

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.

Petitioners

v.

VL COLLECTIVE IP LLC

Patent Owner

Case IPR2023-00924
U.S. Patent 7,436,980

PATENT OWNER'S RESPONSE

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	THE '980 PATENT	1
A.	Overview of the '980 Patent.....	1
B.	Priority Date	7
III.	THE ASSERTED REFERENCES	10
A.	Zhao (EX1006).....	10
B.	Steffens (EX1007).....	13
C.	Ozer (EX1008)	15
D.	TLP (EX1009).....	16
IV.	LEVEL OF ORDINARY SKILL	18
V.	CLAIM CONSTRUCTION	18
VI.	PETITIONERS HAVE NOT DEMONSTRATED THAT ANY CHALLENGED CLAIM IS UNPATENTABLE	25
A.	The challenged claims are patentable over Zhao (Ground I).....	26
1.	Claims 1/9	26
2.	Claims 4/12	34
3.	Claims 8/16	38
4.	All dependent claims are patentable over Zhao.....	41
B.	The challenged claims are patentable over Steffens in view of Zhao (Ground II).	41
1.	Claims 1/9	41
2.	Claims 4/12	46
3.	Claims 8/16	49

4.	All dependent claims are patentable over Steffens in view of Zhao.	49
C.	The challenged claims are patentable over Ozer in view of Zhao (Ground III).	50
1.	Claims 1/9	50
2.	Claims 4/12	50
3.	Claims 8/16	53
4.	All dependent claims are patentable over Ozer in view of Zhao.....	54
D.	The challenged claims are patentable over TLP (Ground IV).	54
1.	Petitioners fail to demonstrate that TLP qualifies as prior art.....	54
2.	Claims 1/9	58
3.	Claims 3/11	64
4.	Claims 8/16	65
5.	All dependent claims are patentable over TLP.....	66
VII.	CONCLUSION.....	66

TABLE OF AUTHORITIES

Cases

<i>Abbott Labs. v. Baxter Pharm. Prods., Inc.</i> , 334 F.3d 1274 (Fed. Cir. 2003)	58
<i>Ariad Pharms., Inc. v. Eli Lilly & Co.</i> , 598 F.3d 1336 (Fed. Cir. 2010)	8
<i>Ashworth Bros., Inc. v. Laitram, L.L.C.</i> , IPR2020-00594, Paper 60 (PTAB Aug. 23, 2021).....	passim
<i>Cent. Admixture Pharmacy Servs., Inc. v. Advanced Cardiac Sols., P.C.</i> , 482 F.3d 1347 (Fed. Cir. 2007)	22
<i>Cisco Sys., Inc. v. C-Cation Techs., LLC</i> , IPR2014-00454, Paper 12 (PTAB Aug. 29, 2014).....	35
<i>Digital Biometrics, Inc. v. Identix, Inc.</i> , 149 F.3d 1335 (Fed. Cir. 1998)	22
<i>In re Stepan Co.</i> , 868 F.3d 1342 (Fed. Cir. 2017)	38
<i>In re Van Os</i> , 844 F.3d 1359 (Fed. Cir. 2017)	passim
<i>Resonate Inc. v. Alteon Websystems, Inc.</i> , 338 F.3d 1360 (Fed. Cir. 2003)	22
<i>Resqnet.com, Inc. v. Lansa, Inc.</i> , 346 F.3d 1374 (Fed. Cir. 2003)	19
<i>ResQNet.com, Inc. v. Lansa, Inc.</i> , 594 F.3d 860 (Fed. Cir. 2010)	57
<i>Sanofi-Synthelabo v. Apotex, Inc.</i> , 550 F.3d 1075 (Fed. Cir. 2008)	passim
<i>Sanofi-Synthelabo</i> , 550 F.3d at 1086	37

<i>Seal-Flex, Inc. v. Athletic Track and Court Const.</i> , 172 F.3d 836 (Fed. Cir. 1999)	21
<i>SIMO Holdings Inc. v. Hong Kong uCloudlink Network Tech. Ltd.</i> , 983 F.3d 1367 (Fed. Cir. 2021)	19
<i>Sjolund v. Musland</i> , 847 F.2d 1573 (Fed. Cir. 1988)	23
<i>SRI Int’l, Inc. v. Internet Sec. Sys.</i> , 511 F.3d 1186 (Fed. Cir. 2008)	56, 57
<i>Stamps.com Inc. v. Endicia, Inc.</i> , 437 F. App’x 897 (Fed. Cir. 2011)	58
<i>Trading Techs. Int’l, Inc. v. eSpeed, Inc.</i> , 595 F.3d 1340 (Fed. Cir. 2010)	8
<i>Univ. of Strathclyde v. Clear-Vu Lighting LLC</i> , 17 F.4th 155 (Fed. Cir. 2021)	passim
<i>W.L. Gore & Assocs., Inc. v. Garlock, Inc.</i> , 721 F.2d 1540 (Fed. Cir. 1983)	25, 28
<i>York Prod., Inc. v. Cent. Tractor Farm & Fam. Ctr.</i> , 99 F.3d 1568 (Fed. Cir. 1996)	19
 Regulations	
37 C.F.R. § 42.6(a)(3)	35

PATENT OWNER'S EXHIBIT LIST

Exhibit	Description
2001	Declaration of Jaime F. Cardenas-Navia in Support of Motion for Admission Pro Hac Vice
2002	Zou et al., Object Detection in 20 Years - Survey
2003	Email dated December 29, 2023
2004	Declaration of Eli Saber in Support of Patent Owner's Response
2005	Curriculum Vitae of Eli Saber, Ph.D.
2006	Deposition of Dr. Chandrajit Bajaj

I. INTRODUCTION

Petitioners have not met their burden of demonstrating that any challenged claim is obvious. Each of Petitioners' four asserted Grounds falls flat because each fails to show that every limitation of element 1/9[e] in independent claims 1 and 9 was disclosed or rendered obvious by the asserted references. Element 1/9[e] is critical because its limitations, when properly considered in full and in the context of the preceding elements, capture key aspects of the purpose and function of the invention as described in the '980 patent's specification—aspects not disclosed or rendered obvious by any of Petitioners' asserted references. But the law is clear that Petitioners cannot prevail if they have not demonstrated that each claim—considered as a whole, with all of its limitations—would have been obvious over the asserted prior art. Because Petitioners have not, the Board should issue a final written decision finding all challenged claims patentable.

Petitioners' Ground IV is fatally flawed for the independent reason that the sole reference upon which it is based (“TLP”) has not been shown to qualify as prior art to the '980 patent. Under controlling precedent, the arguments and evidence with which Petitioners claim TLP constitutes prior art is insufficient.

II. THE '980 PATENT

A. Overview of the '980 Patent

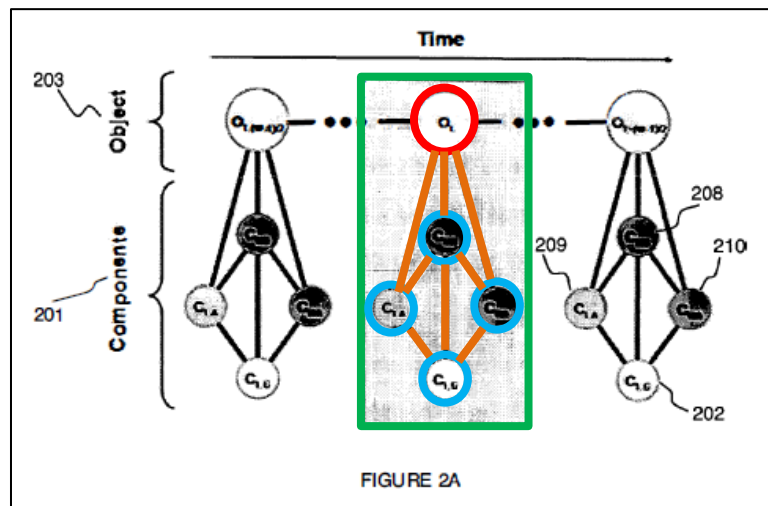
Object detection is an important problem related to computer vision

applications. EX1001, 1:15-21. While the '980 patent suggests it may have been relatively “common to represent objects as collections of features” for detecting them, there was “a large variation in how many features are needed and how these features are detected and represented.” EX1001, 1:15-21. The inventors recognized a need for a new way of “implementing *component detection* and performing inference *over space and time*.” EX1001, 1:22-24.¹ Their solution was “a probabilistic framework for automatic component-based detection and tracking of objects in images and/or video [that] combines object detection with tracking in a unified framework,” where “[o]bject detection *considers the consistency of the detection over time*.” EX1001, 2:43-54.

Under their framework, object detection is formulated as an inference in a spatio-temporal graphical model. EX1001, 2:55-3:2; EX2006, 67:5-10, 44:15-19, 64:2-5. For example, the patent teaches that “object detection and tracking is formulated as an inference in 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**.” EX1001, 2:55-3:2, Fig. 2A (annotated below). “**For a single frame, objects are represented using a two-layer spatial graphical model**,” where “[t]he fine, component, layer 201 includes a set of loosely connected

¹ All emphasis added in this Response unless otherwise noted.

parts, e.g., 202,” and “[t]he course, object, layer 203 corresponds to **an entire appearance model of the object** and is **connected to all constituent components**, e.g., 202.” EX1001, 4:18-24; EX2006, 82:14-83:5. The patent explicitly models the **object** as a distinct entity from its corresponding **components** – in the same graph and for each point in time. EX2004, ¶48.



EX1001, Fig. 2A (annotated)

In the patent’s graphical object model, “[d]irected edges between nodes represent learned **spatial [probabilistic constraints]** and **temporal probabilistic constraints**.” EX1001, 2:59-60 (annotated). The “edges” are lines that connect nodes to each other, which correspond to spatial or temporal constraints. EX1001, 2:59-60, 4:15-17; EX2004, ¶49. “To integrate **temporal constraints** the spatial graphical models are extended over time to an arbitrary length temporal window,” and “[t]he resulting **spatio-temporal graphical models** are shown in FIGS. 2A and

2B.” EX1001, 4:34-45. The patent explicitly models the **object** as a distinct entity from its corresponding **components** at multiple points in time. See EX1001, 4:37-42; EX2004, ¶49.

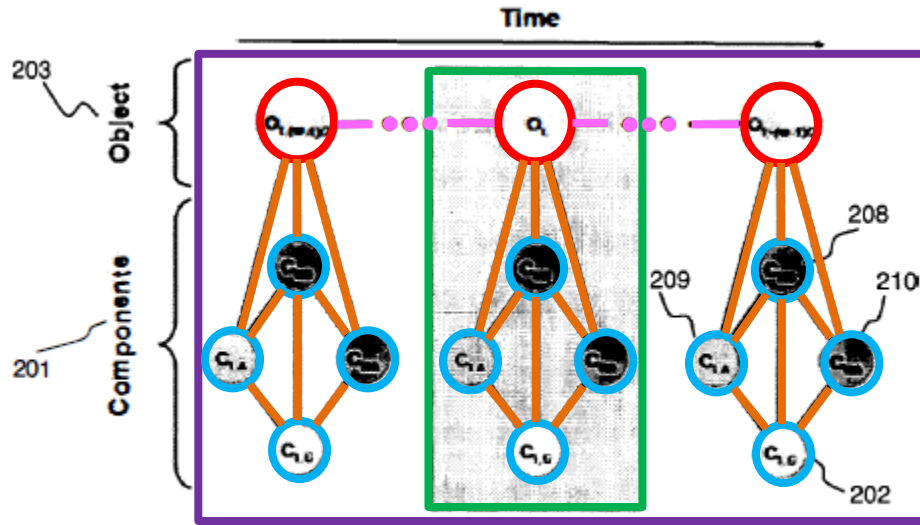
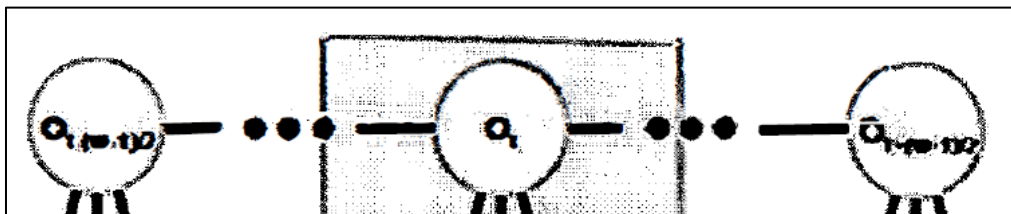


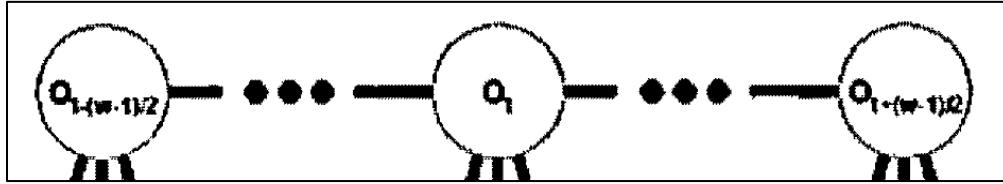
FIGURE 2A

EX1001, Fig. 2A (annotated)

The patent further explains that “[a]n object... is modeled as a spatio-temporal directed graphical model,” where “[e]ach node in the graph represents either the object or a component of the object at time t .” EX1001, 3:47-53. For example, “an object at a current time is denoted as O_t with an earlier model denoted as $O_{t-(w-1)/2}$ and a later model denoted as $O_{t+(w-1)/2}$.” EX1001, 4:45-48; EX2004, ¶50.



EX1001, Fig. 2A (excerpted)



EX1003, Fig. 2(d) (excerpted)

An exemplary “joint probability distribution” for the spatio-temporal model explicitly represents object-level nodes and component-level nodes, as well as object-to-object temporal relationships, object-to-component spatial relationships, and component-to-component spatial relationships. EX1001, 3:53-4:8; EX2004, ¶51. Those relationships correspond to “edges” in Figures 2A and 2B. EX1001, 4:49-51, 4:15-17, 2:59-60; EX2004, ¶51.

$$P(X_0^O, X_0^{C_0}, X_0^{C_1}, \dots, X_0^{C_N}, \dots, X_T^O, X_T^{C_0}, X_T^{C_1}, \dots, X_T^{C_N}) = \frac{1}{Z} \prod_{ij} \psi_{ij}(X_i^O, X_j^O) \prod_{ik} \psi_{ik}(X_i^O, X_i^{C_k}) \prod_{ikt} \psi_{kt}(X_i^{C_k}, X_i^{C_t}) \prod_i \phi_i(X_i^O) \prod_{ik} \phi_i(X_i^{C_k})$$

EX1001, 3:55-65

The patent teaches determining a probability that the **object** (as a distinct entity from its components) is in each of a plurality of images. EX2004, ¶52; *see also id.* ¶¶39-42. For example, the terms “ $\phi_i(X_i^O)$ and $\phi_i(X_i^{C_k})$ denote the local evidence for the object and component states respectively,” which is based on the probability that

the object or component is in the image. EX1001, 3:53-4:8; EX2004, ¶57. “An object is considered present” if the probability (or “confidence”) that the object is present is greater than or equal to a threshold $h(I) \geq \frac{1}{2} \sum_{k=1}^K \alpha_k$. EX1001, 5:25-6:13; EX2004, ¶57.

The patent teaches that detection of an object in any given image is “formulated as an inference in” the above-described spatio-temporal model, EX1001, 2:55-3:2, 3:47-5:24; EX2004, ¶58, for example, using “non-parametric belief propagation.” EX1001, 2:55-3:2, 3:47-5:24. In this example, the inference can be performed using a “message passing” technique “between nodes to propagate information about what states neighboring nodes would be.” EX1001, 6:22-7:5, Fig. 4; EX2004, ¶59. A message $m_{ij}(X_j)$ between two nodes i and j may be formulated as follows:

$$m_{ij}(X_j) = \int \psi_{ij}(X_i, X_j) \phi_i(X_i) \prod_{k \in \{A_i \setminus j\}} m_{ki}(X_i) dX_i$$

where “ $\phi_i(X_i)$ is the local evidence (or likelihood) associated with the node i, and $\psi_{ij}(X_i, X_j)$ is the compatibility potential between the states of node i and j.” EX1001, 6:50-64; EX2004, ¶59. Accordingly, messages between object-level nodes in different frames are based on $\phi_i(X_i^O)$, which is based on the probability that the object is in the image, as well as $\psi_{ij}(X_i^O, X_j^O)$, which is the temporal compatibility between object nodes in two different frames i and j. EX1001, 6:22-64, 5:25-6:13,

3:47-4:8; EX2004, ¶59. The “messages” can be approximated using a “particle set” formed by producing “a set of detection results that score above the $\frac{1}{2} \sum_{k=1}^K \alpha_k$ threshold.” EX1001, 6:29-37, 7:23-44; EX2004, ¶60.

The patent teaches that inference may be performed over “the entire image sequence” or over fewer images using “[a] w-frame windowed smoothing algorithm” where “w is an odd integer ≥ 1 .” EX1001, 7:6-22. Accordingly, the “inference” for detecting an object in any given image is based on determinations of whether an object is present in multiple images. EX2004, ¶61. This provides “temporal integration” and “temporal consistency,” which “produce[s] more consistent results over time.” EX1001, 7:6-22; EX2004, ¶57.

B. Priority Date

Petitioners and their expert do not dispute that the claims of the ’980 patent are entitled to the patent’s priority date of May 27, 2004.” Paper 3 (“Petition”) at 5; EX1004, ¶33; Paper 11 at 48. The ’980 patent claims priority to U.S. Provisional Application No. 60/574,799, filed on May 27, 2004, and incorporates its content by reference. EX1001, 1:4-6. The table below compares the claims to exemplary disclosures in the provisional application, which demonstrates, as Dr. Saber explains, that the provisional application reasonably conveys to the POSITA that the inventors had possession of the claimed subject matter at the time the provisional application was filed. *See Trading Techs. Int’l, Inc. v. eSpeed, Inc.*, 595 F.3d 1340,

1359 (Fed. Cir. 2010); *Ariad Pharms., Inc. v. Eli Lilly & Co.*, 598 F.3d 1336, 1351 (Fed. Cir. 2010); EX2024, ¶ 30.

Claim Element	Exemplary Disclosures in U.S. Prov. App. No. 60/574,799
1/9[pre]	EX1003 at 1 (Abstract), 1-2 (section “1 Introduction”). EX2024, ¶31.
1/9[a]	EX1003 at 1 (Abstract), 1-2 (section “1.1 Related Work”), 5-6 (section “2 Graphical Object Models”), 6 (“section “2.1 Building a Graphical Model”), 7 (Figure 2 and caption). EX2024, ¶32.
1/9[b]	EX1003 at 1 (Abstract), 1-2 (section “1 Introduction”), 10-11 (section “3 Experiments”). EX2024, ¶33.
1/9[c]	EX1003 at 1 (Abstract), 1-2 (section “1 Introduction”), 5-6 (section “2 Graphical Object Models”), 6 (“section “2.1 Building a Graphical Model”), (section “2.3 Non-parametric BP). EX2024, ¶34.
1/9[d]	EX1003 at 1 (Abstract), 1-2 (section “1 Introduction”), 8-9 (section “2.3 Non-parametric belief propagation”); 9-10 (section “2.4 AdaBoost Image Likelihoods”), 10 (section “2.5 Proposal Process”). EX2024, ¶35.

1/9[e]	EX1003 at 1 (Abstract), 1-2 (section “1 Introduction”), 8-9 (section “2.3 Non-parametric belief propagation”); 9-10 (section “2.4 AdaBoost Image Likelihoods”), 10 (section “2.5 Proposal Process”). EX2024, ¶36.
2/10	EX1003 at 1 (Abstract), 1-2 (section “1 Introduction”), 8-9 (section “2.3 Non-parametric belief propagation”); 9-10 (section “2.4 AdaBoost Image Likelihoods”), 10 (section “2.5 Proposal Process”). EX2024, ¶37.
3/11	EX1003 at 1 (Abstract), 1-2 (section “1 Introduction”), 5-6 (section “2 Graphical Object Models”), 6 (“section “2.1 Building a Graphical Model”), 7 (Figure 2 and caption), 8-9 (section “2.3 Non-parametric belief propagation”). EX2024, ¶38.
4/12	EX1003 at 5 (section “1.1 Related work”), 6 (“section “2.1 Building a Graphical Model”), 7 (Figure 2 and caption); 8-9 (section “2.3 Non-parametric belief propagation”); 10 (section “3.1 Vehicle Spatio-Temporal Tracking”). EX2024, ¶39.
5/13	EX1003 at 3 (“Abstract”), 5 (section “1.1 Related work”), 6 (“section “2.1 Building a Graphical Model”), 7 (Figure 2 and

	caption); 8-9 (section “2.3 Non-parametric belief propagation”); 10 (section “3.1 Vehicle Spatio-Temporal Tracking”). EX2024, ¶40.
6/14	EX1003 at 1 (Abstract), 1-2 (section “1 Introduction”), 5 (section “1.1 Related work”), 5-6 (section “2 Graphical Object Models”), 8 (section “2.3 Non-parametric BP”), 10 (section “2.5 Proposal Process”). EX2024, ¶41.
7/15	EX1003 at 1 (Abstract), 1-2 (section “1 Introduction”), 5-6 (section “2 Graphical Object Models”), 6-7 (section “2.2 Learning Spatial and Temporal Constraints”), 9-10 (section “2.4 AdaBoost Image Likelihoods”). EX2024, ¶42.
8/16	EX1003 at 1 (Abstract), 1-2 (section “1 Introduction”), 5-6 (section “2 Graphical Object Models”), 9-10 (section “2.4 AdaBoost Image Likelihoods”). EX2024, ¶43.

III. THE ASSERTED REFERENCES

A. Zhao (EX1006)

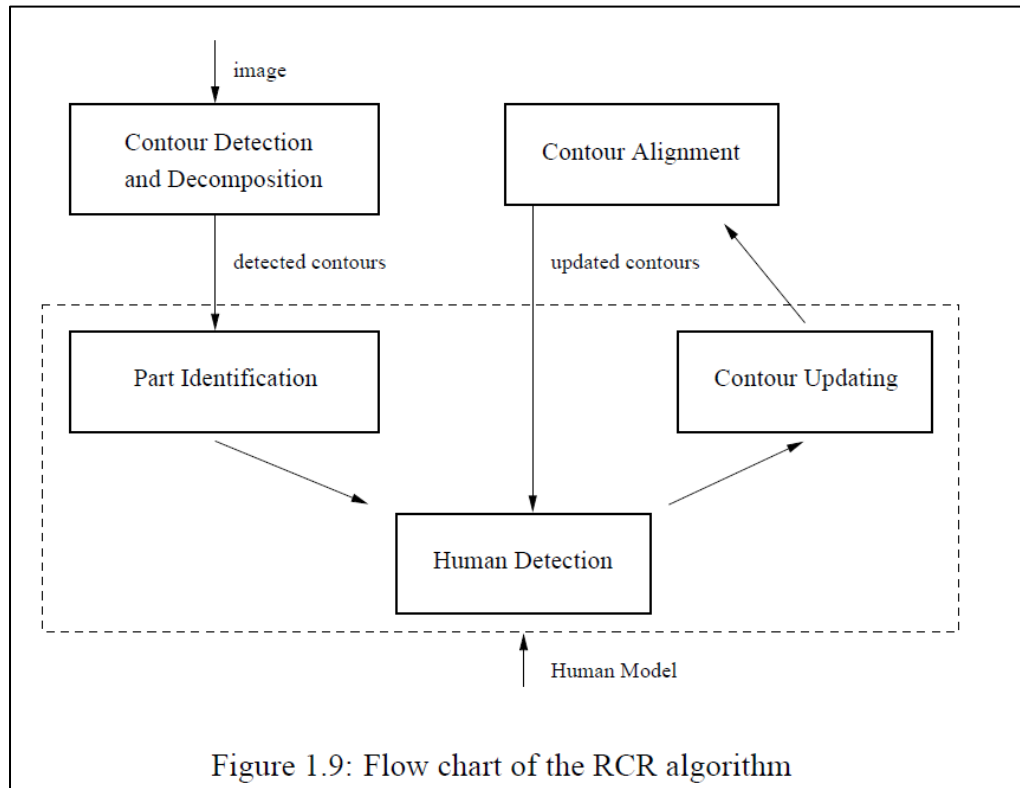
Zhao suggests that detecting humans “from a single image” is “inherently difficult for a computer.” EX1006, 20; EX2004, ¶72. Zhao also poses the following dilemma: “High performance object detection depends on reliable contour

extraction, but contour extraction is an under-constrained problem without knowledge about the objects to be detected.” EX1006, 69; EX2004, ¶¶73-74, 81. Zhao acknowledges that “without prior knowledge of the identification of the object, this goal is hard to achieve.” EX1006, 23; EX2004, ¶74. Zhao “proposes a recursive context reasoning (RCR) algorithm to solve the above dilemma.” EX1006, 69; EX2004, ¶74.

Zhao’s RCR algorithm uses a “TRS-invariant probabilistic model designed to encode the shapes of the body parts and the context information,” where the context information is “the size and spatial relationships between parts.” EX1006, 69; EX2004, ¶¶75-78, 82. It uses a “Bayesian framework ... to perform human detection and part identification under partial occlusion.” EX1006, 69; EX2004, ¶¶79-80, 82. It then uses a “contour updating procedure ... to integrate the human model and the identified body parts to predict the shapes and locations of the parts missed by the contour detector,” where “the refined contours are used to reevaluate the Bayesian similarity measure and determine if the detected contour is a person or not.” EX1006, 69; EX2004, ¶¶82-84. Zhao teaches using “geometrical relationships” and “spatial and size relationships” for updating and predicting the parameters of detected and undetected parts. EX1006, 71, 91; EX2004, ¶85. Zhao’s approach is different from the invention of ’980 patent, as Zhao’s recursive approach reevaluates the probability that a human is present in a given image—it does not detect a human

in a given image based on probabilities of the human being in other images.

EX2004, ¶83.

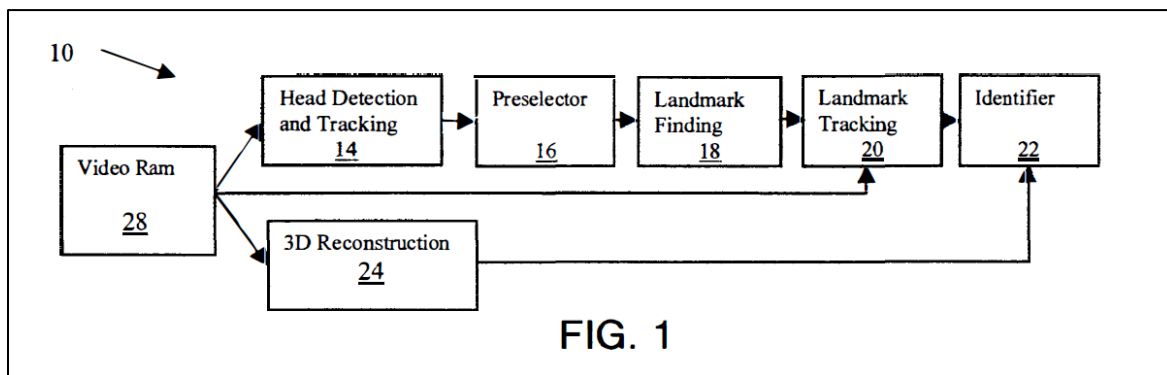


EX1006 at 26

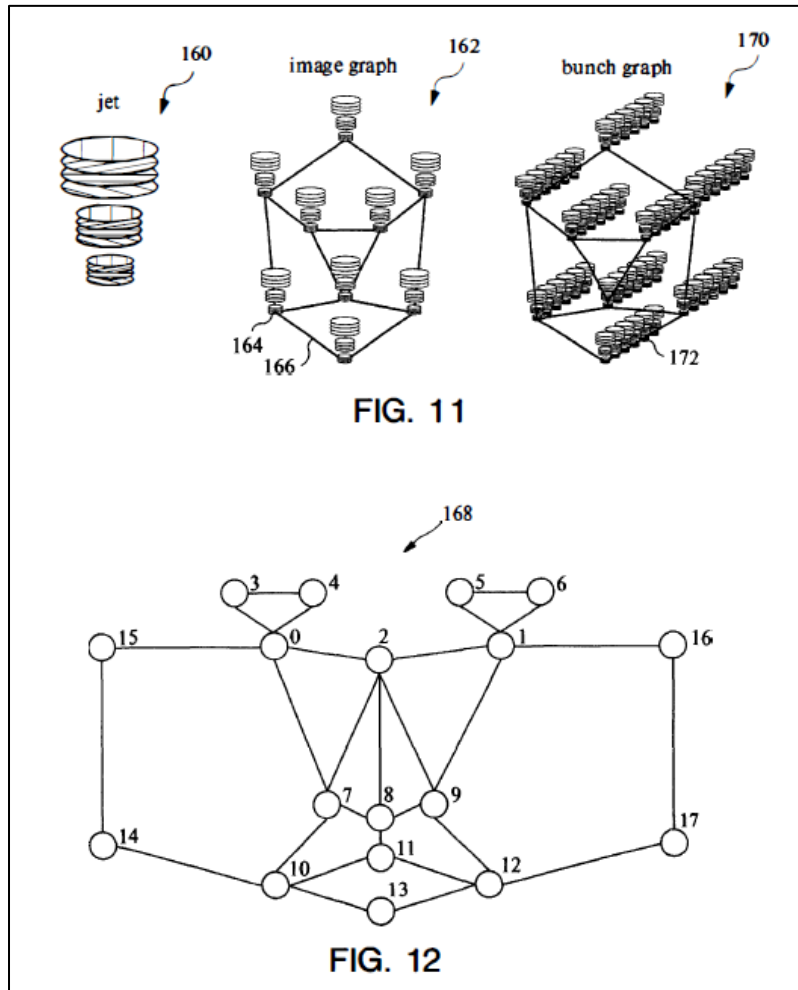
Zhao’s “Bayesian Similarity Measure” or “BSM” is derived from the body model and “evaluates the resemblance between the contour *C* and the human model.” EX1006, 54-55; EX2004, ¶79. Zhao defines human detection as: “given a detected contour *C*, determine whether this contour represents the silhouette of a person or not,” EX1006, 64; EX2004, ¶80, where “the contour *C* corresponds to a person if $P(\text{person}|C) > \text{threshold}$.” EX1006, 64. “Otherwise, *C* is not a person.” *Id.*

B. Steffens (EX1007)

Steffens relates to “vision-based object detection and tracking ... for detecting objects in video images, such as human faces, and tracking and identifying the objects in real time.” EX1007, 1:14-17, 3:36-37; EX2004, ¶86. The steps of Steffens’ process are summarized in Fig. 1, reproduced below. EX1007 at 3:47-4:61; EX2004, ¶86.



The “landmark locations on an image correspond[] to readily identifiable features,” *i.e.*, facial features of a head. EX1007, 3:34-46; EX2004, ¶87. “The facial landmarks and features of the head may be located using an elastic graph matching technique.” EX1007, 7:51-53; EX2004, ¶87. The graph matching technique involves “determining the similarity between [an] image graph and [a] model graph.” EX1007, 7:51-67; EX2004, ¶87. The model graph may be a “bunch graph,” in which “a whole bunch of jets 172 (a bunch jet) are attached to each node.” EX1007, 9:22-45, Figs. 11-12; EX2004, ¶87.



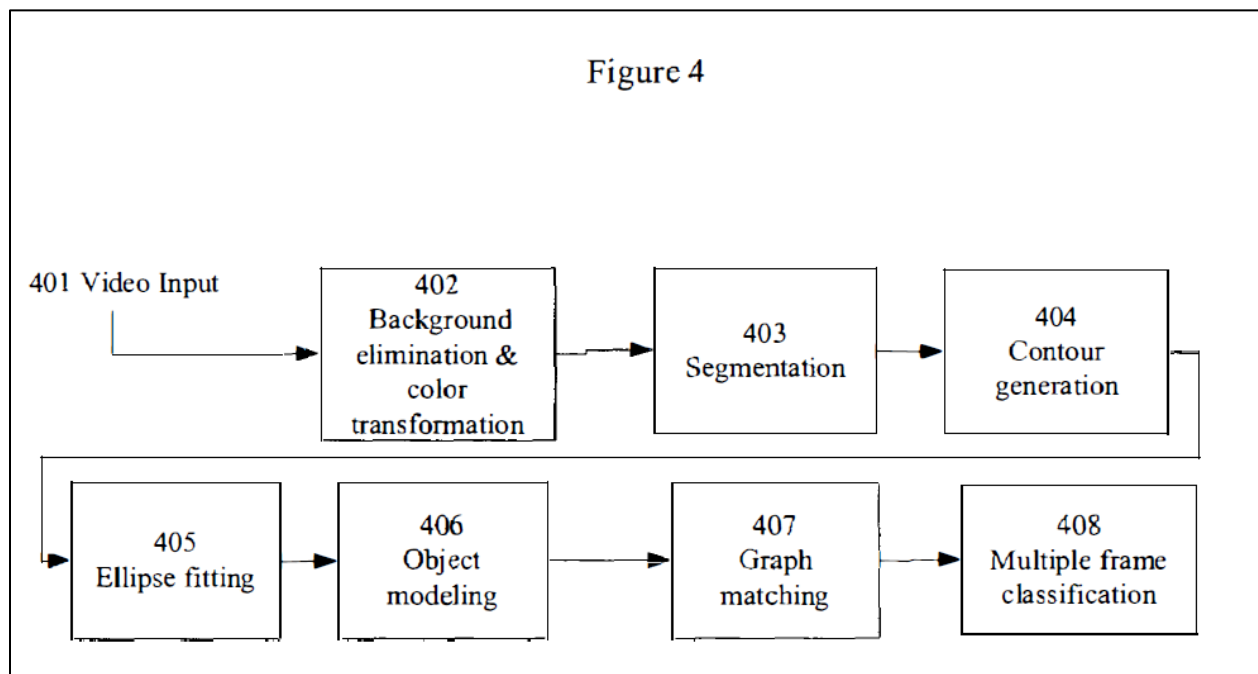
“When matching a bunch graph to an image, each jet extracted from the image is compared to all jets in the corresponding bunch attached to the bunch graph and the best-matching one is selected.” EX1007, 9:22-36; EX2004, ¶88. “The jets are composed of wavelet transforms.” EX1007, 3:42-44; EX2004, ¶89. “In order to find a face in an image frame, the graph is moved and scaled over the image frame until a place is located at which the graph matches best. EX1007, 9:37-41; EX2004, ¶89.

Steffens’ approach is different from the invention of ’980 patent, as Steffens’s

graph matching technique performs matching based on matching image graph data for a given image to a model graph—it does not detect an object in a given image based on probabilities of the object being in other images.

C. Ozer (EX1008)

Ozer relates to a “method and apparatus to detect the presence of articulated objects, e.g., human body, and rigid objects and to identify their activities.” EX1008, 1:15-21; EX2004, ¶90. It has “two levels, namely low and high levels,” where “[t]he low-level part performs object detection and extracts parameters for the abstract graph representation of the image being processed,” and “[t]he high level part uses dynamic programming to determine the activities of the object parts, and uses a distance classifier to detect specific activities.” EX1008, 3:55-64; EX2004, ¶90.



EX1008, Figure 4

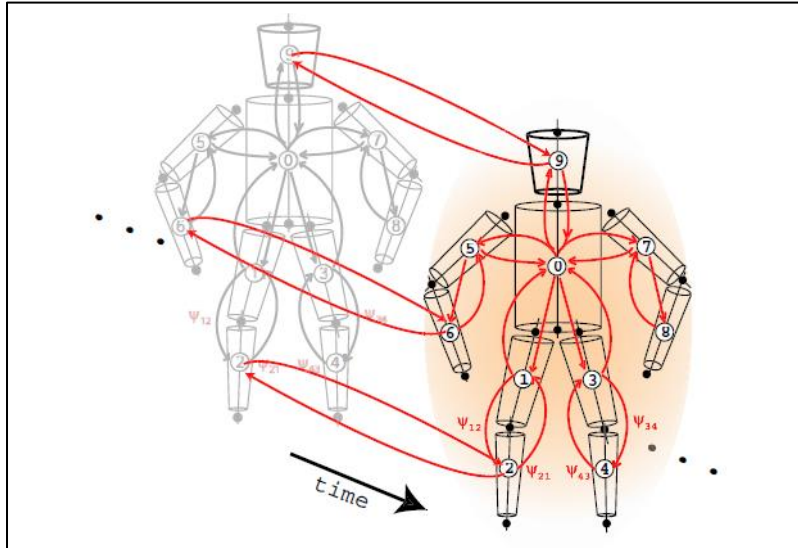
The “low-level” analysis includes the steps 402-407 shown in Figure 4 above, which generate “classified object parts.” EX1008, 6:42-7:35; EX2004, ¶91. The “high-level” analysis thereafter “checks direction of the movements of the object parts for a number of frames and ... and chooses the pattern with the highest probability as the recognized activity in these frames.” EX1008, 3:44-53, 7:36-42; EX2004, ¶¶92-93.

Ozer uses a stores data for “the current frame” in buffers, and that data is then used to “classify the object regions.” EX1008, 5:10-20; EX2004, ¶149. The data from the current frame is used to extract “foreground regions” of the frame and segment “the object of interest” into parts “based on the combination of color components and statistical shape features.” EX1008, 6:52-58; EX2004, ¶149. The “[l]ow-level part performs object detection and extracts parameters for the abstract graph representation of the frame being processed.” EX1008, 3:65-67; EX2004, ¶149. Ozer’s approach is different from the invention of ’980 patent, as Ozer’s graph matching technique performs matching based on data for a single “current” frame—it does not detect an object in a given image based on probabilities of the object being in other images.

D. TLP (EX1009)

TLP describes a “method for tracking human bodies” using “a loose-limbed body model . . . in which limbs are connected via learned probabilistic constraints.”

EX1009, 1; EX2004, ¶94. “Each node in the graphical model represents ... a limb (Figure 1).” *Id.*



EX1009 at 1

TLP teaches that “[m]ost current techniques model the body as a kinematic tree,” and that “[w]hen such algorithms lose track (as they always do), the dimensionality of the state space makes it difficult to recover.” EX1009, 1; EX2004, ¶95. However, “[w]hile the full body pose is hard to recover directly, the location and pose of individual limbs is much easier to compute.” EX1009, 1; EX2004, ¶95. Accordingly, TLP teaches that “the body is represented by a graphical model in which each graph node corresponds to a body part (upper leg, torso, etc.),” where “[e]dges in the graphical model correspond to position and angle relationships between adjacent body parts in space and time, as illustrated in Figure 1.” EX1009, 2; EX2004, ¶95.

TLP's approach is different from the invention of '980 patent, as TLP's tracking technique does not model the object as a distinct entity from its corresponding components or detect the object in a given image based on probabilities of the object being in other images.

IV. LEVEL OF ORDINARY SKILL

Petitioners set forth a level of ordinary skill. Petition, 11. Without agreeing or acquiescing to Petitioners' definition of the level of ordinary skill, this Response establishes that no challenged claim is unpatentable even under Petitioners' definition.

V. CLAIM CONSTRUCTION

In light of the intrinsic evidence, the POSITA would have interpreted the following claim term of the '980 patent as follows:

Term	Patent Owner' Construction
“detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object” (claims 1 and 9)	detecting the object in any image based on comparing the probabilities for each image of the plurality of images to a threshold for detecting the object

Claim 1² of the '980 patent is shown below:

1. 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.

EX1001, claim 1.

Limitation 1[b] requires a “plurality of images.” Here, “plurality” has its ordinary meaning of “at least two.” *SIMO Holdings Inc. v. Hong Kong uCloudlink Network Tech. Ltd.*, 983 F.3d 1367, 1377 (Fed. Cir. 2021) (“The phrase ‘a plurality of’ means ‘at least two of.’”); *York Prod., Inc. v. Cent. Tractor Farm & Fam. Ctr.*, 99 F.3d 1568, 1575 (Fed. Cir. 1996) (“The term [plurality] means, simply, ‘the state of being plural.’”); *Resqnet.com, Inc. v. Lansa, Inc.*, 346 F.3d 1374, 1382 (Fed. Cir. 2003) (“As previously noted, “plurality” ordinarily means “at least two.”). The patent consistently uses the word “plurality” in the ordinary sense. See EX1001,

² For purposes of this claim construction analysis, claim 1 is representative of claims 1 and 9.

Abstract, 1:28-36, 1:48-52, 2:7-18, 5:45-50. Thus, the POSITA would understand limitation 1[b] to require at least two images. EX2004, ¶65.

The POSITA would have understood the limitations following element 1[b] to refer to element 1[b]’s “plurality of images” such that each step builds on the performance of a previous step. EX2004, ¶66. Limitation 1[c] requires “measuring the object as a collection of components in *each image*,” *i.e.*, in each of the previously recited plurality of images. *Id.* Limitation 1[d] then requires “determining a probability that the object is in *each image*,” *i.e.*, for each of the previously recited plurality of images. *Id.* Limitation 1[e] then requires “detecting the object in *any image*,” *i.e.*, any singular image, “upon comparing the probabilities for *each image*,” *i.e.*, for each of the previously recited plurality of images, “to a threshold for detecting the object.” *Id.* The POSITA would have understood that after providing a plurality of images, and after measuring the object in and determining the probability that the object is in “each” of the images, the final step of detecting the object “in any image” is based on the object probabilities determined for “each image” of the plurality, thus requiring a multi-image analysis for detecting the object in any given image. *Id.*

The POSITA would have also considered that object detection based on a multi-image analysis is fundamental to invention’s purpose of providing “[o]bject detection that considers the consistency of the detection *over time*,” meaning based

on multiple images. EX1001, 2:43-3:2, 4:1-2. EX2004, ¶67; *Sequoia*, 66 F.4th at 1236. It is also fundamental to the invention’s function of providing detection with “temporal integration” and “temporal consistency” over multiple image frames and producing “consistent results over time.” EX1001, 7:6-22; EX2004, ¶67; *Medrad*, 401 F.3d at 1319.

The Board’s institution decision determined that element 1[e] does not require a multi-image analysis. Paper 11 at 26-27. It observed that “Element 505 of Figure 5 discloses ‘detecting the object in any image upon comparing the probabilities for *the image* to a threshold for detecting the object,’ rather than “*each image*” as recited in the claims. Paper 11 at 27. But the words “*the image*” in Figure 5 do not override the context of the specification as a whole, which, overwhelmingly supports Patent Owner’s construction. *Adams*, 383 U.S. at 49 (claims *and* specification to be read “with a view to ascertaining the invention”); *Phillips*, 415 F.3d at 1313, 1316 (claims are read “in the context of the entire patent”); *Seal-Flex, Inc. v. Athletic Track and Court Const.*, 172 F.3d 836, 845 (Fed. Cir. 1999) (One “cannot alter the plain meaning of the claim language by referring to parts of the specification out of context.”). Furthermore, the specification describes step 505 as “detecting the object in any image upon comparing the probabilities for *each* image to a threshold for detecting the object,” which matches elements 1/9[e] and aligns with the POSITA’s understanding of the invention in light of the specification as a whole. EX1001,

7:53-55; EX2004, ¶68.

Nor can the words “*the image*” in Figure 5 override the unambiguous meaning of “*each image*” in the claims. Limitation 1/9[e] clearly requires “comparing the probabilities” for *two or more* images to a threshold. When construing claims, “[t]he actual words of the claim are the controlling focus.” *Digital Biometrics, Inc. v. Identix, Inc.*, 149 F.3d 1335, 1344 (Fed. Cir. 1998). Though claims “must be read in view of the specification,” *Phillips*, 415 F.3d at 1315, the specification cannot “be used to rewrite, the chosen claim language.” *Resonate Inc. v. Alteon Websystems, Inc.*, 338 F.3d 1360, 1364 (Fed. Cir. 2003); *Cent. Admixture Pharmacy Servs., Inc. v. Advanced Cardiac Sols., P.C.*, 482 F.3d 1347, 1355 (Fed. Cir. 2007) (“Claims mean precisely what they say.”).

While the specification describes a scenario where only one image is analyzed—where the number of frames *w* over which the analysis is performed is equal to 1 (*i.e.*, where “*w=1*”)—the patent distinguishes that scenario from the invention. EX2004, ¶69. The patent states that “with *w=1*, the algorithm resembles *single frame* component-based fusion,” EX1001, 7:20-22, which considers only the connected components in a single image. EX2004, ¶69; *see* EX2006, 97:9-21, 101:1-102:3. The patent then distinguishes the *w=1* scenario from the invention and demonstrates that it does not achieve the purpose or function of the invention. EX1001, 7:56-8:3; EX2004, ¶69. The specification describes a real-world

comparison of its multi-image analysis (a 3-frame analysis) against a “simple fusion detection” that was run “independently at each frame” (the $w=1$ scenario), which showed that the multi-image analysis “combines the evidence for accurate estimates throughout the sequence,” while the single frame “simple fusion detection” had “poor” performance. EX1001, 7:56-8:3; EX2004, ¶69. The POSITA would have understood the specification to be contrasting the claimed invention with a known technique to highlight a function of the claimed invention. EX2004, ¶69.

It is common for a patent’s specification to discuss existing technologies to distinguish the old from the new in the claims. *See Sjolund v. Musland*, 847 F.2d 1573, 1579-80 (Fed. Cir. 1988) (“One difference between the invention and what was old is immediately apparent: claim 7 of the ’071 patent claims a ‘**plurality**’ of panels -- that is to say, **more than one** -- whereas the old practice was to use a single framed panel.”)). That is the case here. *See* EX2004, ¶69. To interpret the claims as encompassing no more than detecting an object based on a probability for a single image would contradict the plain language of the claims and ignore the invention’s purpose and function described in the specification. *Medrad*, 401 F.3d at 1319; *Sequoia*, 66 F.4th at 1236; EX2004, ¶69.

Reading claims 1 and 9 as covering the $w = 1$ scenario would also read in configurations that are incapable of achieving the invention’s function and purpose. EX1001, 2:49-50, 7:6-22; EX2004, ¶70. When $w=1$, the detection inference would

be over the contents of a single frame, as the window would extend a temporal distance of “0” in either direction from time t .³ EX2004, ¶70. Thus, it would consider only the spatial relationships among the object and its components in that single frame—it would not be an inference over time or a spatio-*temporal* model. *Id.*; see EX2006, 103:22-104:3. In contrast, when $w=3$, the inference would be over the contents of three images, for example, at $t-1$, t , and $t+1$, as the window would extend a temporal distance of “1” in either direction from time t .⁴ EX2004, ¶70.

The specification further supports the POSITA’s understanding that 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 the threshold. For example, the patent teaches that “object detection” is “formulated as an inference” in a “spatio-temporal model” that covers more than “a single frame” because it is “extended over time” and therefore includes instances

³ The model at “ $t-(w-1)/2$ ” when $w=1$ would be at $t-(1-1)/2 = t-(0)/2 = t-0$.

EX1001, 4:45-48; EX2004, ¶56. The model at “ $t+(w-1)/2$ ” when $w=1$ would be at $t+(1-1)/2 = t-(0)/2 = t+0$. *Id.*

⁴ The model at “ $t-(w-1)/2$ ” when $w=3$ would be at $t-(3-1)/2 = t-(2)/2 = t-1$.

EX1001, 4:45-48; EX2004, ¶56. The model at “ $t+(w-1)/2$ ” when $w=3$ would be at $t+(3-1)/2 = t+(2)/2 = t+1$. *Id.*

of the object at a “current time” and “earlier” or “later” times. EX1001, 2:55-3:2, EX1001, 4:18-48, Figs. 2A and 2B; EX2004, ¶71. The detection inference is defined as an estimation of “belief” in the model over a w-frame window, which compares the likelihood of object presence in each image to a threshold and sends “messages” based on that likelihood (*e.g.*, when it scores above the threshold) to the object in the image under consideration for detecting the object. EX1001, 5:25-7:44; EX2004, ¶71. In light of this context, the POSITA 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. EX2004, ¶71.

VI. PETITIONERS HAVE NOT DEMONSTRATED THAT ANY CHALLENGED CLAIM IS UNPATENTABLE

Petitioners do not show that Zhao, Ozer, Steffens, or TLP discloses every element of claims 1/9. *See* EX2006, 13:13-14:3. Petitioners’ attempt to fill the gaps with assertions of obviousness, but they fail to show that every limitation of claims 1/9 (*i.e.*, the claims as a whole) would have been obvious. Accordingly, Petitioners have not satisfied the basic requirements of proving obviousness. *W.L. Gore & Assocs., Inc. v. Garlock, Inc.*, 721 F.2d 1540, 1548 (Fed. Cir. 1983) (“Each claimed invention must be considered as a whole.”) *Ruiz*, 357 F.3d at 1275 (“[S]ection 103 specifically requires consideration of the claimed invention ‘as a whole.’”).

A. The challenged claims are patentable over Zhao (Ground I).

1. Claims 1/9

a. Elements 1/9[e]

Petitioners fail to show that Zhao discloses or renders obvious “detecting the object *in any image* upon comparing the probabilities *for each image* to a threshold for detecting the object,” which captures critical features of how the invention accomplishes its purpose of providing “[o]bject detection [that] considers the consistency of the detection over time” and its function of providing “temporal consistency” over multiple image frames and producing “consistent results over time.” EX1001, 2:49-50, 7:6-22; *See* § VI, *supra*; EX2004, ¶¶99-100. The POSITA would have understood that repeating a “single-frame” or “simple fusion” analysis “independently at every frame” does not meet the requirements of this element. EX1001 at 7:20-22, 7:56-8:3; *see* § VI, *supra*; EX2004, ¶100. Yet, as explained below, Zhao determines the probability that a human is in a single image and compares only that probability to a threshold for detecting the human in that image—the detection is not based on the probability that the human is present in another image. EX2004, ¶100.

i. Petitioners fail to address every limitation of elements 1/9[e] and therefore fail to show they were disclosed or rendered obvious by Zhao.

Petitioners’ analysis of element 1[e] is based on their analysis of element 1[d].

See Petition, 26-27. Element 1[d] requires, “determining a probability that the object is in each image.” EX1001, claim 1. Petitioners argue that Zhao satisfied element 1[d] using Zhao’s “Bayesian Similarity Measure” that “evaluates the resemblance between a contour and a model based on the best match between their body parts.” Petition, 25. Petitioners point to Zhao’s “simple decision rule” where the probability that the “contour C corresponds to a person” in an image is calculated as $P(\text{person}|C) > \text{threshold}$ and its reformulation as $P(C|H^*, \text{person})P(\text{person}|H^*)$. *Id.* Petitioners equate “determining the probability that a contour corresponds to a person” in Zhao with element 1[d]’s requirement of “determining a probability that the object is in *each* image.” *Id.* See EX2004, ¶102.

Petitioners rely on the same probability determination for their element 1[e] arguments. Specifically, in attempting to show that “Zhao compares the probability to a threshold to detect the object (*i.e.*, human),” Petitioners point to both “ $P(\text{person}|C) > \text{threshold}$ ” and “ $P(C|H^*, \text{person})P(\text{person}|H^*) > \text{threshold}$.” Petition, 26-27. See EX2004, ¶103.

But Petitioners never explain, in either their element 1[d] or element 1[e] analyses, how comparing the probability that a contour corresponds to a person to a threshold (*i.e.*, determining $P(\text{person}|C) > \text{threshold}$) or $P(C|H^*, \text{person})P(\text{person}|H^*) > \text{threshold}$) allegedly discloses or renders obvious “detecting the object *in any image* upon comparing the probabilities *for each image* to a

threshold for detecting the object.” EX2004, ¶104. That is, Petitioners do not explain how Zhao discloses or renders obvious this element as a whole. *Sanofi-Synthelabo v. Apotex, Inc.*, 550 F.3d 1075, 1086 (Fed. Cir. 2008) (“The determination of obviousness is made with respect to the subject matter as a whole, not separate pieces of the claim.” (citing cases)); *W.L. Gore*, 721 F.2d at 1548; *Ruiz*, 357 F.3d at 1275.

Instead, Petitioners purport to show that that Zhao “determine[es] the probability that the object is in each image” and that it also “compares the probability to a threshold to detect the object.” Petition, 26-27; EX2004, ¶105. But the claim requires more. EX2004, ¶105; *see* § VI, *supra*. Petitioners never explain how Zhao’s RCR algorithm (including its use of the BSM) would allegedly disclose or render obvious “detecting the object *in any image*” *i.e.*, in any given image, “upon comparing the probabilities *for each image* to a threshold for detecting the object.” EX2004, ¶105; *Univ. of Strathclyde v. Clear-Vu Lighting LLC*, 17 F.4th 155, 160 (Fed. Cir. 2021) (“An obviousness determination generally requires a finding that ‘all claimed limitations are disclosed in the prior art.’”). Nor do Petitioners even argue, let alone meet their burden of proving, that it would have been obvious to modify Zhao to arrive at the claimed invention as a whole. EX2004, ¶105; *Arendi*, 832 F.3d at 1361 (obviousness over a single reference requires showing that “it would have been obvious to modify that reference to arrive at the patented

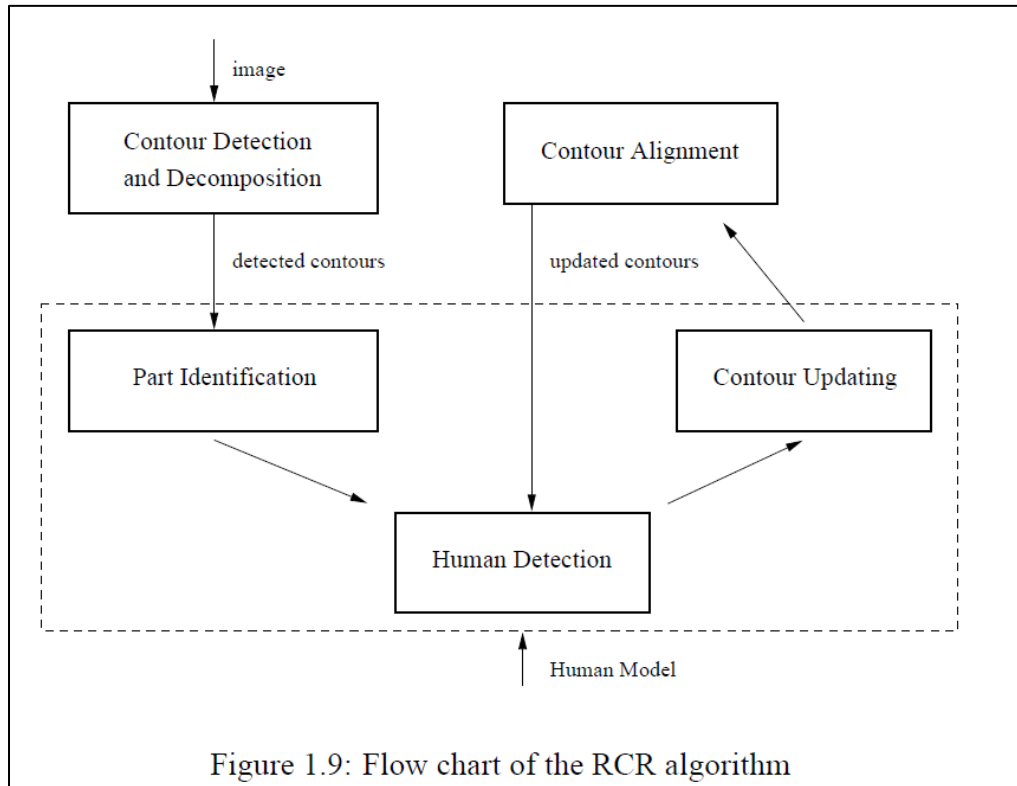
invention”); *In re Van Os*, 844 F.3d 1359, 1361 (Fed. Cir. 2017) (the proper inquiry assesses “whether there would have been a motivation to combine elements from prior art references *in the manner claimed*”).

Dr. Bajaj’s further analysis for element 1/9[e] does not resolve the deficiency. EX2004, ¶106. Dr. Bajaj opines that “employing a threshold in a probability measure was known to the POSITA” and “would be used in a probabilistic measure to perform any type of detection.” EX1004 ¶99. But whether the use of a threshold for determining something as “detected” or “not detected” was known in the art does not show that the POSITA would have found it obvious to modify Zhao to arrive at the specific limitations of the claimed invention as a whole. EX2004, ¶106; *Arendi*, 832 F.3d at 1361; *Van Os*, 844 F.3d at 1361; *Univ. of Strathclyde*, 17 F.4th at 160; *Sanofi-Synthelabo*, 550 F.3d at 1086. Specifically, it does not indicate to the POSITA that Zhao discloses or renders obvious the particular way of using a threshold described in claim 1/9[e], namely “detecting the object *in any image* upon comparing the probabilities *for each image* to a threshold for detecting the object.” *Id.* Dr. Bajaj’s opinion that it does is based on an incorrect claim interpretation that ignores these express claim limitations, given his opinion that claims 1/9 cover the “w=1” scenario. *See* § VI, *supra*; EX2006, 18:8-14, 97:22-99:15; EX2004, ¶106. Thus, neither Petitioner nor Dr. Bajaj shows that Zhao (or any other reference) discloses or renders obvious the express requirement in claims 1/9[e] (even under

their preferred interpretation) of “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.”

ii. Zhao does not disclose or render obvious elements 1/9[e].

The POSITA would have understood that Zhao’s RCR algorithm, which utilizes the Bayes Similarity Measure that Petitioners rely on for elements 1/9[e], performs a human detection analysis for a single image that is not based on probabilities of the human being in other images (or comparing them to a detection threshold). EX2004, ¶107. Zhao acknowledges that “without prior knowledge of the identification of the object, this goal [of shape decomposition] is hard to achieve.” EX1006 at 23; EX2004, ¶107. But Zhao does not disclose a way of incorporating prior (or later) knowledge of the identification of the object into a detection framework. EX2004, ¶107. Zhao expressly assumes such information is unavailable: “In fact contour extraction and shape decomposition are under-constrained problems without the knowledge about the objects to be detected.” EX1006, 26, 69; EX2004, ¶107. Instead, Zhao developed “a recursive context reasoning (RCR) algorithm (see Fig. 1.9) to solve the above dilemma.” EX1006, 26; EX2004, ¶107.



EX1006, 26

Zhao’s RCR algorithm operates on a single image, shown as an input in the upper left corner of Zhao’s Figure 1.9. EX2004, ¶108; *see* EX2006, 106:2-107:17. The RCR algorithm performs a number of operations on that image, including extracting a contour from the image (*i.e.*, a silhouette of an object), decomposing the contour into segments (“ribbons”), identifying body parts in the contour using a TRS-invariant body model, and determining if the contour corresponds to a person using a BSM. *See* § VIII.A, *supra*; EX2004, ¶108.

But 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.

EX2004, ¶109. Instead, Zhao's RCR algorithm uses "a contour **updating** procedure" to "predict the contours of missed parts" and align them to generate "refined contours," which are "used to **reevaluate** the likelihood of a person being presented in the image." EX1006, 26; EX2004, ¶109. In other words, Zhao's RCR algorithm simply updates the extracted contours in the image and then "reevaluates," *i.e.*, repeats, its determination of whether a person is present in the image (*i.e.*, it repeats the BSM for the same image). EX1006, 26, 64-65; EX2004, ¶109. The POSITA would have understood that Zhao's process of updating contours and re-running the BSM evaluation on the same image does not disclose or render obvious detecting an object (*i.e.*, the person) in a given image based on the probability of the object (*i.e.*, the person) being detected in another image or comparing that probability to a detection threshold. EX2004, ¶109.

Nor does the RCR algorithm involve determining or considering the probability of the object (the person) being present in another image or comparing that probability to a detection threshold. EX2004, ¶110. Steps 1-4 of the RCR algorithm ("Contour extraction," "Contour decomposition," "Body part identification," and "Human detection") operate on the single image input into the RCR algorithm, and their collective performance analyzes information extracted from *that image* and results in a determination of whether a detected contour *C* in *that image* corresponds to a person or not. See EX1006 at 17-71; EX2004, ¶110.

Knowledge of whether the object (the person) is in other images is not involved. See EX1006, 23,69; EX2004, ¶110.

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. EX2004, ¶111. The purpose of these steps is “to refine the shapes and locations of the identified body parts,” *i.e.*, the parts already identified in the image under consideration, “and to predict the shapes and locations of the missed body parts.” EX1006, 72; EX2004, ¶111. Zhao teaches performing those updates and predictions based on “geometrical relationships” and “spatial and size relationships,” not based on the probability of the object (the person) being present in other images. EX1006, 71, 91; *see also* EX1006, 72 (using vector (a, l, x, y, θ)), 74 (using vector $(\hat{a}_j, \hat{l}_j, \hat{x}_j, \hat{y}_j)$); EX2004, ¶111.

While Zhao states that “[m]otion information obtained from previous frames can be used to predict the orientations of the limbs and to solve [limb] ambiguity,” the POSITA 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. EX1006, 91; EX2004, ¶112. As explained above, Zhao teaches updating and predicting the

parameters of detected and undetected parts based on “geometrical relationships” and “spatial and size relationships,” not based on the probability of the object (the person) being present in other images. EX1006, 71, 91, 72, 74; EX2004, ¶112. Zhao’s example of considering “how the left knee angle changes with time” confirms to the POSITA that Zhao’s consideration of “motion information” from “previous frames” is based on geometrical and spatial relationships, not the probability that the object (the person) is present in another image or comparing that probability to a detection threshold. EX2004, ¶112.

2. Claims 4/12

Claim 4/12 depends on claim 1/9, respectively. The arguments for claims 1/9 are incorporated by reference. *See* § VII.A.1, *supra*.

Petitioners fail to show that Zhao discloses or renders obvious the specific requirements of these claims. These claims require “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.” EX2004, ¶¶113-115. Thus, the claims require detecting the object in a current (first) image according to measurements of the object from a later (second) image and a later (third) image, *i.e.*, a three-frame analysis for detecting an object in a given image. EX2004, ¶115. Zhao does not disclose such a technique or render it obvious. *Id.* Nor have Petitioners shown that this technique was “widespread and commonly

understood” or that it would have been obvious to modify Zhao to achieve it. *Id.*

Notably, the Petition does not explain how Zhao allegedly describes the limitations of these claims. Petition, 29. This is sufficient reason to find the claims patentable. *See, e.g., Ashworth Bros., Inc. v. Laitram, L.L.C.*, IPR2020-00594, Paper 60 at 44 (PTAB Aug. 23, 2021) (“Arguments and information that are not presented and developed in a brief, but instead are incorporated by reference, are not entitled to consideration, as it is improper to incorporate by reference arguments from one document into another document.” (citing 37 C.F.R. § 42.6(a)(3) and ; *Cisco Sys., Inc. v. C-Cation Techs., LLC*, IPR2014-00454, Paper 12 at 7–10 (PTAB Aug. 29, 2014) (informative))).

Petitioners defer entirely to Dr. Bajaj’s declaration, which is improper. *Ashworth Bros.*, IPR2020-00594, Paper 60 at 44. But Dr. Bajaj fails to show that these claims are disclosed or would have been rendered obvious by Zhao.

Referring to Zhao’s “MAP (maximum a posteriori) probability likelihood estimation of a body part’s parameter vector,” Dr. Bajaj claims that “Zhao explicitly discloses” the use of MAP probability estimates “to enable measurements of the human object as a collection of components determined from a prior image *and a later image* relative to the current image.” EX1004 ¶108; EX2004 ¶116. However, Dr. Bajaj identifies no disclosure in Zhao that even suggests, let alone “explicitly discloses,” the use of later images at all, including in connection with MAP

probability estimations. EX1004 ¶108; EX2004 ¶116. Dr. Bajaj cites no portion of Zhao to support his assertion, thus he has not shown that Zhao discloses or renders obvious every limitation of these claims. EX2004 ¶116. Dr. Bajaj’s interpretation of Zhao is also incorrect. EX2004 ¶116.

Dr. Bajaj asserts that Zhao discloses a step “to predict the outlines of the unidentified body part from the human model and the identified body parts” using MAP estimation. EX1004 ¶108; *see* EX2004 ¶117. Dr. Bajaj then states that “[i]f more than two main body parts have been MAP identified, then the TRS ... transformation that transforms the coordinates from the spatio-temporal human model coordinate system to the image coordinate system is also calculated as part of the coarse-to-fine backtracking search.” *Id.*

Zhao’s description of its MAP estimation, however, does not disclose or suggest the use of information from a later image. EX1006, 74-75; EX2004 ¶118. The transformation in Zhao’s MAP estimation simply transforms from the model’s coordinate system to the coordinate system of a given image. EX1006, 74; EX2004, ¶118. Claims 4/12, however, require something different, namely “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.” EX1006, 74; EX2004, ¶118. This is not disclosed or suggested by Zhao’s MAP estimation.

Zhao's use of "[m]otion information obtained from *previous* frames ... to predict the orientations of the limbs and to solve [limb] ambiguity" also does not disclose or suggest the use of measurements from "a *later* image relative to the current image." EX1004, ¶107; EX2004 ¶119. Dr. Bajaj identifies no disclosure in Zhao of the use of measurements from a later image. EX1004 ¶¶106-110; EX2004 ¶119. Petitioners have therefore failed to show that Zhao discloses or renders obvious every element of these claims as a whole. *Univ. of Strathclyde*, 17 F.4th at 160; *Sanofi-Synthelabo*, 550 F.3d at 1086.

And although Petitioners and Dr. Bajaj suggest that analyzing data in later frames was "widespread and commonly understood," citing three non-asserted references, they provide no explanation as to how any of those references demonstrate that it was "widespread and commonly understood" to detect an object in a current image according to measurements of the object "from a prior image and a later image relative to the current image." Petition, 29; EX1004, ¶109; EX2004, ¶119. Petitioners and Dr. Bajaj also provide no reason or rationale as to why the POSITA would have found it obvious to modify Zhao to arrive at the specific limitations of the claimed invention as a whole. EX2004, ¶119; *Arendi*, 832 F.3d at 1361; *Van Os*, 844 F.3d at 1361. Pointing out that part of the claim may have been known in the art does not satisfy the test for obviousness. *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 418 (2007) ("[A] patent composed of several elements is not

proved obvious merely by demonstrating that each of its elements was, independently, known in the prior art.”); *In re Stepan Co.*, 868 F.3d 1342, 1345-46 (Fed. Cir. 2017) (“An obviousness determination requires finding both ‘that a skilled artisan would have been motivated to combine the teachings of the prior art ... and that the skilled artisan would have had a reasonable expectation of success in doing so.’”).

3. Claims 8/16

Claim 8/16 depends on claim 1/9, respectively. The arguments for claims 1/9 are incorporated by reference. *See* § VII.A.1, *supra*.

Petitioners’ assertion that “the joint probability used in Zhao ... is easily representable in the form of the joint probability distribution formulated by Claim 8” is incorrect because claim 8/16’s joint probability distribution includes elements for a spatio-temporal model that do not exist in Zhao’s framework. EX2004, ¶¶120-122. Neither Petitioners nor Dr. Bajaj provides an explanation as to how it would have been obvious to modify Zhao to arrive at the particular probability distribution of claims 8/16. EX2004, ¶122.

The probability distribution of claims 8/16 is defined in the ’980 patent’s specification at 3:47-4:8. EX2004, ¶123.

$$\begin{aligned}
 P(X_0^O, X_0^{C_0}, X_0^{C_1}, \dots, X_0^{C_N}, \dots, X_T^O, X_T^{C_0}, X_T^{C_1}, \dots, X_T^{C_N}) = \\
 \frac{1}{Z} \prod_{ij} \psi_{ij}(X_i^O, X_j^O) \\
 \prod_{ik} \psi_{ik}(X_i^O, X_i^{C_k}) \prod_{ikl} \psi_{kl}(X_i^{C_k}, X_i^{C_l}) \prod_i \phi_i(X_i^O) \prod_{ik} \phi_i(X_i^{C_k}).
 \end{aligned}$$

EX1001 3:55-65

X_t^O is the state of the object node “O” at time t, and $X_t^{C_n}$ is the state of the object’s n-th component node “C_n” at time t, where the object has N components. EX1001, 3:53-4:8; EX2004, ¶123. Accordingly, X_t^O and $X_t^{C_n}$ explicitly correspond to an object node and distinct component nodes at multiple distinct points in time EX2004, ¶123; see EX2006, 70:9-17, 80:15-21.

The probability distribution also includes terms that represent spatial or temporal compatibility between specific node types. See § II, *supra*; EX2004, ¶124. Those terms include $\psi_{ij}(X_i^O, X_j^O)$, which represents “the temporal compatibility of object state between frames i and j,” which corresponds to *temporal constraints between two object nodes in different frames or images, i.e., at different points in time*. EX1001, 6:22-64, 5:25-6:13, 3:47-4:8; EX2004, ¶124.

Petitioners fail to show that Zhao discloses a joint probability distribution that has a temporal dimension or that explicitly represents object-level nodes at multiple points in time. Zhao states that “the probability distributions of the three model

matrices A , S , U can be estimated separately” and “provide metrics to evaluate the shape, size, relationship, and configuration similarities between the detected contour and the human body model.” EX1006, 45; EX2004, ¶125.

The POSITA would have understood that Zhao’s “probability distributions” for its human body model have no temporal dimension and do not represent temporal compatibility between objects in different images, as they are based on the non-temporal model matrices of Zhao’s TRS-invariant model. See claim 4, *supra*; EX2004, ¶126. Accordingly, the POSITA would not have understood Zhao’s probability distribution to be “representable in the form of the joint probability distribution formulated by Claim 8,” or that it would have been obvious to formulate it that way. Petition, 34; EX2004, ¶126.

Petitioners and Dr. Bajaj also have not shown that the joint probability formulation of claims 8/16 “would have been well known” or that EX1012 discloses that formulation. Petition, 35; EX2004, ¶127. Indeed, EX1012 states that “We apply the [non-parametric belief propagation] algorithm to infer **component** interrelationships in a **parts-based** face model, allowing location and reconstruction of occluded **features**.” EX1012, 1, 3, 6. Petitioners and Dr. Bajaj identify no disclosure in EX1012 of object-level representation in combination with component-level representation. Nor do they provide a reason or rationale as to why the POSITA would have found it obvious to modify Zhao to arrive at the specific

probability distribution of claims 8/16 as a whole. EX2004, ¶127; *Arendi*, 832 F.3d at 1361; *Van Os*, 844 F.3d at 1361. Pointing out that “pairwise Markov field representation of a spatio-temporal graphical model” in general may have been known does not satisfy the test for obviousness. *KSR*, 550 U.S. at 418.

4. All dependent claims are patentable over Zhao.

Claims 2-8 depend on claim 1 and claims 10-16 depend on claim 9. Claims 1 and 9 are patentable over Zhao for the reasons explained above. Accordingly, by their dependency on claim 1 and 9, respectively, claims 2-8 and 10-16 are patentable over Zhao.

B. The challenged claims are patentable over Steffens in view of Zhao (Ground II).

1. Claims 1/9

a. Elements 1/9[d]

The arguments for claims 1/9 for Zhao under Ground I are incorporated by reference. *See* § VII.A.1, *supra*.

Petitioners acknowledge that Steffens’ “‘similarity’ measure” is not formulated as calculating a “probability” that an object is in an image. Petition, 44; EX2004, ¶128. Petitioners assert that because “probabilistic frameworks” were known, it would have been obvious to combine the teachings of Steffens with Zhao’s Bayesian Similarity Measure. Petition, 44-45; EX2004, ¶128. Petitioners fail to show that it would have been obvious to combine Steffens and Zhao to achieve this

claim element. EX2004, ¶128.

The POSITA would have understood that Steffens' wavelet-based approach is fundamentally different from, and incompatible with, Zhao's contour-based approach. EX2004, ¶129. Zhao teaches that while "[m]ost human tracking and motion analysis systems employ a simple segmentation procedure," *i.e.*, "contour extraction," there are other approaches, including "wavelet analysis" in which "no contours are required to extract from the image first." EX1006, 20, 81-82, 86; EX2004, ¶129. Zhao's Bayesian similarity model for human detection, however, is fundamentally based on contour extraction, *see* EX2006, 111:21-112:4, which the POSITA would have viewed as incompatible with Steffens' wavelet-based approach. EX1006, 64; EX2004, ¶129. Thus, the POSITA would not have looked to Zhao's contour-based Bayesian similarity measure approach to modify Steffens' wavelet-based bunch graph approach. EX2004, ¶129.

Furthermore, Zhao teaches away from combining Steffens' wavelet-based bunch graph approach with Zhao. EX2004, ¶130. Zhao observes that a wavelet-based approach "has to search the whole image at multi-scales for humans" and thus "would be an extremely computationally expensive procedure, and it may cause multiple responses from a single human." EX1006, 82-83; EX2004, ¶130. Zhao further observes that another wavelet-based approach "has to search for pedestrians in different scales in the whole image, so it is very slow." EX1006, 88; EX2004,

¶130. Like the wavelet-based approaches disparaged by Zhao, in Steffens' approach, "[t]he jets are composed of wavelet transforms," and "[i]n order to find a face in an image frame, the graph is moved and scaled over the image frame until a place is located at which the graph matches best. EX1007, 3:42-44, 9:37-41; EX2004, ¶130. Thus, the POSITA would not have been motivated to combine Zhao and Steffens and would have been led away from the combination. EX2004, ¶130.

b. Elements 1/9[e]

i. Petitioners fail to address every limitation of elements 1/9[e] and therefore fail to show they were disclosed or rendered obvious by Steffens and Zhao.

The arguments for claims 1/9 for Zhao under Ground I are incorporated by reference. *See* § VII.A.1, *supra*.

Relying on their arguments for claim 1/9[d], Petitioners assert that "it would have been obvious to the POSITA to use a probabilistic framework" and that "as part of a probabilistic framework, it would have been obvious to the POSITA to detect the image by comparing the probabilities for each image to a threshold for detecting the object." Petition, 45. As with their arguments for Zhao, Petitioners fail to consider this element as a whole. EX2004, ¶131; *Sanofi-Synthelabo*, 550 F.3d at 1086. Petitioners never explain why "detecting the object *in any image*" *i.e.*, in any given image, "upon comparing the probabilities *for each image* to a threshold for detecting the object" would have allegedly been obvious in view of Steffens and

Zaho. Univ. of Strathclyde, 17 F.4th at 160; EX2004, ¶131. Nor do Petitioners show that it would have been obvious to combine Steffens and Zhao to arrive at the claimed invention as a whole. EX2004, ¶131; *Arendi*, 832 F.3d at 1361; *Van Os*, 844 F.3d at 1361.

Petitioners argue that this element would have been obvious because “the use of a ‘threshold’ was widely known.” Petition, 45. But that does not satisfy the test for obviousness. *KSR*, 550 U.S. at 418. Petitioners do not explain why that would have rendered obvious this element as a whole—the difference between element 1/9[d] and element 1/9[e] is not just the use of a threshold. See § VI, *supra*; *Univ. of Strathclyde*, 17 F.4th at 160; *Sanofi-Synthelabo*, 550 F.3d at 1086; *Arendi*, 832 F.3d at 1361; *Van Os*, 844 F.3d at 1361; EX2004, ¶132. Petitioners defer entirely to Dr. Bajaj’s declaration, which is improper. Petition, 45; *Ashworth Bros.*, IPR2020-00594, Paper 60 at 44. But Dr. Bajaj also fails to show that this element would have been obvious. EX2004, ¶132.

Rather than addressing the actual claim language, Dr. Bajaj claims that “employing a threshold in a probability measure was known to the POSITA” because “there are many examples of using thresholds in probabilistic frameworks.” EX104 ¶ 147; EX2004, ¶132. But whether the concept of “employing a threshold in a probability measure” was known in the art does not show that the POSITA would have found the specific requirements of claim 1/9[e] obvious over the

combination of Steffens and Zhao. EX2004, ¶132. That the use of “a threshold in a probability measure” was known does not indicate to the POSITA that Steffens or Zhao discloses or renders obvious the particular way of using a threshold described in claim 1[e]/9[e], namely “detecting the object *in any image* upon comparing the probabilities *for each image* to a threshold for detecting the object,” or that it was otherwise known in the art. EX2004, ¶132.

Petitioners also cite their Ground I arguments for Zhao in support of an argument that that the combination of Steffens and Zhao discloses a “Bayesian Similarity Measure” that “determines the probability that an object is in each image of a video” and “further includes comparing the probabilities for each image to a threshold for detecting an object.” Petition, 45. For the reasons explained above regarding Zhao under Ground I (*see* § VII.A.1, *supra*), Petitioners fail to show that Zhao discloses or renders obvious this element.

ii. Steffens in view of Zhao does not disclose or render obvious elements 1/9[e].

The arguments for claims 1/9 for Zhao under Ground I are incorporated by reference. *See* § VII.A.1, *supra*.

Dr. Bajaj’s further analysis for element 1/9[e] in paragraphs 148-149 does not resolve the deficiency in the Petition. EX2004, ¶133. Dr. Bajaj asserts that the POSITA would have found it obvious to combine Steffens and Zhao to achieve this

claim element because they both “use a similarity measure,” where Zhao’s is a Bayesian Similarity Measure. EX1004 ¶ 148. For the same reasons explained above with respect to claim 1/9[d], the POSITA would not have found it obvious to combine Steffens’s wavelet-based approach with Zhao’s contour-based BSM to achieve this claim element. *See* § VII.B.1.a, *supra*; EX2004, ¶133.

Additionally, for the reasons explained above for Ground I, Zhao does not disclose or render obvious elements 1/9[e]. *See* § VII.A.1, *supra*. Petitioners and Dr. Bajaj also do not show that Steffens discloses this claim element. EX2004, ¶134. Accordingly, Petitioners have not shown that claims 1/9[e] are disclosed or rendered obvious by the combination of Steffens and Zhao. EX2004, ¶134.

2. Claims 4/12

Claim 4/12 depends on claim 1/9, respectively. The arguments for claims 1/9 are incorporated by reference. *See* § VII.B.1, *supra*.

Petitioners argues that in view of Steffens’ reference to tracking an object “between frames,” “the POSITA 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.” Petition, 48; EX2004, ¶137. According to Petitioners, this is because the POSITA would have allegedly found it obvious that “‘between frames’ would include doing so according to a ‘prior image’ and a ‘later image.’” *Id.* at 48-49; EX2004, ¶137. Notably, Petitioners do not explain why the POSITA would have

understood tracking “between frames” in the context of Steffens to involve detecting an object in a current frame “according to a ‘prior image’ and a ‘later image’”—Petitioners impermissibly defer entirely to Dr. Bajaj’s declaration. Petition, 49; *Ashworth Bros.*, IPR2020-00594, Paper 60 at 44.

Dr. Bajaj’s analysis for claims 4/12 does not resolve the deficiency. EX2004, ¶138. Steffens detects objects using a graph matching technique that involves “determining the similarity between [an] image graph and [a] model graph.” EX1007, 7:51-67; EX2004, ¶138; *see* EX1004, ¶125. But the POSITA would have understood that Steffens’ “elastic graph matching technique” for detecting an object performs matching based on matching image graph data from only a single (current) image to a model graph. EX2004, ¶138. Steffens states that “[i]n the elastic graph matching technique, *a captured image* (block 140) is transformed into Gabor space using a wavelet transformation (block 142),” that “[*t*]he transformed image (block 144) is represented by 40 complex values, representing wavelet components, per each pixel *of the original image*,” that a model graph “is positioned over *the transformed image* at varying model node positions to locate a position of optimum similarity (block 146),” and that “[*t*]he search for the optimum similarity may be performed by positioning the model graph in the upper left hand corner of *the image*, extracting the jets at the nodes, and determining the similarity between *the image graph* and the model graph.” EX1007, 7:51-67; EX2004, ¶138. The POSITA would

have understood that the graph matching detection technique described by Steffens detects an object in a given (single) image by comparing a graph for that image with a given model graph—it does not detect the object based on graph data from other images, whether earlier or later. EX2004, ¶138.

Petitioners and Dr. Bajaj are incorrect that Steffens’ reference to tracking “between frames” discloses *detecting* an object in a current image according to measurements from later images or that it would have been obvious. Steffens confirms to the POSITA that its tracking technique operates in only the forward direction. EX2004, ¶139. Steffens states that “facial features may be tracked over consecutive frames,” but makes clear that tracking is performed from earlier nodes to later nodes, *i.e.*, in the forward direction from I_{n-1} to I_n and then from I_n to I_{n+1} . EX1007, 10:46-61; EX2004, ¶139. Steffens does not disclose or suggest tracking backwards in time or detecting an object in a current frame according to graph information for the object from a later frame. EX2004, ¶139. Steffens also does not disclose detecting an object in a current frame based on information from both a previous and a future frame. *Id.*

Furthermore, although Dr. Bajaj suggests that analyzing data in later frames was “widespread and commonly understood,” citing three non-asserted references, he provides no explanation as to how any of those references demonstrate that it was “widespread and commonly understood” to detect an object in a current image

according to measurements of the object “from a prior image and a later image relative to the current image.” EX1004, ¶158; EX2004, ¶140. Petitioners provide no such explanation in the Petition. Petition, 49.

3. Claims 8/16

Claim 8/16 depends on claim 1/9, respectively. The arguments for claims 1/9 are incorporated by reference. *See* § VII.B.1, *supra*.

Petitioners rely on their arguments for combining Steffens and Zhao with respect to elements 1/9 (d-e), as well as their arguments for claims 8/16 under Ground I for Zhao. Petition, 52. For the reasons explained above with respect to claims 1/9[d] and 1/9[e], it would not have been obvious to the POSITA to combine Steffens and Zhao to achieve those elements. *See* § VII.B.1, *supra*; EX2004, ¶143. Further, for the reasons explained above under Ground I, Zhao does not disclose or render obvious the joint probability distribution of claims 8/16. *See* § VII.A.3, *supra*; EX2004, ¶143.

4. All dependent claims are patentable over Steffens in view of Zhao.

Claims 2-8 depend on claim 1 and claims 10-16 depend on claim 9. Claims 1 and 9 are patentable over Steffens in view of Zhao for the reasons explained above. Accordingly, by their dependency on claim 1 and 9, respectively, claims 2-8 and 10-16 are patentable over Steffens in view of Zhao.

C. The challenged claims are patentable over Ozer in view of Zhao (Ground III).

1. Claims 1/9

a. Elements 1/9[e]

The arguments for claims 1/9 for Zhao under Ground I are incorporated by reference. *See* § VII.A.1, *supra*.

Petitioners rely on their Ground I arguments that this element was allegedly “disclosed or at least rendered obvious by Zhao.” Petition, 60; EX2004, ¶144. For the reasons explained above with respect to claim 1/9[e] under Ground I, Zhao does not disclose or render obvious “detecting the object *in any image* upon comparing the probabilities *for each image* to a threshold for detecting the object.” *See* § VII.A.1, *supra*; EX2004, ¶144. Also, Petitioners do not otherwise argue that Ozer discloses or renders obvious the limitations of this element. EX2004, ¶144. Accordingly, Petitioners have not shown that this element was disclosed or would have been rendered obvious to the POSITA by the combination of Ozer and Zhao. EX2004, ¶144.

2. Claims 4/12

Claim 4/12 depends on claim 1/9, respectively. The arguments for claims 1/9 are incorporated by reference. *See* § VII.C.1, *supra*.

Petitioners fail to explain how Ozer’s method of “Classifying Over Multiple Frames” allegedly discloses or would have rendered obvious claims 4/12. Petition,

63. Petitioners quote Ozer’s description of its method for “Classifying Over Multiple Frames,” which, rather than detecting an *object*, describes a process of checking “*movements*,” calculating “the probabilities of the *activities*,” and “choos[ing] the pattern with the highest probability as the recognized *activity*.” *Id.* Petitioners do not explain how or why Ozer’s activity classification discloses or would have rendered obvious claims 4/12. EX2004, ¶147. The Petition thus fails on this ground.

Petitioners again impermissibly defer entirely to Dr. Bajaj’s declaration, which is improper. *Id. Ashworth Bros.*, IPR2020-00594, Paper 60 at 44. Dr. Bajaj’s analysis for claims 4/12 does not resolve the deficiency. EX2004, ¶148. Ozer makes clear that its technique involves two separate processes, specifically, “a method with two levels, namely low and high levels, is used,” where “[t]he low-level part performs object detection and extracts parameters for the abstract graph representation of the image being processed,” and “[t]he high level part uses dynamic programming to determine the activities of the object parts, and uses a distance classifier to detect specific activities.” EX1008, 3:55-64; EX2004, ¶148.

The POSITA would have understood that Ozer’s object detection is based on information in a single image, not prior or later images. EX2004, ¶149. Ozer teaches that a “video capturer” element “converts the frames sent from the camera from analog format to digital format and stores the color components of *the current frame*

in three different buffers,” which is used to “classify the object regions.” EX1008, 5:10-20; EX2004, ¶149. Ozer’s detection process uses that data from the current frame to extract “foreground regions” of the frame and segment “the object of interest” into parts “based on the combination of color components and statistical shape features.” EX1008, 6:52-58; EX2004, ¶149. Additionally, Ozer states, “[l]ow-level part performs object detection and extracts parameters for the abstract graph representation *of the frame being processed* in real time.” EX1008, 3:65-67; EX2004, ¶149. Furthermore, Ozer states that “this system builds a model of the object of interest *on each video frame*.” EX1008, 6:19-24; EX2004, ¶149. These teachings of Ozer confirm to the POSITA that Ozer’s detection process uses data extracted from only a current frame and does not disclose or suggest detecting an object in the current frame based on parameters extracted from prior or later frames. EX2004, ¶149.

Dr. Bajaj’s (and Petitioners’) reliance on Ozer’s “Classifying Over Multiple Frames” is misplaced, as that part of the process is the separate “high-level” process for classifying “activities,” not the object. EX1008, 3:55-64, 4:44-53 (“The high-level part ... classifies activity patterns” and detects “specific gestures and activities.”); EX2004, ¶150. As Ozer states, only “[a]fter the detection of the object parts [is] the system ready [] to recognize the activities of each object part and the overall activity of the object.” EX1008, 4:35-37; EX2004, ¶150. From this the

POSITA would understand that Ozer's object classification step is for detecting an object, *see* EX2006, 41:8-20, 43:7-44:7, whereas Ozer's subsequent activity tracking step for classifying activities is not part of Ozer's object detection process. EX2004, ¶150. Indeed, the "[o]utput" of detection is produced before activity classification even begins, which confirms that the activity classifier does not result in basing object detection in a current frame on data extracted from prior or later frames. EX1008, 7:34-45; EX2004, ¶150.

Although Dr. Bajaj suggests that analyzing data in later frames was "widespread and commonly understood," citing three non-asserted references, he provides no explanation as to how any of those references demonstrate that it was "widespread and commonly understood" to detect an object in a current image according to measurements of the object "from a prior image and a later image relative to the current image." EX1004, ¶194; EX2004, ¶151.

3. Claims 8/16

Claim 8/16 depends on claim 1/9, respectively. The arguments for claims 1/9 are incorporated by reference. *See* § VII.C.1, *supra*.

Petitioners rely on their arguments for combining Ozer and Zhao, as well as their arguments for claims 8/16 under Ground I for Zhao. Petition, 65. For the reasons explained above with respect to claims 1/9[d] and 1/9[e], it would not have been obvious to the POSITA to combine Steffens and Zhao. *See* § VII.C.1, *supra*;

EX2004, ¶154. Further, for the reasons explained above under Ground I, Zhao does not disclose or render obvious the joint probability distribution of claims 8/16. *See* § VII.A.3, *supra*; EX2004, ¶154.

4. All dependent claims are patentable over Ozer in view of Zhao.

Claims 2-8 depend on claim 1 and claims 10-16 depend on claim 9. Claims 1 and 9 are patentable over Ozer in view of Zhao for the reasons explained above. Accordingly, by their dependency on claim 1 and 9, respectively, claims 2-8 and 10-16 are patentable over Ozer in view of Zhao.

D. The challenged claims are patentable over TLP (Ground IV).

1. Petitioners fail to demonstrate that TLP qualifies as prior art.

The Petition alleges that TLP “was widely disseminated and made publicly available for the IEEE Computer Society Conference on Computer Visions and Pattern Recognition by at least April 9, 2004,” citing paragraphs 65-66 of Dr. Bajaj’s declaration. Petition, 6. Specifically, Dr. Bajaj asserts that “TLP was widely disseminated and made publicly available *through the submission process*” in which “papers are widely disseminated and made publicly available to academics *for peer review*.” EX1004, ¶66 (emphasis added). Petitioners’ and Dr. Bajaj’s theory that TLP was publicly accessible through the 2004 CVPR peer review process is their only public accessibility theory for TLP. As explained below, Petitioners fail to

show that TLP was either publicly disseminated or publicly available at the relevant time under their sole theory.

TLP itself bears no indicia of publication or dissemination (or dates thereof) and Petitioners identify none. EX1009; Petition, 12. Nor do Petitioners identify the source from where TLP was obtained. *Id.* Petitioners rely solely on evidence extrinsic to TLP, none of which satisfies the legal test. Petitioners rely on two paragraphs of Dr. Bajaj’s declaration that purportedly describe the paper submission and peer review process for the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (“2004 CVPR”). Petition, 12; EX1004, ¶¶65-66. Dr. Bajaj’s testimony, however, fails to show that TLP was disseminated or publicly available before May 27, 2004, as part of the 2004 CVPR peer review process or otherwise.

Dr. Bajaj asserts that papers for the 2004 CVPR were due by “November 19, 2003.” EX1004, ¶66. Dr. Bajaj asserts, without any factual support, that “[a]fter submission, the papers are widely disseminated and made publicly available to academics for peer review,” and that “TLP was widely disseminated and made publicly available through the submission process, and no later than April 9, 2004.” EX1004, ¶66. However, Dr. Bajaj’s unsupported assertion that TLP was “widely disseminated” and made “publicly available” through submission and peer review, if accepted, is insufficient to meet the legal standard.

The Federal Circuit has found that submitting a paper to a “peer-review committee” for a conference “suggest[s] an *absence* of actual public accessibility” and “no intent to publicize” where there is no evidence that the paper “was accessible to anyone other than the peer-review committee.” *SRI Int’l, Inc. v. Internet Sec. Sys.*, 511 F.3d 1186, 1197 (Fed. Cir. 2008). In *SRI*, a paper called the “Live Traffic paper” was submitted to a conference by placing it on an open FTP server for retrieval. *Id.* at 1190. In finding the paper not publicly accessible, the court relied on the fact that that the Live Traffic paper “was still subject to pre-publication review” and was posted on the FTP server solely to facilitate peer review in preparation for later publication.” *Id.* at 1197. The court also found that “the peer-review feature also suggests no intent to publicize,” and that “without additional evidence as to the details of the [paper’s] accessibility,” the evidence was insufficient to demonstrate public accessibility. *Id.* at 1197-98. The court further found that the papers’ submission “does not evidence that the Live Traffic paper was accessible to anyone other than the peer-review committee, thus further suggesting an absence of actual public accessibility.” *Id.* at 1197.

TLP was allegedly submitted to the 2004 CVPR peer review committee for pre-publication review, but, as in *SRI*, there is no evidence that TLP was accessible to anyone other than the peer review committee. EX1004, ¶66; EX1010. Dr. Bajaj’s conclusory assertion that TLP was “widely disseminated” and “publicly available to

academics *for peer review*” lacks any explanation of how TLP was purportedly disseminated to anyone outside the peer review committee. EX1004, ¶66; *SRI Int’l*, 511 F.3d at 1196-97. The insufficiency of Petitioners’ evidence in this case is even clearer than in *SRI*. Whereas the Live Traffic paper in *SRI* was “‘posted’ on an open FTP server and might have been available to anyone with FTP know-how and knowledge of the EMERALD subdirectory” but was found not to be publicly accessible, *SRI Int’l*, 511 F.3d at 1197, here Petitioners provide no evidence (nor do they assert) that anyone other than the 2004 CVPR peer review committee would have been capable of accessing the TLP submission before May 27, 2004. EX1004, ¶66. More so than in *SRI*, this “suggests no intent to publicize” TLP and “an absence of actual public accessibility” of TLP. *SRI Int’l*, 511 F.3d at 1197.

Petitioners’ assertion that TLP “was disclosed during prosecution” and “is cited as prior art on the face of the ’980 Patent and its provisional application” also fails to show public accessibility. Petition, 6. In *SRI*, the “Live Traffic paper” was cited in an information disclosure statement of the patents-in-suit during prosecution. *SRI Int’l*, 511 F.3d at 1190. The paper was nevertheless found not to qualify as prior art. *See id.* at 1194-98. Moreover, it is well settled that a patentee’s inclusion of a reference in an information disclosure statement does not convert it into printed publication prior art. *ResQNet.com, Inc. v. Lansa, Inc.*, 594 F.3d 860, 866 (Fed. Cir. 2010); and *Abbott Labs. v. Baxter Pharm. Prods., Inc.*,

334 F.3d 1274, 1279 (Fed. Cir. 2003); *see Stamps.com Inc. v. Endicia, Inc.*, 437 F. App'x 897, 917 (Fed. Cir. 2011) (Bryson, J, concurring in part). Here, the information disclosure statement listing TLP includes the date of “2004” (no month or day), which does not demonstrate public availability before the '980 patent's undisputed priority date of May 27, 2004. *See* EX1002, 40. Moreover, the information disclosure statement associates the “2004” date with “CVPR,” but the 2004 CVPR began after the patent's priority date on June 27, 2004. EX1010.

Petitioners have failed to meet their burden of proving that TLP qualifies as prior art to the '980 patent. Accordingly, Petitioners have failed to demonstrate that any challenged claim is unpatentable over TLP.

2. Claims 1/9

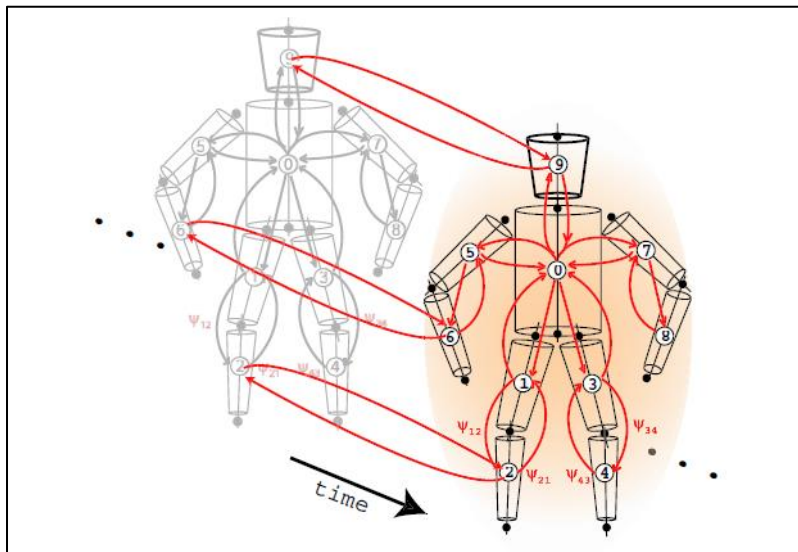
In all cases, the TLP-based invalidity grounds are unpersuasive.

a. Elements 1/9[d]

Petitioners argue that TLP “disclosed this element or at least rendered it obvious” based on TLP's “inference algorithm,” which defines “ $\phi_i(X_i)$ to be the likelihood of observing the image measurements conditioned on the pose of limb i .” Petition, 70. Petitioners argue that this “probabilistic method is **equivalent** to determining a probability an object is in each image.” Petition, 71. But Petitioners provide no explanation as to how this “probabilistic method,” which determines a “**limb** likelihood” (*i.e.*, a **component** likelihood), is “equivalent” to determining the

probability that the **object** is in each image. EX2004, ¶155; EX1009, 4; *Arendi*, 832 F.3d at 1361; *Van Os*, 844 F.3d at 1361; *Univ. of Strathclyde*, 17 F.4th at 160; *Sanofi-Synthelabo*, 550 F.3d at 1086.

Regardless, Petitioners are incorrect, as TLP’s image likelihood model does not determine a probability that the “object” (*i.e.*, human) is in each image, only its constituent components. EX2004, ¶156. Indeed, TLP makes clear that in its model, “[e]ach node in the graphical model represents ... a **limb** (Figure 1),” *i.e.*, a component. EX1009, 1; EX2004, ¶156; EX2006, 119:11-17, 120:9-121:3.



EX1009 at 1

The POSITA would have understood that TLP’s model does not include nodes that represent the object (*i.e.*, the human, *see* Petition, 66), only components referred to as “limbs” or “parts.” EX2004, ¶156. TLP teaches that “the body is represented by a graphical model in which each graph node corresponds to a body **part** (upper

leg, torso, etc.),” where “[e]dges in the graphical model correspond to position and angle relationships between adjacent body *parts* in space and time, as illustrated in Figure 1.” EX1009, 2; EX2004, ¶156. However, there is no “body” node. EX2004, ¶156. TLP states that “[d]irected edges between nodes represent statistical dependencies and these constraints between *limbs* are used to form messages that are sent to neighboring nodes in space and time.” EX1009, 1; EX2004, ¶156. Accordingly, when TLP states that “each node has an associated likelihood defined over a rich set of image cues using a learned Gibbs model,” the POSITA would have understood TLP as referring to each limb or part (not the human) having an associated likelihood. EX1009, 1-2; EX2004, ¶156. This would have been clear to the POSITA, as TLP states that “[w]e define $\phi_i(X_i)$ to be the likelihood of observing the image measurements conditioned on the pose of *limb* i ,” or a “*limb* likelihood.” EX1009, 4; EX2004, ¶156.

b. Elements 1/9[e]

i. Petitioners fail to address every limitation of elements 1/9[e] and therefore fail to show they were disclosed or rendered obvious by TLP.

Relying on their arguments for claim 1/9[d], Petitioners argue that “it would have been known to the POSITA to further compare the probabilities to a ‘threshold’ to detect the object.” Petition, 71; EX2004, ¶157. Petitioners argue that “[t]o infer belief in a model, the POSITA would have known to compare a determined

probability against a threshold.” *Id.* As with their analysis of Zhao, Steffens, and Ozer, however, Petitioners fail to consider this element as a whole. *Sanofi-Synthelabo*, 550 F.3d at 1086. Petitioners never explain why “detecting the object *in any image*” i.e., in any given image, “upon comparing the probabilities *for each image* to a threshold for detecting the object” would have been obvious in view of TLP. EX2004, ¶157. *Univ. of Strathclyde*, 17 F.4th at 160; EX2004, ¶157. Nor do Petitioners show that it would have been obvious to modify TLP to arrive at the claimed invention as a whole. EX2004, ¶157; *Arendi*, 832 F.3d at 1361; *Van Os*, 844 F.3d at 1361. The Petition thus fails on this ground.

Petitioners argue that “to infer the belief in a model, it would have been known to the POSITA *to compare a determined probability against a threshold.*” Petition, 71. Elements 1/9[e], however, require more than just comparing a determined probability against a threshold, and Petitioners do not explain how TLP discloses or renders obvious the element as a whole. *Id.*; EX2004, ¶157. Petitioners do not explain why their incomplete characterization of the claim would have allegedly been obvious—they again impermissibly defer entirely to Dr. Bajaj’s declaration. *Id.*; *Ashworth Bros.*, IPR2020-00594, Paper 60 at 44.

Dr. Bajaj’s analysis cannot and does not resolve the deficiency. Dr. Bajaj’s vague and generic assertion that “compar[ing] a determined probability against a threshold” was known in the art does not show that the POSITA would have found

the specific requirements of claim 1[e]/9[e] obvious over TLP. EX2004, ¶158. It does not show that the POSITA would have understood TLP to disclose or render obvious the particular way of using a threshold described in claim 1[e]/9[e], namely “detecting the object *in any image* upon comparing the probabilities *for each image* to a threshold for detecting the object,” or that it was otherwise known in the art. As explained below, it is not disclosed or rendered obvious by TLP. EX2004, ¶158; *Univ. of Strathclyde*, 17 F.4th at 160; *Sanofi-Synthelabo*, 550 F.3d at 1086; *Arendi*, 832 F.3d at 1361; *Van Os*, 844 F.3d at 1361.

ii. TLP does not disclose or render obvious elements 1/9[e].

TLP does not disclose or render obvious “detecting the ***object*** in any image upon comparing the probabilities for each image to a threshold for detecting the ***object***.” EX2004, ¶159. As explained above, TLP’s model includes nodes only for ***limbs*** or body ***parts***—it has no object-level nodes. See analysis of claim 1[d], *supra*. Alternatively, if these limbs or parts are considered objects, then it has no components. Accordingly, the “image likelihood model” that Petitioners rely on only determines the likelihood that a body ***part*** or ***limb*** is present in an image. EX2004, ¶159. Indeed, TLP discloses detecting only body ***parts*** for its likelihood model, and it does so with “simple eigen-template detectors” for detecting discrete body ***parts***, *i.e.*, “for the head, upper arms, and lower legs.” EX1009, 5; EX2004,

¶159. The POSITA would have understood that TLP’s model has no object level nodes, object-level likelihood model, or object-level detectors, and therefore does not disclose or render obvious detecting an **object** based on comparing the probabilities that the **object** is present in other images to a detection threshold. EX2004, ¶159.

While TLP uses non-parametric belief propagation with message passing, the POSITA would understand that the messages are between *body parts*, not between object nodes (because there are none), and that they encode only part-level information (not object-level information)—and TLP certainly does not disclose both objects and components, as all claims of the ’980 patent require. EX1009, 4-5; EX2004, ¶160. Accordingly, detection in any given image in TLP is not based on comparing probabilities that an **object** is present in other images to a detection threshold. EX2004, ¶160. The POSITA would understand this to be the case because, as explained above, the image likelihood model that TLP uses in message passing is for body **parts**, *i.e.*, it produces a “**limb** likelihood” for individual limbs (not an object likelihood for an object). EX1009, 4; EX2004, ¶160. Furthermore, the POSITA would understand that TLP’s belief propagation uses only a “**limb** proposal distribution based on local image measurements” (because there is only a limb-level likelihood model), where the local image measurements are of body **parts** (*i.e.*, head, upper arms, and lower legs) using only body **part** detectors in the form

of “simple eigen-template detectors.” EX1009, 4-5; EX2004, ¶160.

The POSITA would have further understood that TLP leads away from the claimed invention. EX2004, ¶161. TLP teaches that a “full body pose is hard to recover directly” in the event of a tracking loss, while “the location and pose of individual limbs is much easier to compute,” which TLP considers to be a “significant advantage” over known tracking techniques. EX1009, 1; EX2004, ¶161. Consistent with that teaching, TLP teaches a framework based on detecting and assembling individual limbs, as explained above, and therefore leads the POSITA away from the claimed invention towards its limb-based detection framework. EX2004, ¶161.

3. Claims 3/11

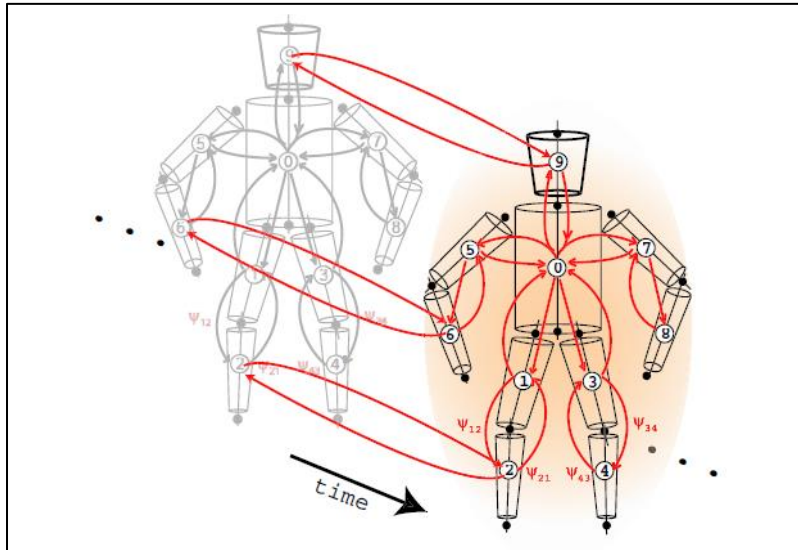
Claims 3 and 11 require “nodes corresponding to each of the collection of components **and** to the **object**.” EX2004, ¶164. Petitioners acknowledge that in TLP’s model “each graph node corresponds to a body part (upper leg, torso, etc.),” but argue that “the nodes correspond to each of the collection of components **and the object**.” Petition, 66-67, 72; EX2004, ¶164. However, Petitioners do not identify an object node in TLP—they identify only TLP’s part/limb nodes, which demonstrates (1) the absence of object nodes in TLP and (2) that TLP does not include both component and object. *Id.* Furthermore, for the reasons explained above with respect to elements 1/9[d] and 1/9[e] and below with respect to claims

8/16, TLP does not disclose or render obvious object nodes. EX2004, ¶164.

4. Claims 8/16

Petitioners do not argue that TLP discloses claims 8/16. Petition, 75-76. Petitioners argue that “TLP rendered Claims 8/16 obvious,” asserting that the joint probability distribution formulated in those claims “would have been well known to the POSITA” and is purportedly described in an article cited by TLP for its concept of non-parametric belief propagation. *Id.*; EX2004, ¶167. Petitioners’ argument fails for several reasons.

First, TLP does not disclose or render obvious a model with object-level nodes or object-level constraints, *see* § VII.D.2, *supra*, which are required by the probability distribution of claims 8/16. *See* §§ II.A and VII.A.3, *supra*; EX2004, ¶168; EX2006, 70:9-17, 80:15-21. TLP’s Figure 1 confirms the absence of object-level nodes and the absence of object-level temporal constraints (directed edges). EX2004, ¶168.



EX1009 at 1

Furthermore, Petitioners have not shown that the joint probability formulation of claims 8/16 “would have been well known” or that EX1012 discloses that formulation. See § VII.A.3, *supra*; EX2004, ¶¶169-170.

5. All dependent claims are patentable over TLP.

Claims 2-8 depend on claim 1 and claims 10-16 depend on claim 9. Claims 1 and 9 are patentable over TLP for the reasons explained above. Accordingly, by their dependency on claim 1 and 9, respectively, claims 2-8 and 10-16 are patentable over TLP.

VII. CONCLUSION

For the reasons stated above, every challenged claim should be found patentable.

Dated: March 13, 2024

By: /Christine E. Lehman/
Christine E. Lehman
PTO Reg. No. 38,535
Counsel for Patent Owner

CERTIFICATE OF COMPLIANCE

The undersigned hereby certifies that this Patent Owner's Response complies with the type-volume limits of 37 C.F.R. § 42.24 and consists of 13,577 words, excluding the table of contents, a table of authorities, certificate of service, this certificate of compliance, or table of exhibits, and claim listing. The undersigned has relied on the word count feature of the word-processing system used to prepare this paper (Microsoft Word).

Dated: March 13, 2024

Respectfully submitted,

/Christine E. Lehman/

Christine E. Lehman

PTO Reg. No. 38,535

Counsel for Patent Owner

CERTIFICATE OF SERVICE

The undersigned hereby certifies that a copy of the foregoing **PATENT OWNER'S RESPONSE** was served on March 13, 2024, via electronic mail to Meta_VideoLabs@kirkland.com, as agreed to by counsel for Petitioners, upon the following counsel:

W. Todd Baker
todd.baker@kirkland.com
KIRKLAND & ELLIS LLP
1301 Pennsylvania Avenue, N.W.
Washington, D.C. 20004

Ellisen Shelton Turner
ellisen.turner@kirkland.com
KIRKLAND & ELLIS LLP
2049 Century Park East
Los Angeles, CA 90067

Jonathan D. Brit
jonathan.brit@kirkland.com
KIRKLAND & ELLIS LLP
601 Lexington Avenue
New York, NY 10022

Jennifer R. Bush
jbush@fenwick.com
FENWICK & WEST LLP
801 California Street
Mountain View, CA 94041
Telephone: (650) 988-8500

Dated: March 13, 2024

/Christine E. Lehman/
Christine E. Lehman

APPENDIX A

No.	Claim Language
1[pre]	A computer implemented method for object detection comprising:
1[a]	providing a spatio-temporal model for an object to be detected;
1[b]	providing a video comprising a plurality of images including the object;
1[c]	measuring the object as a collection of components in each image;
1[d]	determining a probability that the object is in the image; and
1[e]	detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.
2	The computer implemented method of claim 1, wherein providing the spatio-temporal model for the object to be detected comprises providing detectors for each of the collection of components.
3	The computer implemented method of claim 1, wherein the spatiotemporal model is a graphical model comprising nodes corresponding to each of the collection of components and to the object.
4	The computer implemented method of claim 1, wherein determining the probability that the object is in each image comprises 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.
5	The computer implemented method of claim 1, wherein providing the spatio-temporal model for the object to be detected further comprises providing a temporal window defining a plurality of images in which measurements of components detected therein are passed to components detected in the current image.
6	The computer implemented method of claim 1, wherein determining the probability that the object is in each image comprises determining the probability for a position and a size of the object in each image.
7	The computer implemented method of claim 1, wherein the threshold is provided for the object to be detected, wherein the threshold is determined empirically.
8	The computer implemented method of claim 1, wherein a joint probability distribution for the spatio-temporal model with N components is:

	$P(X_0^O, X_0^{C_0}, X_0^{C_1}, \dots, X_0^{C_N}, \dots, X_T^O, X_T^{C_0}, X_T^{C_1}, \dots, X_T^{C_N}) =$ $\frac{1}{Z} \prod_{ij} \psi_{ij}(X_i^O, X_j^O)$ $\prod_{ik} \psi_{ik}(X_i^O, X_i^{C_k}) \prod_{ikl} \psi_{kl}(X_i^{C_k}, X_i^{C_l}) \prod_i \phi_i(X_i^O) \prod_{ik} \phi_i(X_i^{C_k}).$
9[pre]	A computer readable medium embodying a computer program to perform method steps for object detection, the method steps comprising:
9[a]	providing a spatio-temporal model for an object to be detected;
9[b]	providing a video comprising a plurality of images including the object;
9[c]	measuring the object as a collection of components in each image;
9[d]	determining a probability that the object is in the image; and
9[e]	detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.
10	The method of claim 9, wherein providing the spatio-temporal model for the object to be detected comprises providing detectors for each of the collection of components.
11	The method of claim 9, wherein the spatio-temporal model is a graphical model comprising nodes corresponding to each of the collection of components and to the object.
12	The method of claim 9, wherein determining the probability that the object is in each image comprises 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.
13	The method of claim 9, wherein providing the spatio-temporal model for the object to be detected further comprises providing a temporal window defining a plurality of images in which measurements of components detected therein are passed to components detected in the current image.
14	The method of claim 9, wherein determining the probability that the object is in each image comprises determining the probability for a position and a size of the object in each image.
15	The method of claim 9, wherein the threshold is provided for the object to be detected, wherein the threshold is determined empirically.

16

The method of claim 9, wherein a joint probability distribution for the spatio-temporal model with N components is:

$$P(X_0^O, X_0^{C_0}, X_0^{C_1}, \dots, X_0^{C_N}, \dots, X_T^O, X_T^{C_0}, X_T^{C_1}, \dots, X_T^{C_N}) =$$

$$\frac{1}{Z} \prod_{ij} \psi_{ij}(X_i^O, X_j^O) \quad \vdash$$

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