 2004. Reply 7 (citing Ex. 1004 ¶¶ 65–66). Dr. Bajaj testifies that TLP was presented at the 2004 Conference on Computer Vision and Pattern Recognition. Ex. 1004 ¶ 66. Dr. Bajaj testifies that the deadline for submitting the manuscript for TLP was April 9, 2004. *Id.* Dr. Bajaj testifies that "TLP was widely disseminated and made publicly available through the submission process, and no later than April 9, 2004." *Id.*

Petitioner contends that evidence of record shows that TLP was publicly available on January 31, 2004. Reply 7 (citing Ex. 1050 ¶¶ 40–48). Dr. Hall-Ellis testifies that a co-author of TLP, Michael J. Black, posted TLP on ResearchGate, a public website, and that the ResearchGate website indicates a publication date of TLP as January 2004. Ex. 1050 ¶¶ 41, 45, 47; *see* Ex. 1049. Dr. Hall-Ellis testifies that she examined entries to ResearchGate posted by Michael J. Black from 2003 to 2005. Ex. 1050 ¶ 47. Dr. Hall-Ellis testifies that the publication date of each entry closely aligned with an official publication date for the document. *Id.* Dr. Hall-Ellis testifies that in her opinion, TLP was publicly accessible on January 31, 2004. *Id.* ¶ 48.

Patent Owner, relying on the testimony of Dr. Saber, contends that public disclosure of TLP during the review process would have been unethical. PO Sur-Reply 27 (citing Ex. 2007 ¶ 63). Patent Owner contends that the evidence shows no intent to publicize TLP during pre-publication review, and that TLP was not accessible to anyone other than the peer-review committee, thus further suggesting an absence of actual public accessibility before May 27, 2004. *Id.* (citing Ex. 2007 ¶ 64).

Patent Owner contends that Dr. Hall-Ellis (a) confirmed that ResearchGate does not identify "Jan 2004" as a publication date, (b) did not have personal knowledge about how "Jan 2004" came to appear on the ResearchGate website, and (c) "would never take these [ResearchGate dates] solely on face value." *Id.* at 24–25 (citing Ex. 2011, 19:2–22:8, 48:6–59:13) (alteration in original). Patent Owner contends that even if Michael J. Black did indicate the "Jan 2004" date on the ResearchGate website, he did so on March 2, 2014, nearly a decade after the relevant time. *Id.* at 24. Patent Owner contends that there is no evidence that Dr. Black meant "Jan 2004" as a publication date, or that he knew what a printed publication is under the patent laws. *Id.*

We agree with Patent Owner for the reasons given by Patent Owner. We find that submitting TLP for peer review on April 9, 2004, does not show that TLP was accessible to anyone other than the peer-review committee. We find that the "Jan 2004" date on the ResearchGate website was, at best, entered by Dr. Black in 2014 without indicating that Dr. Black considered "Jan 2004" to be a publication date as opposed to a copyright date or another kind of date. Therefore, we find that Petitioner has not shown that TLP is prior art to the '980 patent.

We find that Petitioner has not shown, by a preponderance of the evidence, that TLP would have rendered claims 1–16 obvious.

VIII. CONCLUSION

In summary, we determine that a preponderance of the evidence establishes that claims 1–3, 5–7, 9–11, and 13–15 are unpatentable, but does not establish that claims 4, 8, 12, 16 are unpatentable, as shown in the following table:[10]

Claim(s)	35 U.S.C. §	Reference(s)/ Basis	Claims Shown Unpatentable	Claims Not Shown Unpatentable
1–16	103(a)	Zhao	1–3, 5–7, 9–11, 13–15	4, 8, 12, 16
1–16	103(a)	Steffens, Zhao	1–3, 5–7, 9–11, 13–15	4, 8, 12, 16
1–16	103(a)	Ozer, Zhao	1–3, 5–7, 9–11, 13–15	4, 8, 12, 16
1–16	103(a)	TLP		1–16
Overall Outcome			1–3, 5–7, 9–11, 13–15	4, 8, 12, 16

[10] Should Patent Owner wish to pursue amendment of the challenged claims in a reissue or reexamination proceeding subsequent to the issuance of this decision, we draw Patent Owner's attention to the April 2019 *Notice Regarding Options for Amendments by Patent Owner Through Reissue or Reexamination During a Pending AIA Trial Proceeding*. *See* 84 Fed. Reg. 16,654 (Apr. 22, 2019). If Patent Owner chooses to file a reissue application or a request for reexamination of the challenged patent, we remind Patent Owner of its continuing obligation to notify the Board of any such related matters in updated mandatory notices. *See* 37 C.F.R. § 42.8(a)(3), (b)(2).

IX. ORDER

In consideration of the foregoing, it is hereby

ORDERED that claims 1–3, 5–7, 9–11, and 13–15 of the ’980 patent have been proven by a preponderance of evidence to be unpatentable;

FURTHER ORDERED that claims 4, 8, 12, 16 of the ’980 patent have not been proven by a preponderance of evidence to be unpatentable; and

FURTHER ORDERED that, because this is a final written decision, parties to this proceeding seeking judicial review of the Decision must comply with the notice and service requirements of 37 C.F.R. § 90.2.

MCKONE, *Administrative Patent Judge*, dissenting

I join the Majority's determination that Petitioner has not shown that TLP is prior art to the '980 patent and, thus, that Petitioner has not shown, by a preponderance of the evidence, that claims 1–16 would have been obvious over TLP. However, I disagree that Zhao teaches claim limitation 1e, "detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object," or the similar limitation of independent claim 9. Accordingly, I would not find that Petitioner has shown that any claim of the '980 patent is unpatentable under Petitioner's remaining grounds, obviousness over Zhao, Steffens and Zhao, or Ozer and Zhao. Therefore, I dissent-in-part.

First, I would construe "comparing the probabilities for each image to a threshold for detecting the object" as it is written, namely, that to detect an object in a given image, each image of a plurality of images must be compared to the threshold, or, as Petitioner would phrase it (Reply 8), "comparing multiple probabilities from multiple images to a threshold."[11] For example, if a plurality of images consists of two images, in order to detect the object in the second image, the probability for detecting the object in the first image must be compared to the threshold and the probability for detecting the object in the second image must also be compared to the

[11] I do not view this as a new construction proposed by Patent Owner in the Patent Owner Response, as the Majority does. Rather, Patent Owner raised this construction in the Preliminary Response, at 42–43, and the Institution Decision, at 26–28, acknowledged, but did not resolve, Patent Owner's argument. Thus, Petitioner's new Reply arguments were precipitated not by Patent Owner's arguments but instead to support an alternative theory advanced in the Institution Decision assuming Patent Owner's claim construction to be correct.

threshold. Since the Majority does not reach the construction of this term (rather, finds that Zhao teaches the limitation either way), I will not belabor this point. However, I would find support in the Specification, at Exhibit 1001, 6:65–8:8, and in particular 7:6–20 and 7:45–55. Contrary to Petitioner's argument that claim limitation 1e includes w=1 (Reply 1–2), I would not read the case of w=1 to be an example of the invention; rather, it is the degenerate case of the algorithm when a comparison is made with no temporal information. Ex. 1001, 7:20–22. I also would not read the text in Figure 5 to contradict (or override) the text of the Specification at Exhibit 1001, 7:45–55, contrary to Petitioner's reading. *See* Reply 3 (arguing that Figure five, box 505, recites "an image" and "the image").

Second, I disagree with the Majority that Zhao teaches "comparing the probabilities for each image to a threshold for detecting the object," as its algorithm detects an object by comparing only one image to a threshold. To the extent Zhao provides enough detail to discern how its motion information is used, that motion information is used only to disambiguate segments of an image already detected. Zhao summarizes its algorithm as follows:

> The body parts of a person are located in a coarse-to-fine manner using the RCR algorithm. First, the person is segmented from the background . . . through background subtraction . . . . Second, the segmented region is decomposed into ribbons and these ribbons are matched with the modeled body parts including the extended parts . . . . 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 . . . . Fourth, the predicted outlines of the body parts are aligned with the edge features . . . .

Ex. 1006, 91. Each of these steps is conducted on a single image, starting with a course identification of the object in an image and continuing with recursive (the first "R" in RCR) refinement of the object in the same image.

Chapter 4 of Zhao describes the RCR algorithm in more detail. *Id.* at 74–80. There is no description of using motion information or other temporal information in detecting an image in a frame in Zhao's detailed description of the RCR algorithm. *Id.* Rather, Zhao describes detecting an image by aligning predicted body parts with edge features detected in an image and comparing the alignment to a threshold. *Id.* at 76–77. The predicted body parts are estimated from human models and, since the process is iterative, body parts in the image already identified. *Id.* at 60–62. Zhao does not describe using motion or any other temporal information to predict body parts. *Id.* The Bayesian similarity measurements underlying the iterative comparisons are described in more detail in Zhao's Chapter 3. *Id.* at 54–60. As Dr. Saber confirms, Zhao's description of Bayesian similarity measurements does not describe using motion or any other temporal information in its comparisons. Ex. 1051, 173:5–8 ("In Zhou, the use of Bayes, Bayes' theorem, is done in the spatial manner. There is no temporal discussion or any formulation that talks about the use of Bayes in a temporal manner."). Dr. Saber confirms that Zhao's RCR algorithm operates on a single image and does not detect an object in an image based on the probability of the object being detected in a previous image.
Ex. 2004 ¶¶ 107–111; Ex. 2007 ¶¶ 36–38.

The Petition characterizes Zhao's algorithm as

> Begin[ning] with a simple decision rule to detect a human that compares the probability of an extracted contour matching a person with a threshold: "the contour C corresponds to a person if
>
> $$P(person|C) > threshold.$$
>
> Otherwise, C is not a person." By applying Bayes' formula and conditioning the hypothesis for body part identification, the decision rule is reformulated as: "the contour C corresponds to a person if
>
> $$P(C|H^*,person)P(person|H^*) > threshold$$

Pet. 27 (citing Ex. 1006, 64–65; Ex. 1004 ¶¶ 98–99). The Petition does not allege that Zhao's RCR algorithm takes into account a probability determination for a previous frame, and Dr. Bajaj's testimony makes no mention of it. Pet. 27; Ex. 1004 ¶¶ 98–99.

The Institution Decision introduced a theory of Zhao in which

> 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. 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.

Dec. 27–28 (citing Ex. 1006, 91, 93, Fig. 5.5 Ex. 1004 ¶¶ 79–82, 87). Here, the Institution Decision refers to description in Zhao that

> 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.

Ex. 1006, 91. Dr. Bajaj testifies that this use of motion information can help disambiguate body parts and, thus, is evidence that Zhao uses a spatio-temporal model. Ex. 1004 ¶¶ 79–82, 87. But Dr. Bajaj does not testify in detail as to how such motion information might be used with Zhao's algorithm and does not testify that Zhao uses a comparison of other images to a threshold when comparing a current image to the threshold. *Id.* Dr. Saber testifies that Zhao does not describe any details of how his algorithm uses motion. Ex. 1051, 170:3–172:1.

In the Reply, Petitioner (Reply 10) cites to disclosure in Zhao that

> Motion information obtained from previous frames can be used to predict the orientations of the limbs and to solve the ambiguity. Fig. 5.6 presents a full cycle of a walking person. In 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.

Ex. 1006, 91–92. According to Petitioner, a skilled artisan "would understand that if the cycle of a person walking involves prediction from previous frames, not only is there a temporal component, the RCR algorithm contour updating procedure and prediction procedure would contain comparison of that probability in the previous frame to a detection threshold." Pet. 10–11. The Majority adopts this theory.

Patent Owner cites no evidentiary support for this "understand[ing]." Dr. Bajaj does not weigh in on this issue.[12] Zhao itself does not support this theory. It states little more than "motion information obtained from previous frames can be used to predict the orientations of the limbs and to solve the ambiguity." Ex. 1006, 91. This does not explain how the motion information is generated or used. Dr. Saber, on the other hand, testifies that a skilled artisan "would have understood that such 'motion information' does not include (nor is it based on) the probability of the object (the person) being present in another image or comparing that probability to a detection threshold." Ex. 2004 ¶ 112 (citing Ex. 1006, 91); *see also* Ex. 2007 ¶ 38; Ex. 1051, 179:6–180:21.[13] Dr. Saber's testimony is the only evidence we have on this point and it refutes, rather than supports, the theory espoused by Petitioner and the Majority.

The Majority reasons that, in Zhao's RCR algorithm, an object in a previous frame is detected by comparison of a probability to a threshold, motion information is determined from the detected object (thus, contains prior knowledge of the identification of the object in a previous frame), and the RCR algorithm uses the motion information to align the contours of the detected object in the current frame to the object model. Perhaps this is how

---

[12] Petitioner cites generally to about 30 paragraphs of Dr. Bajaj's Supplemental Declaration. Reply 9–11 (citing Ex. 1052 ¶¶ 6–35). This testimony, however, is in support of Petitioner's argument that claim limitation 1e does not require comparing multiple probabilities from multiple images to a threshold. This testimony does not mention Zhao.
[13] Petitioner cites generally to about 80 pages of cross-examination testimony of Dr. Saber. Reply 11 (citing Ex. 1051, 167–251). Here, Dr. Saber repeatedly testifies that Zhao's mention of motion does not teach comparing multiple probabilities from multiple images to a threshold.

Zhao's algorithm works, but the evidence is far from clear. Zhao's detailed description of the RCR algorithm is silent on the use of motion. Ex. 1006, 70–80.[14] Dr. Bajaj does not provide testimony that would support the Majority's inference. Dr. Saber testifies that "the RCR algorithm does not detect a human in the image under consideration based on the probability of the human being detected in **other** images." Ex. 2004 ¶ 109; *see also id.* ¶¶ 110 ("Nor does the RCR algorithm involve determining or considering the probability of the object (a person) being present in another image or comparing that probability to a detection threshold."), 111 ("The contour updating and alignment operations of the RCR algorithm (Steps 5 and 6) operate on the same image as steps 1–4 (not different images) and likewise do not involve the probability of the object (the person) being present in another image or comparing that probability to a detection threshold."), 112 ("While Zhao states that '[m]otion information obtained from previous frames can be used to predict the orientations of the limbs and to solve the ambiguity' when limbs are misaligned or confused, a [person of ordinary skill in the art] would have understood that such 'motion information' does not include (nor is it based on) the probability of the object (the person)

[14] Zhao states that "Because of cluttered backgrounds, the alignment may be distracted by other objects. To avoid such situations, other cues such as stereo, motion, and the intensity pattern can be used to constrain the search of the body parts to be within the region of similar attributes. For example in Fig. 4.3(1), the search for the arms is constrained to be within a region having similar disparity as the torso region." Ex. 1006, 77. This does not appear to be a description of the use of information from a previous frame in evaluating a current frame, and neither party argues that it is. Rather, it appears to suggest making inferences based on how the object, such as a human, is expected to behave.

being present in another image or comparing that probability to a detection threshold.”). Dr. Saber was consistent with this testimony on cross-examination. Ex. 1051, 167–251. When I weigh the evidence, I find it to be one-sided, with all of the expert testimony weighing against Petitioner or avoiding the issue altogether. Thus, I would not find that Petitioner has met its burden to prove, by a preponderance of the evidence, that Zhao teaches claim limitation 1e.

Moreover, even if the Majority is correct about how Zhao's RCR algorithm works, all that shows is that the detection of object presence in a previous image can be used to disambiguate segments (e.g., limbs) of an already detected object (e.g., a person), not that it can be used to detect the object in the image in the first place, as required by the claim language. For this additional reason, I would not find that Petitioner has met its burden.

As to the combination of Ozer and Zhao, the Majority finds that claim limitation 1e is taught by Zhao. I disagree for the reasons given above. Thus, I would not find that Petitioner has met its burden to prove, by a preponderance of the evidence, that the combination of Zhao and Ozer teaches claim limitation 1e.

As to the combination of Steffens and Zhao, Petitioner argues, as to claim limitation 1e, that a skilled artisan would have combined Zhao's probabilistic framework with Steffens's teachings and that Zhao teaches comparing the probabilities for each image to a threshold for detecting an object. Pet. 45; Reply 16–17 (“For the reasons already explained as to Ground I, the combination of Zhao and Steffens renders obvious element 1/9[e].”); Ex. 1004 ¶ 148. The Majority, however, finds that Steffens, not Zhao, teaches claim limitation 1e. This is not the theory advanced by

Petitioner, either in the Petition or the Reply. Pet. 45; Reply 16–17. This also is not the theory presented in the Institution Decision. Rather, the Institution Decision noted that "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," and agreed with that argument "for the reasons given in our analysis of claim 1." Dec. 39.

I would not depart from the theory presented in the Petition and construct a new theory for the first time in the Final Written Decision. *See SAS Inst., Inc. v. Iancu*, 584 U.S. 357, 358 (2018) ("[T]he petitioner's petition, not the Director's discretion, should guide the life of the litigation."); *Axonics, Inc. v. Medtronic, Inc.*, 75 F.4th 1374, 1380 (Fed. Cir. 2023) ("An IPR is an expedited administrative procedure, driven by the invalidity theories presented in a petition."). This is unfair to Patent Owner, who does not have an opportunity to respond or present opposing evidence. *See Axonics*, 75 F.4th at 1381 ("In a formal adjudication under the APA, such as an IPR proceeding, the Board must inform the parties of 'the matters of fact and law asserted,' 5 U.S.C. § 554(b)(3), and 'give all interested parties opportunity for . . . the submission and consideration of facts [and] arguments,' *id.* § 554(c). The Board must also permit parties 'to submit rebuttal evidence, and to conduct such cross-examination as may be required for a full and true disclosure of the facts.' *Id.* § 556(d)."). Rather, I would evaluate the theory Petitioner presented and, for the reasons give above, I would find that Petitioner has not met its burden to show that Zhao teaches claim limitation 1e.

For the foregoing reasons, I dissent.

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

Jennifer R. Bush
FENWICK & WEST LLP
jbush@fenwick.com

FOR PATENT OWNER:

Christine E. Lehman
Michael Matulewicz-Crowley
Philip J. Eklem
Jaime F. Cardenas-Navia
REICHMAN JORGENSEN LEHMAN
& FELDBERG LLP
clehman@reichmanjorgensen.com
mmatulewicz-crowley@reichmanjorgensen.com
peklem@reichmanjorgensen.com
jcardenas-navia@reichmanjorgensen.com
RJ_VideoLabs-Facebook@reichmanjorgensen.com