

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

**DECLARATION OF ELI SABER, PH.D. IN SUPPORT OF
PATENT OWNER'S RESPONSE**

TABLE OF CONTENTS

I.	INTRODUCTION	6
II.	BACKGROUND AND QUALIFICATIONS	7
III.	MATERIALS CONSIDERED	12
IV.	LEGAL PRINCIPLES.....	12
A.	Claim Construction	12
B.	Person of Ordinary Skill in the Art (“POSITA”).....	13
C.	Obviousness.....	14
D.	Priority Date of the ’980 patent.....	16
V.	LEVEL OF ORDINARY SKILL	35
VI.	THE ’980 PATENT	36
VII.	CLAIM CONSTRUCTION	48
VIII.	THE ASSERTED REFERENCES	55
A.	Zhao (EX1006).....	55
1.	Contour-based detection	55
2.	TRS-invariant human body model.....	57
3.	Bayesian Similarity Measure for Human Detection	59
4.	Recursive context reasoning (“RCR”) algorithm	60
B.	Steffens (EX1007).....	64
1.	Steffens’ single-frame graph matching detection technique	64
C.	Ozer (EX1008)	67

1.	Ozer’s single-frame object detection technique with activity classification	67
D.	TLP (EX1009).....	71
1.	TLP’s body part detection technique	71
IX.	DR. BAJAJ DOES NOT SHOW THAT ANY CHALLENGED CLAIM WOULD HAVE BEEN OBVIOUS	73
A.	Dr. Bajaj does not demonstrate that Zhao (Ground I) discloses or renders obvious any challenged claim.	73
1.	Claim 1[e]/9[e] “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.”	73
2.	Claim 4/12.....	81
3.	Claim 8/16.....	84
B.	Dr. Bajaj does not demonstrate that Steffens in view of Zhao (Ground II) discloses or renders obvious any challenged claim.	88
1.	Claim 1[d]/9[d] “determining a probability that the object is in each image.”	88
2.	Claim 1[e]/9[e] “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.”	90
3.	Claim 4/12.....	92
4.	Claim 8/16.....	95
C.	Dr. Bajaj does not demonstrate that Ozer in view of Zhao (Ground III) discloses or renders obvious any challenged claim.	96
1.	Claim 1[e]/9[e] “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.”	96

2.	Claim 4/12.....	97
3.	Claim 8/16.....	100
D.	Dr. Bajaj does not demonstrate that that TLP (Ground IV) discloses or renders obvious any challenged claim.	102
1.	Claim 1[d]/9[d] “determining a probability that the object is in each image.”	102
2.	Claim 1[e]/9[e] “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.”	104
3.	Claim 3/11.....	107
4.	Claim 8/16.....	108

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

1. I have been retained as an independent expert consultant on behalf of VL Collective IP LLC (“Patent Owner”) in this *Inter Partes* Review proceeding (“IPR”) before the Patent Trial and Appeal Board (“Board” or PTAB”) of the United States Patent and Trademark Office (“PTO”) regarding U.S. Patent No. 7,436,980 (“the ’980 Patent”) (EX1001).¹ I am being compensated for the time I spend on this matter at my usual hourly rate, in addition to expenses. My compensation in this matter is in no way contingent on the nature of my findings, the presentation of my findings in testimony, or the outcome of this or any other proceeding. I have no other interest in this proceeding.

2. In particular, I have been asked to consider, and provide my opinions regarding, whether the features recited in claims 1-16 (“the Challenged Claims”) of the ’980 patent, which I understand the Petitioners (Meta Platforms, Inc. et al.) have challenged in this IPR, are disclosed by or rendered obvious in view of the alleged prior art references cited in this proceeding. In my view, the alleged prior art references do neither. This declaration is based on the information currently

¹ In this Declaration, I reference documents by exhibit numbers that, as I understand, are attached to exhibits in this *Inter Partes* Review of the ’980 Patent.

available to me. To the extent that additional information becomes available, I reserve the right to continue my investigation and study. My opinions are set forth below.

II. BACKGROUND AND QUALIFICATIONS

3. I have summarized in this section my educational background, career history, publications, and other relevant qualifications. A copy of my curriculum vitae is included as EX2005.

4. I am a Professor in the Department of Electrical and Microelectronic Engineering (EME) and at the Chester F. Carlson Center for Imaging Science at the Rochester Institute of Technology (RIT) in Rochester, New York. I also serve as the Director of the Image, Video and Computer Vision Laboratory. I joined RIT's full-time faculty in 2004. I have over thirty-four (34) years of industry and academic experience, thirty (30) of which are in the fields of Computer Vision, Digital Image/Video Processing, Pattern Recognition and Machine Learning.

5. I received a Ph. D. in Electrical Engineering from the University of Rochester in 1996. My concentration was in Signal/Image/Video Processing, Pattern Recognition, and Computer Vision. I received a Master of Science Degree in Electrical Engineering from the University of Rochester in 1992 in the same concentration as listed above, and I received a Bachelor of Science Degree in

Electrical and Computer Engineering, Summa Cum Laude, from the State University of New York at Buffalo in 1988.

6. Before becoming a full-time professor, I worked for Xerox Corporation from 1988 to 2004. During my 16 years at Xerox, I was – generally speaking – responsible for: a) delivering color management solutions, image processing innovations, architectures, and algorithms for high-end color printing systems, b) xerographic sub-systems for a variety of color products, and c) control systems for toner production facilities. Among other roles, as an Advanced Development Scientist and Manager, I established the Advanced Design Laboratory – an Imaging/Xerographics lab – and provided technical and managerial leadership for the Electrical, Imaging and Xerographics Department. As a Product Development Scientist and Manager, I led the research and development of color management algorithms and image quality metrics for various product platforms. I also led the Image Science, Analysis, and Evaluation area, with twelve to fifteen direct reports and a budget of approximately \$2 million.

7. From 1997 until 2004, I was an adjunct faculty member at the Electrical Engineering Department of the Rochester Institute of Technology and at the Electrical & Computer Engineering Department at the University of Rochester. I was responsible for teaching undergraduate and graduate coursework in signal,

image, video processing, pattern recognition, and communications and performing research in multimedia applications, pattern recognition, image understanding and color engineering.

8. Since joining RIT full time, I have been responsible for teaching several undergraduate and graduate courses in Digital Signal Processing, Digital Image Processing, Digital Video Processing, Engineering Analysis, Random Signal & Noise, Advanced Engineering Mathematics, Matrix Methods, Pattern Recognition, Communications, Modern Control Theory, and Linear Systems.

9. I proposed, developed and introduced the Digital Video Processing course and reestablished the Pattern Recognition course at the Electrical Engineering Department at RIT. These graduate courses were instrumental in providing MS/PhD students with the proper foundation and knowledge in all aspects of the subject areas in order to perform in-depth research in the field and continue to advance the state of the art.

10. The Digital Video Processing course is designed to provide a thorough and comprehensive coverage of the field. Topics covered include the mathematical representation of digital video including motion and time-varying image formation models, spatio-temporal sampling/reconstruction, sampling structure conversions, motion estimation using optical flow, block-based, pel-recursive, Bayesian and point

correspondence type methodologies, segmentation, motion tracking, filtering, restoration, super-resolution and compression.

11. The Pattern Recognition course, on the other hand, covers pertinent concepts in Bayes decision theory, linear and nonlinear classifiers, artificial neural networks, feature selection/generation/extraction and clustering.

12. I am currently working to develop a graduate course titled: “Advanced Neural Networks and Deep Learning: Theory, Mathematics and Algorithms.” The course is intended to provide an in-depth comprehensive study of the theory, mathematics and algorithms for advanced neural networks and their usage in deep learning. It is primarily targeted for upper graduate students that are planning to pursue advanced studies/research in the field. The course provides a thorough study of shallow and deep neural networks architectures and algorithms, activation functions, backpropagation, initialization, optimization strategies, etc. It then progresses to convolutional and recurrent style neural networks and, if time permits, will provide an overview of attention mechanisms/transformers and deep reinforcement learning. The course is being taught for the first time in Spring 2024.

13. My current research focuses on machine learning for object tracking/remote sensing applications; image/video processing for multimedia, military & biomedical applications; and computer vision and three-dimensional

scene reconstruction. As a principal investigator (PI) or coprincipal investigator, I have acquired almost \$6 million in research funding since joining RIT and have managed multiple government grants from the Department of Defense as well as several corporate grants from Hewlett-Packard, Lenel, and Data Physics. I am currently managing (as PI) a multiyear government grant from the Department of Defense titled “Target Detection/Tracking and Activity Recognition from Multimodal Data.” In this project, we propose to develop operational target detection/tracking techniques and activity recognition/understanding methods that can leverage multimodal data in a fusion framework using deep learning algorithms while providing accurate and near real-time performance.

14. In 2012, I was awarded the Prestigious Trustees Scholarship, the highest award at RIT with regard to research recognition.

15. I am the author or co-author of thirty-eight (38) peer-reviewed journal publications.

16. I have also authored or co-authored one hundred and two (102) conference and workshop publications and a book entitled Advanced Linear Algebra for Engineers with MATLAB, published by CRC Press in February 2009, and am a named inventor on multiple U.S. and foreign patents.

17. I am a senior member of the Institute of Electrical and Electronic Engineers and a member of the IEEE Signal Processing Society, the Electrical Engineering Honor Society, and Eta Kappa Nu.

18. A complete statement of my industrial and academic records including a listing of the above publications and patents is included with my curriculum vitae.

III. MATERIALS CONSIDERED

19. In forming the opinions in this Declaration, I reviewed the Petition (Paper 3) and its accompanying exhibits, including Dr. Bajaj's declaration (EX1004), the '980 patent, and the '980 patent's file history. A list of the materials I considered is included as Appendix A.

IV. LEGAL PRINCIPLES

20. I am not an attorney or a legal expert. My understanding of legal principles that I considered in reaching the conclusions expressed in this Declaration were provided to me by counsel. In preparing this Declaration, I have endeavored to faithfully apply these legal standards. I do not intend to offer opinions on the law. Below, I describe the legal principles I considered in reaching my conclusions.

A. Claim Construction

21. I understand that the first step in evaluating obviousness is to properly construe the claims. I further understand that, in this proceeding, the Board must determine the scope of the claims by using the same claim construction standard that

United States District Courts use. Under this standard, words of a claim are given their “ordinary” and “customary” meaning as understood by a person of skill in the art at the time of the invention. I understand that such a person looks to the claims, specification, and file history in understanding the meaning of the claim terms. I have applied this understanding in this Declaration.

22. I further understand that claims must be interpreted in the light of the specification with a view to ascertaining the invention, and that the interpretation to be given a term can only be determined and confirmed with a full understanding of what the inventors actually invented. I understand that it is proper to consider the functions of an invention in seeking to determine the meaning of particular claim language. I further understand that the purpose of the claimed invention stated in the specification can inform the meaning of a claim as understood by a person of ordinary skill in the art.

B. Person of Ordinary Skill in the Art (“POSITA”)

23. I understand that a person having ordinary skill in the art may be defined by factors such as: (a) the levels of education and experience of the inventor and other persons actively involved in the field; (b) the types of problems encountered in the field; (c) prior art solutions to those problems; (d) rapidity with which innovations are made; and (e) the sophistication of the technology.

C. Obviousness

24. I understand that the obviousness inquiry considers whether the differences between a patent claim and the prior art references would have been obvious to a person of ordinary skill in the art at the time of the invention, taking into account (1) the scope and content of the prior art, (2) the differences between the prior art and the claims, (3) the level of ordinary skill in the art, and (4) any objective indicia of non-obviousness. I further understand that a claim can be found to be obvious in light of either a single prior art reference or multiple prior art references.

25. I understand that a claim composed of several elements is not proved obvious merely by demonstrating that each of its elements was independently known in the prior art. I further understand that obviousness cannot be based on the hindsight combination of components selectively culled from the prior art to fit the parameters of the claimed invention.

26. Instead, for a claim to be obvious in light of a single prior art reference or multiple prior art references, I understand that there must be a reason either to modify the single prior art reference or to combine the multiple prior art references to achieve the claimed invention. I understand that this reason may come from a teaching, suggestion, or motivation to combine or may come from the reference or

references themselves, the knowledge or “common sense” of one skilled in the art, or from the nature of the problem to be solved; and may be explicit or implicit from the prior art as a whole. I understand that this inquiry asks whether a person of ordinary skill in the art not only could have made but would have been motivated to make the combinations of prior art to arrive at the claimed invention.

27. I further understand that a reference that teaches away from the claimed invention can preclude a determination that the reference renders the claim obvious. A reference may be said to teach away when a person of ordinary skill, upon reading the reference, would be discouraged from following the path set out in the reference, or would be led in a direction divergent from the path that was taken by the claimed invention, or if it suggests that the line of development flowing from the reference’s disclosure is unlikely to be productive of the result sought by the claimed invention. If when combined, the references would produce a seemingly inoperative device, then they teach away from their combination.

28. I understand that if the teachings of a prior art reference would lead one skilled in the art to make a modification that would render another prior art device inoperable, then such a modification would generally not be obvious. I understand that if a proposed modification would render the prior art device unsatisfactory for

its intended purpose, then the proposed modification would generally not be obvious.

D. Priority Date of the '980 patent

29. I understand that Dr. Bajaj does not dispute that the '980 patent is entitled to a priority date of May 27, 2004. EX1004 ¶33. I further understand that the Petition does not dispute that the '980 patent is entitled to a priority date of May 27, 2004. Petition at 5. Therefore, when I refer to a person of ordinary skill in the art at the time of the invention in my Declaration, I am referring to a person of ordinary skill in the art as of May 27, 2004.

30. Furthermore, as explained below, it is my opinion that that the provisional application reasonably conveys to a POSITA that the inventors had possession of the claimed subject matter at the time the provisional application was filed.

31. The provisional application conveys to a POSITA that the inventors had possession of element 1/9[pre]: “A computer implemented method for object detection.” In the Abstract, the provisional application states, “We present a probabilistic framework for component-based automatic **detection** and tracking of **objects** in images and/or video.” EX1003 at 3. The Abstract further states, that “[o]**bject detection** and tracking is then formulated as an inference over directed

loopy graph, and is solved with non-parametric belief propagation that uses a variation of particle filtering.” *Id.* It also states that “[p]roposal distributions provide ‘bottom-up’ information that is incorporated into the inference process; **this permits automatic object detection** and aids recovery from transient tracking failures.” *Id.* In section “1 Introduction,” the provisional application states that “[b]y combining the **object detection** with tracking in a unified framework we can achieve a more robust solution for both problems.” *Id.* It further states that “[b]y combining the **object detection** with tracking in a unified framework we can achieve a more robust solution for both problems.” *Id.* at 4. A POSITA would have understood that the detection methods referred to are computer-implemented methods. That understanding is further supported by the provisional’s indication that such methods are for use in computer vision processes, which are computer-implemented processes. *Id.* at 4 (“The problem of describing and recognizing categories of objects (e.g. faces, people, cars) has been an active field of **computer vision** for a considerable time.”). In view of these exemplary disclosures in the provisional application, a POSITA would have understood that the inventors were in possession of this element.

32. The provisional application conveys to a POSITA that the inventors had possession of element 1/9[a]: “providing a spatio-temporal model for an object to be

detected.” For example, in the Abstract, the provisional indicates that “We represent objects as **spatio-temporal two-layer graphical models**, where each node corresponds to an object or component of an object at a given time, and the edges correspond to the learned spatial and temporal constraints.” EX1003 at 3. In the section “1.1 Related Work,” the provisional States that “We model objects using **two-layer spatio-temporal graphical model**, and use non-parametric belief propagation to infer the position and scale of the object and components.” *Id.* at 4. In section “2 Graphical Object Models,” the provisional states that “Following the framework of [1, 2] we model an object as a **spatio-temporal directed graphical model**.” *Id.* at 5. In section “2.1 Building a Graphical Model,” the provisional also states that “We can extend this model over time to an arbitrary length **spatio-temporal graphical model** as shown in Fig. 2c and 2d.”). *Id.* at 6. In the caption of Figure 2, the provisional states that “**Spatio-temporal models** are obtained by replicating a spatial model along the temporal domain to a w-length window and then connecting the second-layer nodes across time.” *Id.* at 7. Twice on page 10, the provisional states that “We use a 3-frame **spatio-temporal model**.” *Id.* at 10. In view of these exemplary disclosures in the provisional application, a POSITA would have understood that the inventors were in possession of this element.

33. The provisional application conveys to a POSITA that the inventors had possession of element 1/9[b]: “providing a video comprising a plurality of images including the object.” For example, the Abstract indicate that “We present a probabilistic framework for component-based automatic detection and tracking of objects in **images and/or video**.” EX1003 at 3. The Abstract further states that “We illustrate our method by detecting and tracking two classes of objects, cars and pedestrians, in **video sequences** collected using a single uncalibrated car mounted camera.” *Id.* In section “3 Experiments,” the provisional describes “the results for the **55 frame long sequence**” and indicates that “[w]e use a **3-frame spatio-temporal model** that is shifted in a tracking centric (i.e. only the new frame is reinitialized using the proposal when the window is shifted) way through time.” *Id.* at 10. The provisional also describes how “We have also tested the system on **sequences of moving pedestrians**.” *Id.* In view of these exemplary disclosures in the provisional application, a POSITA would have understood that the inventors were in possession of this element.

34. The provisional application conveys to a POSITA that the inventors had possession of element 1/9[c]: “measuring the object as a collection of components in each image.” In the Abstract, the provisional states that “We present a probabilistic framework for **component-based** automatic detection and tracking of

objects in images and/or video.” EX1003 at 3. In the section “1 Introduction,” the provisional explains that “**Component-based** detection facilitates detection of complex articulated objects, as well as helps in handling partial object occlusions or local illumination changes.” *Id.* at 3-4. In the section “1.1 Related Work,” the provisional states that “[u]nlike all previous approaches however we provide a automatic probabilistic framework that **explicitly models objects using components**, integrates temporal information, and allows for simultaneous tracking.” *Id.* at 4. It further states that “We **model objects using two-layer spatio-temporal graphical model**, and use non-parametric belief propagation to infer the position and scale of the object and components.” *Id.* at 5. In the section “2 Graphical Object Models,” the provisional states that “[f]ollowing the framework of [1, 2] we model an object as a spatio-temporal directed graphical model” where “[e]ach node in this graph represents an object or a component of an object at time t.” *Id.* at 5. In section “2.1 Building a Graphical Model,” the provisional states that “The **first layer** corresponds to **components of the object** and is loosely connected (while maintaining connectivity, we attempt to keep the degree of each node relatively small for efficiency). The **second layer** corresponds to an **entire appearance model and is connected to all constituent components**.” *Id.* at 6. In section “2.3 Non-parametric BP,” the provisional states that “once we **integrate**

information from all components in the second layer node, we can get a very reliable estimate for the overall object size.” *Id.* at 8. In view of these exemplary disclosures in the provisional application, a POSITA would have understood that the inventors were in possession of this element.

35. The provisional application conveys to a POSITA that the inventors had possession of element 1/9[d]: “determining a probability that the object is in the image.” In the sections Abstract and “1 Introduction,” the provisional indicates that “We present a probabilistic framework for automatic component-based detection and tracking of objects in images and/or video” and that “by combining the object detection with tracking in a unified framework we can achieve a more robust solution for both problems.” *Id.* at 3. The Introduction section further states that “Each node also has an associated AdaBoost detector that serves as a basis for the local image likelihood and the proposal process.” *Id.* at 3. In section “2.3 Non-parametric belief propagation,” the provisional states that “ $\phi_i(X_i)$ is the local evidence (or likelihood) associated with the node i .” *Id.* at 9. In section “2.4 AdaBoost Image Likelihoods,” the provisional states that “[t]he inference algorithm outlined in the previous section requires probabilistic image likelihood model $\phi_i(X_i)$,” which is “the likelihood of observing the $[]$ image conditioned on the state X_i of node i .” *Id.* at 9. The provisional further states that “The output of the AdaBoost classifier is a confidence

$h_k(I)$ that the given pattern I is of the desired class. It is customary to consider object present if $h_k(I) \geq \frac{1}{2} \sum_{k=1}^K \alpha_k$.” *Id.* at 10. In section “2.5 Proposal Process,” the provisional states that “[t]o form a proposal distribution for a component, we run corresponding AdaBoost detector over an image at a number of scales to produce a set of detection results $\mathfrak{R}$ that score above the $\frac{1}{2} \sum_{k=1}^K \alpha_k$ threshold.” *Id.* at 10. In view of these exemplary disclosures in the provisional application, a POSITA would have understood that the inventors were in possession of this element.

36. The provisional application conveys to a POSITA that the inventors had possession of element 1/9[e]: “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.” In addition to the disclosures for element 1/9[d], the provisional application further states in the Abstract that “Object detection and tracking is then formulated as an inference over directed loopy graph, and is solved with non-parametric belief propagation that uses a variation of particle filtering. This type of an object model allows object-detection to make use of temporal consistency (over an arbitrary size temporal window), and provides mechanism for simultaneous robust tracking of the object.” EX1003 at 3. In section “1 Introduction,” the provisional application further states that “object detection can be made more reliable by considering the consistency of the detection over time,” and that “[o]bject detection and tracking are formulated as inference in

a two-layer graphical model in which object components are connected via learned spatial and temporal probabilistic constraints.” *Id.* at 4. In section “2.3 Non-parametric belief propagation,” the provisional application states that “Inferring the state of the object and its components in our framework is defined as estimating a belief in a graphical model,” where “We use a form of non-parametric belief propagation,” and “The approach is a generalization of particle filtering [15] which allows inference over an arbitrary graphs rather than a simple chain.” *Id.* at 8. “In this generalization the ‘message’ used in a standard belief propagation is approximated with a kernel density ... and the conditional distribution used in standard particle filtering is replaced by product of incoming messages,” where “the product over all incoming messages into the node tends to produce a very tight distribution in the state space,” where “we integrate information from all components in the second-layer node,” and send messages between “the set of neighbors of node i ” based on “the local evidence (or likelihood) associated with the node i ” ($\phi_i(X_i)$) and “the potential designating the compatibility between the states of node i and j .” *Id.* at 8-9. The provisional application further states that “We use w -frame windowed smoothing algorithm for out tracking and object detection, where w is an odd integer ≥ 1 (see Fig. 2cd). There are two ways one can do windowed smoothing, what we call an object-detection centric way, and the tracking

centric way. In an object-detection centric way, we re-initialize all nodes every time we shift a window, hence the temporal integration is only applied in the window of size w . In the tracking centric way we only initialize the nodes associated with a new frame, which tends to force temporal consistency from before $t-w(w-1)/2$.” *Id.* at 9. In section “2.4 AdaBoost Image Likelihoods,” the provisional further states that “The output of the AdaBoost classifier is a confidence $h_k(I)$ that the given pattern I is of the desired class. It is customary to consider object present if $h_k(I) \geq \frac{1}{2} \sum_{k=1}^K \alpha_k$.” *Id.* at 10. In section “2.5 Proposal Process,” the provisional states that “[t]o form a proposal distribution for a component, we run corresponding AdaBoost detector over an image at a number of scales to produce a set of detection results $\mathfrak{R}$ that score above the $\frac{1}{2} \sum_{k=1}^K \alpha_k$ threshold.” *Id.* at 10. In view of these exemplary disclosures in the provisional application, a POSITA would have understood that the inventors were in possession of this element.

37. The provisional application conveys to a POSITA that the inventors had possession of claim 2/10: “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.” In the sections Abstract, the provisional indicates that “AdaBoost detectors are used to define the likelihood and form proposal distributions for components. Proposal distributions

provide ‘bottom-up’ information that is incorporated into the inference process; this permits automatic object detection and aids recovery from transient tracking failures.” *Id.* at 3. The Introduction section further states that “Each node also has an associated AdaBoost detector that serves as a basis for the local image likelihood and the proposal process.” *Id.* at 3. In section “2.3 Non-parametric belief propagation,” the provisional states that “ $\phi_i(X_i)$ is the local evidence (or likelihood) associated with the node i .” *Id.* at 9. In section “2.4 AdaBoost Image Likelihoods,” the provisional states that “The inference algorithm outlined in the previous section requires probabilistic image likelihood model $\phi_i(X_i)$,” which is “the likelihood of observing the I image conditioned on the state X_i of node i .” *Id.* at 9. The provisional further states that “we train boosting based detectors [7] for each component.” *Id.* at 9. The provisional further states that “The output of the AdaBoost classifier is a confidence $h_k(I)$ that the given pattern I is of the desired class. It is customary to consider object present if $h_k(I) \geq \frac{1}{2} \sum_{k=1}^K \alpha_k$.” *Id.* at 10. In section “2.5 Proposal Process,” the provisional states that “[t]o form a proposal distribution for a component, we run corresponding AdaBoost detector over an image at a number of scales to produce a set of detection results $\mathfrak{R}$ that score above the $\frac{1}{2} \sum_{k=1}^K \alpha_k$ threshold.” *Id.* at 10. In view of these exemplary disclosures in the

provisional application, a POSITA would have understood that the inventors were in possession of this element.

38. The provisional application conveys to a POSITA that the inventors had possession of claim 3/11: “The computer implemented method of claim 1, wherein the spatio-temporal model is a graphical model comprising nodes corresponding to each of the collection of components and to the object.” For example, in the Abstract, the provisional states that “We represent objects as spatio-temporal two-layer graphical models, where each node corresponds to an object or component of an object at a given time,” and that “[t]he two layer structure of the graphical model allows inference over the entire object as well as individual components.” EX1003 at 3. In the section “1 Introduction,” the provisional explains that “[e]ach node in a graphical model corresponds to a position and scale of the component (for the first layer) or an object as a whole (for the second layer) in an image at a given time instant.” *Id.* at 4. In the section “1.1 Related Work,” the provisional states that “[u]nlike all previous approaches however we provide a automatic probabilistic framework that explicitly models objects using components, integrates temporal information, and allows for simultaneous tracking.” *Id.* at 4. It further states that “We model objects using two-layer spatio-temporal graphical model, and use non-parametric belief propagation to infer the position and scale of the object and

components.” *Id.* at 5. In the section “2 Graphical Object Models,” the provisional states that “[f]ollowing the framework of [1, 2] we model an object as a spatio-temporal directed graphical model” where “[e]ach node in this graph represents an object or a component of an object at time t.” *Id.* at 5. In section “2.1 Building a Graphical Model,” the provisional states that “The first layer corresponds to components of the object and is loosely connected (while maintaining connectivity, we attempt to keep the degree of each node relatively small for efficiency). The second layer corresponds to an entire appearance model and is connected to all constituent components.” *Id.* at 6. In section “2.3 Non-parametric BP,” the provisional states that “once we integrate information from all components in the second layer node, we can get a very reliable estimate for the overall object size.” *Id.* at 8. Figure 2c and 2d show examples of a two-layer graphical model with object nodes and component nodes. *Id.* at 7. In view of these exemplary disclosures in the provisional application, a POSITA would have understood that the inventors were in possession of this element.

39. The provisional application conveys to a POSITA that the inventors had possession of claim 4/12: “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.” For example, in section “1.1 Related work,” the provisional states that “Unlike our approach however their system does not integrate component-based object detection, and is limited to temporal propagation in only one direction (forward in time).” EX1003 at 5. In section “2.1 Building a Graphical Model,” the provisional states “We can extend this model over time to an arbitrary length spatio-temporal graphical model as shown in Fig. 2c and 2d.” *Id.* at 6. In Figure 2, the provisional shows objects and components modeled at $O_{t-(w-1)/2}$ (for the object at an earlier time), O_t (for the object at a current time), and $O_{t+(w-1)/2}$ (for the object at a later time). *Id.* at 7. In section “2.3 Non-parametric belief propagation,” the provisional states that the disclosed approach “allows inference over an arbitrary graphs rather than a simple chain,” and that “[w]e use w-frame windowed smoothing algorithm for our tracking and object detection where w is an odd integer ≥ 1 (see Fig. 2cd).” *Id.* at 8-9. In section “3.1 Vehicle Spatio-Temporal Tracking,” the provisional states that “We use a 3-frame spatio-temporal model that is shifted in a tracking centric (i.e. only the new frame is reinitialized using the proposal when the window is shifted) way through time.” *Id.* at 10. In view of these exemplary disclosures in the provisional application, a POSITA would have understood that the inventors were in possession of this element.

40. The provisional application conveys to a POSITA that the inventors had possession of claim 5/13: “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.” For example, the Abstract states that “This type of an object model allows object-detection to make use of temporal consistency (over an arbitrary size temporal window), and provides mechanism for simultaneous robust tracking of the object.” EX1003 at 3. In section “1.1 Related work,” the provisional states that “Unlike our approach however their system does not integrate component-based object detection, and is limited to temporal propagation in only one direction (forward in time).” EX1003 at 5. In section “2.1 Building a Graphical Model,” the provisional states “We can extend this model over time to an arbitrary length spatio-temporal graphical model as shown in Fig. 2c and 2d.” *Id.* at 6. In Figure 2, the provisional shows objects and components modeled at $O_{t-(w-1)/2}$ (for the object at an earlier time), O_t (for the object at a current time), and $O_{t+(w-1)/2}$ (for the object at a later time). *Id.* at 7. In section “2.3 Non-parametric belief propagation,” the provisional states that the disclosed approach “allows inference over an arbitrary graphs rather than a simple chain,” and that “[w]e use w-frame windowed smoothing

algorithm for our tracking and object detection where w is an odd integer ≥ 1 (see Fig. 2cd).” *Id.* at 8-9. In section “3.1 Vehicle Spatio-Temporal Tracking,” the provisional states that “We use a 3-frame spatio-temporal model that is shifted in a tracking centric (i.e. only the new frame is reinitialized using the proposal when the window is shifted) way through time.” *Id.* at 10. In the captions of Figures 4 and 5, the provisional states that “[t]he detection and tracking was conducted using a 3-frame smoothing window.” *Id.* at 11. In view of these exemplary disclosures in the provisional application, a POSITA would have understood that the inventors were in possession of this element.

41. The provisional application conveys to a POSITA that the inventors had possession of claim 6/14: “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.” For example, in the Abstract, the provisional states that “We represent objects as spatio-temporal two-layer graphical models, where each node corresponds to an object or component of an object at a given time.” EX1003 at 3. In section “1 Introduction,” the provisional states that “Each node in a graphical model corresponds to a position and scale of the component (for the first layer) or an object as a whole (for the second layer) in an image at a given time instant.” *Id.* at 4. In section “1.1 Related work,”

the provisional states that “We model objects using two-layer spatio-temporal graphical model, and use non-parametric belief propagation to infer the position and scale of the object and components.” *Id.* at 5. In section “2 Graphical Object Models,” the provisional states that “Each node has an associated state vector $X^T = (x,y,s)$ defining the component’s position and scale within an image.” *Id.* at 5. In section “2.3 Non-parametric BP,” the provisional states that “any given component of a vehicle is incapable of estimating the height of the vehicle reliably, however once we integrate information from all components in the second-layer node, we can get a very reliable estimate for the overall object size.” *Id.* at 8. In section “2.5 Proposal Process,” the provisional indicates that the disclosed proposal process “tends to be manageable for the entire appearance model, it is usually large for non-specific component detectors (a few thousand positions can easily be found running a detector at every pixel for 9 consecutive scales).” *Id.* at 10. In view of these exemplary disclosures in the provisional application, a POSITA would have understood that the inventors were in possession of this element.

42. The provisional application conveys to a POSITA that the inventors had possession of claim 7/15: “The computer implemented method of claim 1, wherein the threshold is provided for the object to be detected, wherein the threshold is determined empirically.” For example, in the Abstract, the provisional states that

“the edges correspond to the learned spatial and temporal constraints.” EX1003 at 3. In section “1 Introduction,” the provisional state that “Object detection and tracking are formulated as inference in a two-layer graphical model in which object components are connected via learned spatial and temporal probabilistic constraints,” and that “Directed edges between nodes represent statistical dependencies between components, and are learned using standard machine learning methods.” *Id.* at 4. In section “2 Graphical Object Models,” the provisional states that “The direct edges correspond to the learned spatial and temporal constraints.” *Id.* at 5. In section “2.2 Learning Spatial and Temporal Constraints,” the provisional states that “Given a set of labeled images, with each component associated with the reference point, we use standard iterative Expectation-Maximization (EM) algorithm with K-means initialization to learn $F_{ijm}(\bullet)$ and $G_{ijm}(\bullet)$,” and that “[e]xamples of the learned conditional distributions can be seen in Fig. 3cde.” *Id.* at 6-8. In section “2.4 AdaBoost Image Likelihoods,” the provisional states that “we train boosting based detectors [7] for each component,” and “[w]e expect that training a cascade would improve the performance.” *Id.* at 9. In view of these exemplary disclosures in the provisional application, a POSITA would have understood that the inventors were in possession of this element.

43. The provisional application conveys to a POSITA that the inventors had possession of claim 8/16: “The computer implemented method of claim 1, wherein a joint probability distribution for the spatio-temporal model with N components is:

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

In the Abstract, the provisional states “We present a probabilistic framework for component-based automatic detection and tracking of objects in images and/or video. We represent objects as spatio-temporal two-layer graphical models, where each node corresponds to an object or component of an object at a given time, and the edges correspond to the learned spatial and temporal constraints.” EX1003 at 3. A POSITA would understand that the joint probability distribution in claims 8/16 describes this type of model. In section “1 Introduction,” the provisional states that “Object detection and tracking are formulated as inference in a two-layer graphical model in which object components are connected via learned spatial and temporal probabilistic constraints. Each node in a graphical model corresponds to a position and scale of the component (for the first layer) or an object as a whole (for the second

layer) in an image at a given time instant. Each node also has an associated AdaBoost detector that serves as a basis for the local image likelihood and the proposal process. Directed edges between nodes represent statistical dependencies between components, and are learned using standard machine learning methods.” *Id.* at 4. A POSITA would understand that the joint probability distribution in claims 8/16 describes this type of model. In section “2 Graphical Object Models,” the provisional states that “we model an object as a spatio-temporal directed graphical model. Each node in this graph represents an object or a component of an object at time t . The direct edges correspond to the learned spatial and temporal constraints. Each node has an associated state vector $X^T = (x, y, s)$ defining the component's position and scale within an image.” *Id.* at 5. The provisional further states that “[e]ach directed edge between components i and j has an associated potential function $\psi_{ij}(X_i, X_j)$ that encodes the compatibility between pairs of node states. The potential $\psi_{ij}(X_i, X_j)$ is modeled using a mixture of M_{ij} Gaussians.” *Id.* at 6. A POSITA would understand that the joint probability distribution in claims 8/16 describes this type of model. In section “2.4 AdaBoost Image Likelihoods,” the provisional application states that “[t]he inference algorithm outlined in the previous section requires probabilistic image likelihood model $\phi_i(X_i)$,” which is “the likelihood of observing the [] image conditioned on the state X_i of node i .” *Id.* at 9.

In view of these exemplary disclosures in the provisional application, a POSITA would have understood that the inventors were in possession of this element.

V. LEVEL OF ORDINARY SKILL

44. Dr. Bajaj stated that the level of ordinary skill of a POSITA would have been the following:

A person of ordinary skill in the art (“POSITA”) in the field of the ’980 Patent, as of its earliest possible priority date of May 27, 2004, would have been someone with a bachelor’s degree in computer science, computer engineering, computer vision or visualization, or an equivalent field, and approximately two years of experience in software development for computer vision applications. Additional education might compensate for less experience, and vice-versa.

EX1004 ¶16.

45. I met the level of ordinary skill set forth by Dr. Bajaj before May 27, 2004, and I am familiar with the degree of knowledge that a POSITA would have had and how a POSITA would have understood and interpreted the claims and specification of the ’980 patent.

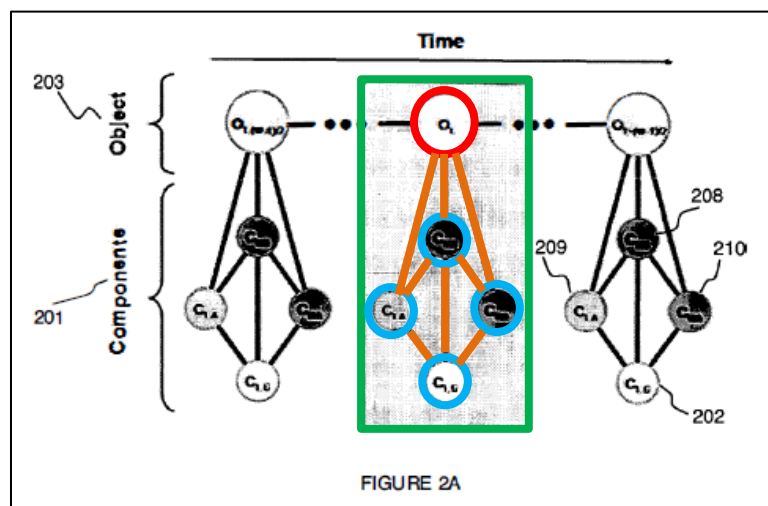
VI. THE '980 PATENT

46. The '980 patent indicates that object detection is an important problem related to computer vision application. EX1001 at 1:15-21. And while the patent suggests that it may have been relatively “common to represent objects as collections of features,” there was “a large variation in how many features are needed and how these feature are detected and represented.” EX1001 at 1:15-21. Given that landscape, the inventors of the '980 patent identified a need for a new way of “implementing component detection and performing inference over space and time.” EX1001 at 1:22-24.

47. The inventors therefore disclosed “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.” EX1001 at 2:43-47. Under the inventors’ new object detection framework, “[o]bject detection considers the consistency of the detection over time.” EX1001 at 2:49-50.

48. To achieve this, the '980 patent teaches a new technique in which object detection is formulated as an inference in a two-layer graphical model using a spatio-temporal framework. EX1001 at 2:55-3:2; *see* EX2006, 67:5-10, 44:15-19, 64:2-5. To this effect, as shown in Figure 2A (annotated below), the '980 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 at 2:55-3:2 (annotated). The '980 patent explains that “[f]or 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 parts, e.g., 202,” and “[t]he coarse, object, layer 203 corresponds to **an entire appearance model of the object** and is **connected to all constituent components**, e.g., 202.” EX1001 at 4:18-24; see EX2006, 82:14-83:5. This indicates to a POSITA that the patent’s object model explicitly models the object (as a distinct entity from its components) as well as the object’s corresponding components in the same model for a given point in time.



EX1001, Fig. 2A (annotated)

49. The '980 patent further explains that in its graphical object model, “[d]irected edges between nodes represent learned **spatial [probabilistic**

constraints] and **temporal probabilistic constraints.**” EX1001 at 2:59-60 (annotated). As shown below, the “edges” in the graph are the lines that connect nodes to each other, which, as the patent explains, correspond to spatial or temporal constraints. EX1001 at 2:59-60, 4:15-17 (“spatial and/or temporal constraints corresponding to the edges in a graph”). “To integrate **temporal constraints**” the ’980 patent teaches that “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 at 4:34-45 (annotated), Figs. 2A (annotated below). The ’980 patent’s provisional application similarly states that “[s]patio-temporal models are obtained by replicating a spatial model along the temporal domain to a w-length window and then connecting the second-layer nodes across time.” EX1003 at 7. This indicates to a POSITA that the patent’s spatio-temporal object model explicitly models the object (as a distinct entity from its components), as well as the object’s corresponding components, at multiple points in time. See EX1001 at 4:37-42 (“Having a two-layer graphical model allows the inference process to reason explicitly about the object as a whole, e.g., 203 ... over time given the overall object appearance.”).

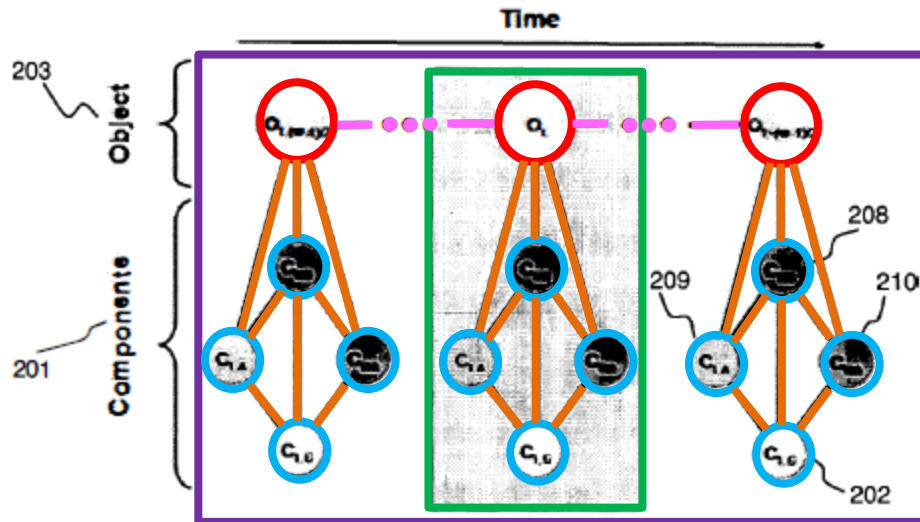
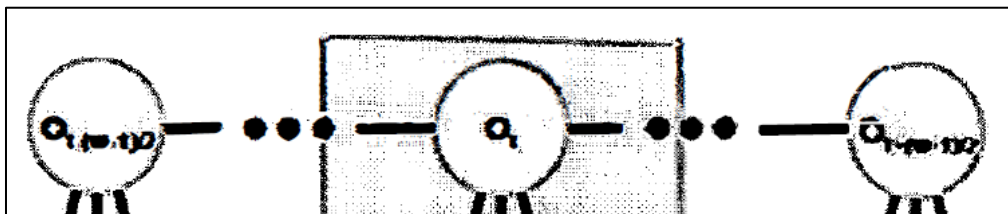


FIGURE 2A

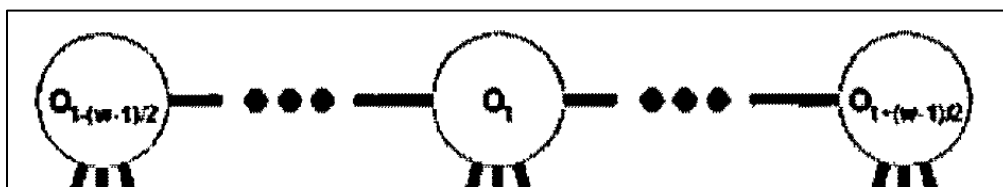
EX1001, Figs. 2A (annotated)

50. The '980 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 ” and that “[N]odes have an associated state vector $X^T = (x, y, s)$.” EX1001 at 3:47-53. This can be seen in Figures 2A and 2B where, as the patent explains, “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 at 4:45-48. Although slightly blurry, the object nodes in Figures 2A and 2B (from left to right) are denoted with $O_{t-(w-1)/2}$ (for the object at an earlier time), O_t (for the object at a current time), and $O_{t+(w-1)/2}$ (for the object at a later time). EX1001, Fig. 2A; *see also* EX1003, Fig. 2(d) (slightly less blurry). This further indicates to a POSITA that the patent’s spatio-temporal object model explicitly

models the object (as a distinct entity from its components), as well as the object's corresponding components, at multiple points in time.



EX1001, Fig. 4 (excerpted)



EX1003, Fig. 2(d) (excerpted)

A POSITA would have understood that in the example above, “w” represents a window of size w, where w is an odd integer greater than or equal to 1. EX1001 7:6-22. A POSITA would have understood that, when w is an odd integer (as in the example above), there can be an object node in the middle of the temporal window (*e.g.*, at time *t*) with an equal number of object nodes on either side between *t* and *t*-(w-1)/2 (for nodes earlier in time) and between *t* and *t*+(w-1)/2 (for nodes later in time). A POSITA would have understood that additional object nodes (*e.g.*, in cases where w is greater than 3) are not shown in Figures 2A and 2B but are represented by the ellipses in the edges connecting the object nodes.

51. Consistent with the '980 patent's teachings of a probabilistic spatio-temporal model, the patent provides an exemplary “joint probability distribution for

VL Collective Exhibit 2004

IPR2023-00924

Meta Platforms, Inc. v. VL Collective IP LLC

Page 40 of 113

this spatio-temporal graphical object model” in the equation shown below. EX1001 at 3:53-65. The joint probability distribution includes explicit representation of 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 at 3:53-4:8. Those relationships correspond to the lines connecting the nodes (the “edges”) in Figures 2A and 2B. EX1001 at 4:49-51, 4:15-17, 2:59-60.

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

EX1001 at 3:55-65

52. In the joint probability equation above, X_t^O is the state of the object “O” at time t, and X_t^C is the state of the object’s n-th component “C_n” at time t, where the object has N components.² EX1001 at 3:53-4:8. Accordingly, as a POSITA would

² The patent explains that “n” is in the set of numbers from 1 through an upper-bound number of components denoted N (i.e., “n ∈ (1,N)”) and that “t” is in the set

have understood, the patent's spatio-temporal model explicitly models the object (as a distinct entity from its components) at multiple points in time, as well as the object's associated components at each point in time.

53. Furthermore, the term $\psi_{ij}(X_i^O, X_j^O)$ represents “the temporal compatibility of object state between frames i and j,” *i.e.*, between two object nodes across frames. EX1001 at 3:53-4:8. A POSITA would have understood that this temporal compatibility corresponds to a directed edge connecting two object nodes in the spatio-temporal model over time (*i.e.*, across frames or images) and represents temporal probabilistic constraints between the object nodes. EX1001 at 4:49-51 (“Each directed edge between components i and j has an associated potential function $\psi_{ij}(X_i, X_j)$ that encodes the compatibility between pairs of node states.”), 4:15-17 (“spatial and/or temporal constraints corresponding to the edges in a graph”), 2:59-60 (“Directed edges between nodes represent learned spatial and temporal probabilistic constraints.”).

54. The term $\psi_{ik}(X_i^O, X_k^C)$ represents “the spatial compatibility of the object and its components at frame i,” *i.e.*, between an object node and a component

of numbers from 1 through an upper-bound number of time units denoted T (*i.e.*, “ $t \in (1, T)$ ”). EX1001 at 3:53-4:8

node in a given frame. EX1001 at 3:53-4:8. A POSITA would have understood that this spatial compatibility corresponds to a directed edge connecting the object node and a component node at a single point in time (i.e., within a frame or image) and represents spatial probabilistic constraints between the object node and component node. EX1001 at 4:49-51, 4:15-17, 2:59-60.

55. The term $\psi_{ik}(X_i^C, X_i^C)$ represents “the spatial compatibility between object components at frame i,” i.e., between two component nodes in a given frame. EX1001 at 3:53-4:8. A POSITA would have understood that this temporal compatibility corresponds to a directed edge connecting component nodes at a single point in time (i.e., within a frame or image) and represents spatial probabilistic constraints between component nodes. EX1001 at 4:49-51, 4:15-17, 2:59-60.

56. The three potential functions described above—i.e., (1) $\psi_{ij}(X_i^O, X_j^O)$ corresponding to **object node-to-object node temporal constraints**; (2) $\psi_{ik}(X_i^O, X_i^C)$ corresponding to **object node-to-component node spatial constraints**; and (3) $\psi_{ik}(X_i^C, X_i^C)$ corresponding to component node-to-component node spatial constraints—as well as the context of the other disclosures in the patent’s specification, indicate to a POSITA that the ’980 patent’s “spatio-temporal model” for detecting an **object** is “spatio-temporal” because it explicitly models the **object**

(as distinct from its components) both spatially, *i.e.*, within a frame or image, and temporally, *i.e.*, across frames or images.

57. The patent teaches determining a probability that the **object** (as a distinct entity from its components) is in each of a plurality of images. Turning back to the joint probability equation above for the patent’s spatio-temporal model, 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 at 3:53-4:8. As the patent explains, $\phi_i(X_i)$ is a likelihood that “models the probability of observing the image conditioned on the state X_i of the node i .” EX1001 at 5:25-30. In other words, $\phi_i(X_i)$ models the probability of observing something in an image based on its state, where $\phi_i(X_i^O)$ models the probability of observing the object in an image (as distinct from its components, modeled separately as $\phi_i(X_i^C_k)$) based on the object’s state. According to the likelihood model of the patent, “[a]n object is considered present if” the probability (or “confidence”) that an object is present is greater than or equal to a threshold: $h(I) \geq \frac{1}{2} \sum_{k=1}^K \alpha_k$. EX1001 at 5:25-6:13.

58. As explained above, the ’980 patent teaches a new technique in which object detection is “formulated as an inference in” the above-described spatio-temporal model. EX1001 at 2:55-3:2, 3:47-5:24; *see* analysis above. The patent

provides an example of performing such inference using “non-parametric belief propagation” (or non-parametric BP”). EX1001 at 2:55-3:2, 3:47-5:24. In this example, “[i]nferring the state of **the object** and its components in a framework is defined as estimating belief in a graphical model” in a way “which allows inference over arbitrary graphs,” *i.e.*, across graphs at different points in time, using “message passing” to “constrain” a node state. EX1001 at 6:22-64 (emphasis added).

59. The patent teaches that with message passing, “[m]essages 405 are passed between nodes to propagate information about what states neighboring nodes would be,” where message passing includes “observing an image conditioned on the state of node i 402,” as well as “determining a potential designating the compatibility between states of nodes i and [j].” EX1001 at 6:50-7:5, Fig. 4. This is shown in the formulation for a message $m_{ij}(X_j)$ between nodes i and j:

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

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 potential designating the compatibility between the states of node i and j.” EX1001 at 6:50-64. In other words, the message is based on both the likelihood model $\phi_i(X_i)$ for a node i and the applicable node-to-node compatibility for the nodes i and j. EX1001 at 6:50-7:5. Accordingly, for object nodes in different

frames, messages passed between them are based on $\phi_i(X_i^O)$ (which determines the probability that the object is in the image) and $\psi_{ij}(X_i^O, X_j^O)$ (which represents the temporal compatibility of the object's state between frames i and j). EX1001 at 6:22-64, 5:25-6:13, 3:47-4:8.

60. The patent further teaches that “the message used in belief propagation is approximated with a kernel density, formed by propagating a particle set through a mixture of Gaussians density, and the conditional distribution used in standard particle filtering is replaced by product of incoming messages.” EX1001 at 6:29-37. “To form a proposal particle set,” “a corresponding AdaBoost detector is run over an image at a number of scales to produce a set of detection results that score above the $\frac{1}{2} \sum_{k=1}^K \alpha_k$ threshold.” EX1001 at 7:23-44. The patent teaches that this is manageable for “the entire appearance model,” *i.e.*, for object level nodes. EX1001 at 7:23-44, 4:22-24 (“The course, object, layer 203 corresponds to an entire appearance model of the object and is connected to all constituent components, e.g., 202.”).

61. The patent teaches that “it is possible to perform inference over the **spatio-temporal model** defined for the entire image sequence” or that “[a] w-frame windowed smoothing algorithm” can be used where “w is an odd integer ≥ 1 .” EX1001 7:6-22 (emphasis added). Accordingly, the patent teaches to a POSITA

that the “inference” for detecting an object in any given image (defined as an estimated “belief in a graphical model”) is based on determinations of whether an object is present in multiple images. See EX1001 4:45-48, 7:6-22, Fig. 2A; EX1003 at 9, Fig. 2(d). A POSITA would therefore have understood that the technique taught by the patent for detecting an object to necessarily span multiple images (*i.e.*, multiple points in time), consistent with the patent’s goal of providing “[o]bject detection that considers the consistency of the detection over time.” EX1001 at 2:43-3:2. As the patent explains, its multi-frame analysis is able to achieve “temporal integration” within the multi-frame window w and “enforce temporal consistency from before $t-(w-1)/2$,” which “produce[s] more consistent results over time.” EX1001 at 7:6-22.

62. The patent contrasts its improved technique with “single-frame component-based fusion,” which it says resembles a scenario where $w=1$. EX1001 at 7:6-22. The patent teaches that in comparison to its improved technique, performance of such a “simple fusion” scheme is “poor,” whereas the inventors’ improved technique “consistently combines the evidence for accurate estimates throughout the sequence.” EX1001 at 7:56-8:3.

VII. CLAIM CONSTRUCTION

63. In my opinion, a POSITA would have interpreted the following claim term of the '980 patent – in light of the surrounding claim language, the patent's specification, and the patent's file history – as follows:

Term	Meaning to a POSITA
“detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object”	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

64. Claim 1³ 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

³ Claim 9 of the '980 includes similar limitation, thus I treat claim 1 as representative of both claims for purposes of claim construction. My opinions for claim 1 apply equally to claim 9.

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

EX1001, claim 1 (emphasis added).

65. Limitation 1[b] requires a “plurality of images.” A POSITA would understand “plurality” in this context to have its ordinary and customary meaning of “at least two.” The ’980 patent consistently uses the word “plurality” in the ordinary and customary sense and never in a way that is inconsistent with that meaning to a POSITA. See EX1001 at Abstract, 1:28-36, 1:48-52, 2:7-18, 5:45-50. Thus, a POSITA would understand limitation 1[b] to require providing a video with at least two images.

66. A POSITA would have understood that the limitations following 1[b] refer to the “plurality of images” in 1[b] in such a way that each step builds on the performance of a previous step. For instance, limitation 1[c] requires, “measuring the object as a collection of components in **each image**,” which a PSOTIA would have understood to require performing a measurement of the object for each of the previously recited plurality of images. Limitation 1[d] then requires “determining a probability that the object is in **each image**,” which a POSITA would have also understood to require determining an object probability for each of the previously recited plurality of images. 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.” Accordingly, a POSITA would have understood that in the claimed process, after providing plurality of images, and after measuring the object 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 multi-image analysis for detecting the object in any given image.

67. A POSITA would have also considered that object detection based on a multi-image analysis is fundamental to the purpose of the invention, which is to provide “[o]bject detection that considers the consistency of the detection **over time,”** meaning based on multiple images. EX1001 at 2:43-3:2, 4:1-2 (describing “temporal compatibility” of object state “between frames i and j”). 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” for object detection. EX1001 at 7:6-22. That purpose and function distinguish the patented invention as improvement over prior techniques. Indeed, the patent explains that its technique of detecting an object using a multi-frame analysis was shown to “combine[] the evidence for accurate estimates

throughout the sequence” of image frames, whereas a single-frame “simple fusion scheme” was shown to exhibit “poor” performance. EX1001 at 7:56-8:3.

68. I understand that the Board’s institution decision determined that element 1[e] does not require a multi-image analysis. Paper 11 at 26-27 (“On this record, Patent Owner’s contention that the scope of this term [“probabilities for each image”] as read in light of the specification requires comparing multiple probabilities from multiple images to a threshold is inconsistent with the specifications’ disclosure of this term.”). However, a POSITA would have understood that, based on the intrinsic evidence, the “plurality of images” introduced in limitation 1[b] is necessarily “two or more images” by its plain meaning, and that it is these “two or more images” that the detection in element 1[e] is based on, as I explained above. See my analysis above. 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 a POSITA’s understanding of the invention in light of the specification as a whole. EX1001 at 7:53-55 (emphasis added).

69. While the specification describes a scenario where only one image is analyzed—namely where the number of frames *w* over which the analysis is performed is equal to 1 (*i.e.*, where “*w*=1”)—a POSITA would have understood that

the patent distinguishes that scenario from the invention. The patent indicates that “with $w=1$, the algorithm resembles **single frame** component-based fusion.” EX1001 at 7:20-22 (emphasis added). As the window w is the sequence of images over which the detection inference is made, the “single-frame” “component-based fusion” considers only the connected components in a single image. See EX2006, 97:9-21, 101:1-102:3. The patent then distinguishes the $w=1$ scenario from the invention and demonstrates that the $w=1$ scenario does not achieve the purpose or function of the invention. The specification describes a real-world comparison of its multi-image analysis (*i.e.*, a 3-frame analysis) against a “simple fusion detection” that was run “independently at each frame” (*i.e.*, the $w=1$ scenario that “resembles single frame component-based fusion”), 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. A POSITA would have understood the specification to be contrasting the claimed invention with a known technique to highlight the function of the claimed invention. Accordingly, interpreting the claims to encompass detecting an object based on a probability for a single image (*i.e.*, the $w=1$ “single frame component-based fusion”) would ignore invention’s purpose and function described in the specification.

70. Reading claims 1 and 9 as covering the $w = 1$ scenario would read in configurations that are incapable of achieving the invention's function and purpose of providing "[o]bject detection [that] considers the consistency of the detection over time" to produce "temporal consistency" over multiple image frames and "consistent results over time." EX1001 at 2:49-50, 7:6-22. When the window size w is only equal to 1 (*i.e.*, in the case of a "single frame component-based fusion"), 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 , given that $t \pm (1-1)/2 = t \pm (0)/2 = t \pm 0$). 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. See EX2006, 103:22-104:3. In contrast, when, for example, the window size w is equal to 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 , given that $t \pm (3-1)/2 = t \pm (2)/2 = t \pm 1$).

71. In light of the specification, a POSITA would have also understood that detecting the object in any image "**upon**" comparing the probabilities for each image to a threshold for detecting the object means detecting the object "**based on**" comparing the probabilities to a threshold. A POSITA would have understood the

claim this way because the patent teaches basing the detection of an object on whether the object is determined to be present in other images, which is how it achieves “consistency of the detection over time” and produces “temporal consistency” over multiple image frames and “consistent results over time.” EX1001 at 2:43-54; 7:6-22. For example, the patent teaches that “object detection” is “formulated as an inference” in a model. EX1001 at 2:55-3:2. The model is a “spatio-temporal model” for object detection that covers more than “a single frame” because it is “extended over time” to multiple frames, thus it includes multiple instances of the object at a “current time” as well as at “earlier” or “later” times. EX1001 at 4:18-48, Figs. 2A and 2B; see § VI of my analysis above. The patent teaches that the detection inference is defined as an estimation of “belief” in the model over a w-frame window, performed by comparing the likelihood of object presence in each image to a threshold and sending “messages” based on that likelihood for each image (*e.g.*, when it scores above the threshold) to the object in the image under consideration for detecting the object. EX1001 at 5:25-7:44; see § VI of my analysis above. A POSITA would have understood that by this process, the detection of an object in an image is **based on** the probabilities of the object being in other images (as determined by comparing the probabilities to a threshold). This understanding is reinforced by the patent’s indication that simply repeating a

“single frame” or “simple fusion” analysis “independently at every frame” produces “poor” performance in comparison to the patent’s multi-image analysis, which produces “accurate estimates throughout the sequence.” EX1001 at 7:20-22, 7:56-8:3.

VIII. THE ASSERTED REFERENCES

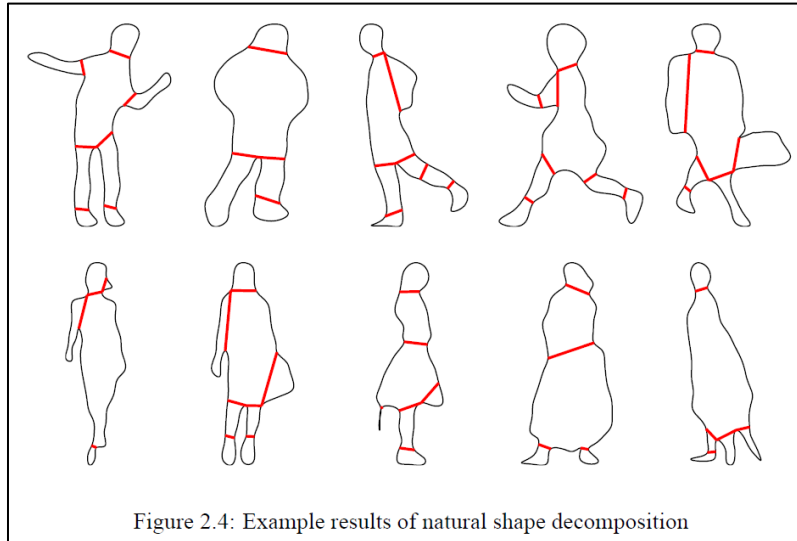
A. Zhao (EX1006)

1. Contour-based detection

72. Zhao acknowledges that “detection” is a “challenging problem[] in computer vision.” EX1006 at 17. Zhao suggests that detecting humans “from a single image” is “inherently difficult for a computer.” EX1006 at 20.

73. Zhao identifies “two approaches to object detection” for detecting a human in a given image: (1) “search the whole image at multi-scales for objects,” and (2) “segment foreground objects from the background, then classify each segmented object as human or non-human.” EX1006 at 20. Zhao expressly “employ[s] the second approach” and states that it “focuses” on how to classify a previously segmented object,” *i.e.*, an object already segmented from the image under consideration, “as human or non-human.” EX1006 at 20, 70 (“the algorithm is independent of how the initial contours are extracted from the image”). Zhao specifically uses “contours,” or “silhouettes,” “of the human body parts as features to detect humans” by “decompos[ing] a silhouette into parts” and “represent[ing] the

shapes and the relationships between the parts.” EX1006 at 22. Zhao states that “Fig. 2.4 shows several example results from the shape decomposition algorithm.” EX1006 at 40.

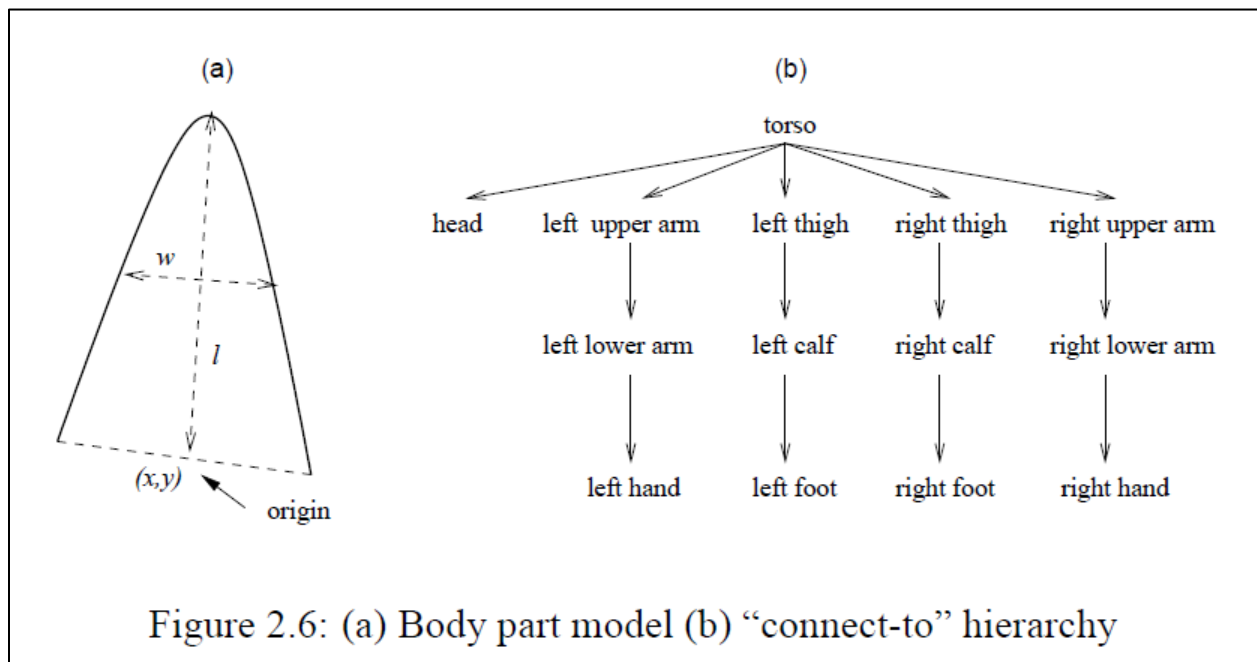


EX1006 at 40

74. Zhao acknowledges that “without prior knowledge of the identification of the object, this goal [of shape decomposition] is hard to achieve.” EX1006 at 23. “In fact,” Zhao explains, “contour extraction and shape decomposition are under-constrained problems without the knowledge about the objects to be detected.” EX1006 at 26, 69. “In contrast, [Zhao] develop[ed] a recursive context reasoning (RCR) algorithm (see Fig. 1.9) to solve the above dilemma.” EX1006 at 26. The RCR algorithm is discussed further below.

2. TRS-invariant human body model

75. Zhao's RCR algorithm relies on "[a] TRS-invariant probabilistic model [that] is designed to encode the shapes of the body parts and the context information – the size and spatial relationships between body parts." EX1006 at 69. Zhao's TRS-invariant human body model "consists of six main body parts (the head, the torso, two arms and two legs) and twelve subparts of the four limbs." EX1006 at 42. "The body parts are modeled with ribbons ... as shown in Fig. 2.6(a)." EX1006 at 42. Zhao defines "ribbon" as "a 'ribbon-like' planar shape [that] is defined by specifying an arc, called the spine or axis, and a geometric figure such as a disk or line segment, called the generator, that 'sweeps out' the shape by moving along the spine, changing size as it moves." EX1006 at 42 footnote 1.



76. In Zhao's TRS-invariant model, "a body part is parameterized with a vector (a, l, x, y, θ) , where $a = w/l$ is the aspect ratio that captures the general shape of a ribbon, (x, y) are the coordinates of the origin in the coordinate frame of the parent part, and θ is the intersection angle between the major axes of this part and its parent part." EX1006 at 43.

77. Zhao further states that "[a]ssuming that the human body model consists of m body parts that are parameterized with vectors $f_1, f_2, \dots, f_m$, where $f_i = (a_i, l_i, x_i, y_i, \theta_i)$. Then, the human body model is parameterized with four model matrices: the aspect ratio vector $A = [a_1, \dots, a_m]$ the length ratio matrix $S = \{s_{ij}\}, i, j = 1, \dots, m$, where $s_{ij} = l_i/l_j$, the relative position vector $X = (x_1, y_1, \dots, x_m, y_m)$, and the orientation or posture vector $\Theta = \{\theta_1, \dots, \theta_m\}$," where "A and S are TRS-invariant." EX1006 at 43.

78. Zhao states that "[i]n this thesis, probability distributions are employed to accommodate shape variation among people and they can also absorb some of the shape variation due to changes of viewpoint." EX1006 at 44. 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 at 45.

3. Bayesian Similarity Measure for Human Detection

79. Zhao uses “a Bayesian Similarity Measure derived from the probabilistic human model.” EX1006 at 54. Zhao states that “[t]he body part identification problem is to match a set of ribbons decomposed from a contour against a set of model body parts described in Chapter 2,” i.e., parts of the TRS-invariant body model. EX1006 at 54. Under Zhao’s framework, “the more body parts being identified, the more likely the extracted contour is a person.” EX1006 at 55, 56. Zhao states that “the similarity measure that evaluates the resemblance between the contour C and the human model is defined as $BSM(C) \triangleq G(H^*)$,” where $G(H) \triangleq P(C|H, person)P(person|H)$ and H^* is the best hypothesis selected as follows: $H^* = \underset{H}{\operatorname{argmax}} P(C|H, person)P(person|H)$. EX1006 at 55.

80. Zhao explicitly formulates human detection as follows: “The human detection problem is: given a detected contour C , determine whether this contour represents the silhouette of a person or not.” EX1006 at 64. Zhao states that “[a] simple decision rule to perform human detection is: the contour C corresponds to a person if $P(person|C) > threshold$.” EX1006 at 64. “Otherwise, C is not a person.” EX1006 at 64. After manipulating $P(person|C)$ according to Bayes rule, conditioning the result on “the hypothesis for body part identification,” and employing certain approximations and assumptions, Zhao reformulates the decision

rule such that “the Contour C corresponds to a person if $P(C|H^*, person)P(person|H^*) > threshold$.” EX1006 at 65. According to Zhao, “[t]his means that the similarity measure proposed in this thesis can be used to perform human detection: the contour C corresponds to a person if $BSM(C) > threshold$.” EX1006 at 65.

4. Recursive context reasoning (“RCR”) algorithm

81. Zhao 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 at 69. Zhao states that “[t]his thesis proposes a recursive context reasoning (RCR) algorithm to solve the above dilemma.” EX1006 at 69.

82. 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 at 69. The RCR algorithm also uses a “Bayesian framework ... to perform human detection and part identification under partial occlusion.” EX1006 at 69. The RCR algorithm also 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

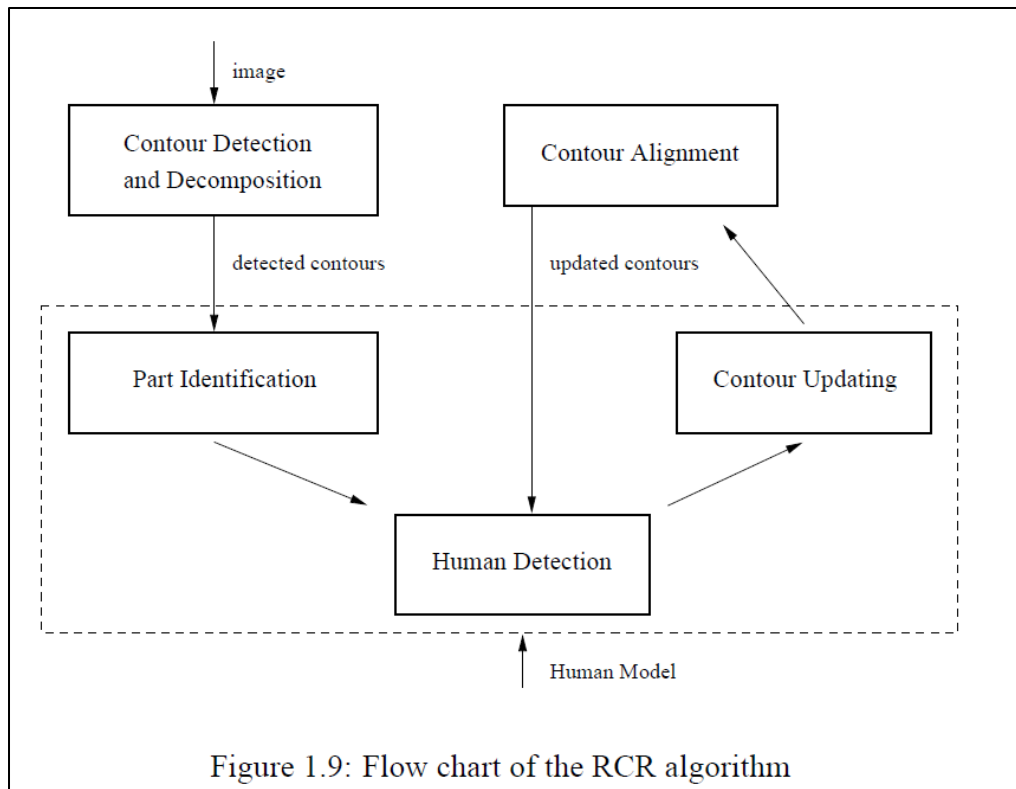
the Bayesian similarity measure and determine if the detected contour is a person or not.”

83. The RCR algorithm has seven steps:

- “**Step 1:** Contour extraction: the algorithm is independent of how the initial contours are extracted from the image. Methods such as depth segmentation or background subtraction can be used for this purpose.” EX1006 at 70-71. For example, “the person is segmented from the background.” EX1006 at 91.
- “**Step 2:** Contour decomposition (described in Section 2.2): decompose the contour C into natural parts.” EX1006 at 70-71. For example, “the segmented region is decomposed into ribbons and these ribbons are matched with the modeled body parts.” EX1006 at 91.
- “**Step 3:** Body part identification (described in Section 3.3): generate a hypothesis $H^* = \underset{H}{\operatorname{argmax}} G(H)$, where $G(H)$ is the goodness function calculated using Eq. (3.9). EX1006 at 70-71.
- “**Step 4:** Human detection (described in Section 3.4): calculate the similarity measure $BSM(C) = G(H^*)$ to determine if the contour C is a person or not based on the decision rule given in (3.19).” EX1006 at 70-71.

- “**Step 5:** Update the locations and outlines of the body parts from the identified body parts and the human model.” EX1006 at 70-71. For example, “the locations of the joints are adjusted to achieve consistency with the modeled spatial and size relationships between the body parts,” and “[t]he locations and the sizes of the missed body parts are inferred from the extended body parts and the detected body parts.” EX1006 at 91.
- “**Step 6:** Align the predicted outlines of the missed body parts to the edge features in the image.” EX1006 at 70-71.
- “**Step 7:** Recalculate the similarity measure and determine if a person is present or not.” EX1006 at 70-71.

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.



EX1006 at 26

84. Zhao’s goal of steps 5 and 6 (*i.e.*, “contour updating” and “contour alignment”) is “to refine the shapes and locations of the identified body parts and to predict the shapes and locations of the missed body parts.” EX1006 at 72. Zhao explains that “the RCR algorithm helps to locate missed parts” and that “[b]ecause more body parts are detected from the image, the Bayesian similarity measure can be estimated more accurately.” EX1006 at 78-80.

85. Zhao states that “[m]otion information obtained from previous frames can be used to predict the orientations of the limbs and to solve the ambiguity” when limbs are misaligned or confused. EX1006 at 91. Zhao teaches the use of

“geometrical relationships” and “spatial and size relationships” for updating and predicting the parameters of detected and undetected parts. EX1006 at 71, 91; see also EX1006 at 72 (using vector (a, l, x, y, θ)), 74 (using vector $(\hat{a}_j, \hat{l}_j, \hat{x}_j, \hat{y}_j)$).

B. Steffens (EX1007)

1. Steffens’ single-frame graph matching detection technique

86. Steffens relates to “vision-based object detection and tracking, and more particularly, to systems for detecting objects in video images, such as human faces, and tracking and identifying the objects in real time.” EX1007 at 1:14-17, 3:36-37 (“The object may be, for example, a head having particular facial characteristics.”). As shown in Fig. 1 below, Steffens’ process includes steps of “receiv[ing] a person’s image from a video source” such as “a stream of digital video image frames” and performing “head detection and tracking process 14”; using a “landmark finding process 18 for detecting the individual facial features”; and using a “landmark tracking process 20” to “track [] the landmarks,” among other things. EX1007 at 3:47-4:61.

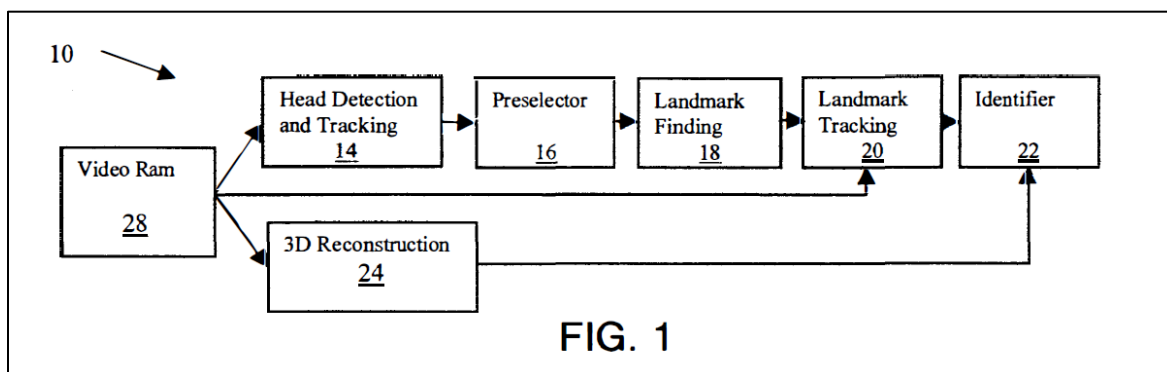
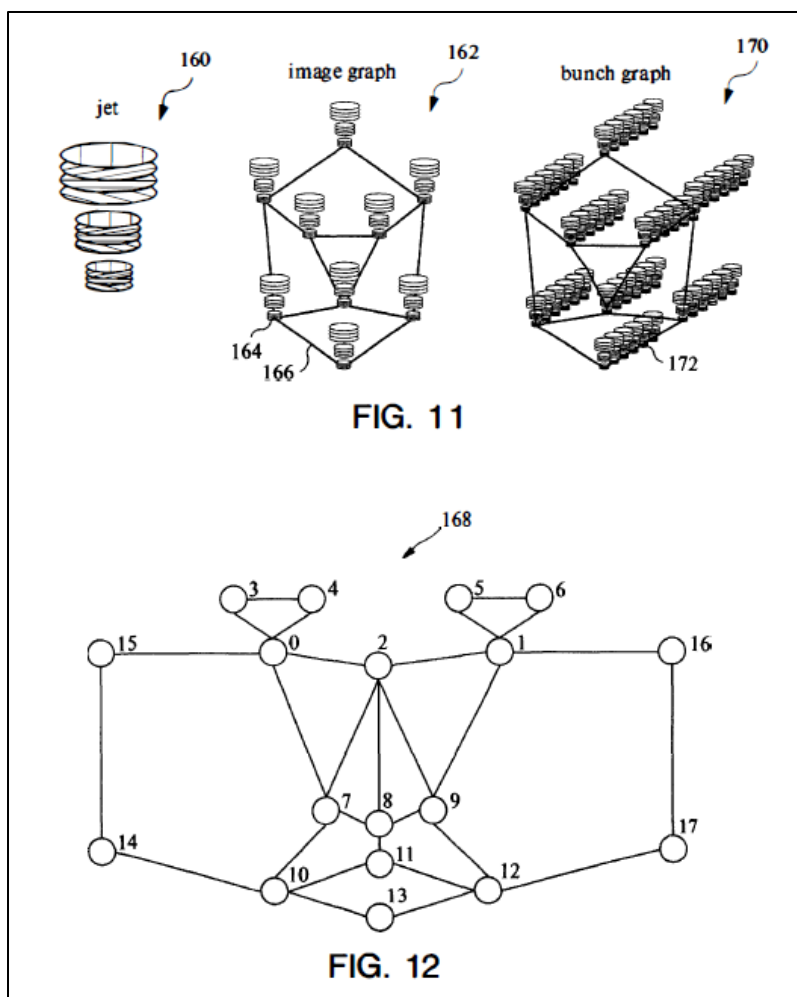


FIG. 1

EX1007, Fig. 1

87. The “landmark locations on an image correspond[] to readily identifiable features,” *i.e.*, facial features of a head. EX1007 at 3:34-46. “The facial landmarks and features of the head may be located using an elastic graph matching technique.” EX1007 at 7:51-53. The graph matching technique involves “determining the similarity between [an] image graph and [a] model graph.” EX1007 at 7:51-67. An “image graph” is “used to describe the aspects of an object (in this context, a face),” in which “nodes 164” in the image graph “refer to points on the object and are labeled by jets 160” and “define the graph topology.” EX1007 at 8:56-67. “A model graph 168,” on the other hand, includes “nodes” corresponding to facial features (*e.g.*, right eye pupil, left eye pupil, top of the nose, etc.). EX1007 at 9:1-21. 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 at 9:22-45, Figs. 11-12.



EX1007, Figs. 11-12

88. “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 at 9:22-36.

89. “The jets are composed of wavelet transforms and are processed at nodes or landmark locations on an image corresponding to readily identifiable features.” EX1007 at 3:42-44. “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 (the best fitting jets within the bunch jets are most similar to jets extracted from the image at the current positions of the nodes). EX1007 at 9:37-41. 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)

1. Ozer's single-frame object detection technique with activity classification

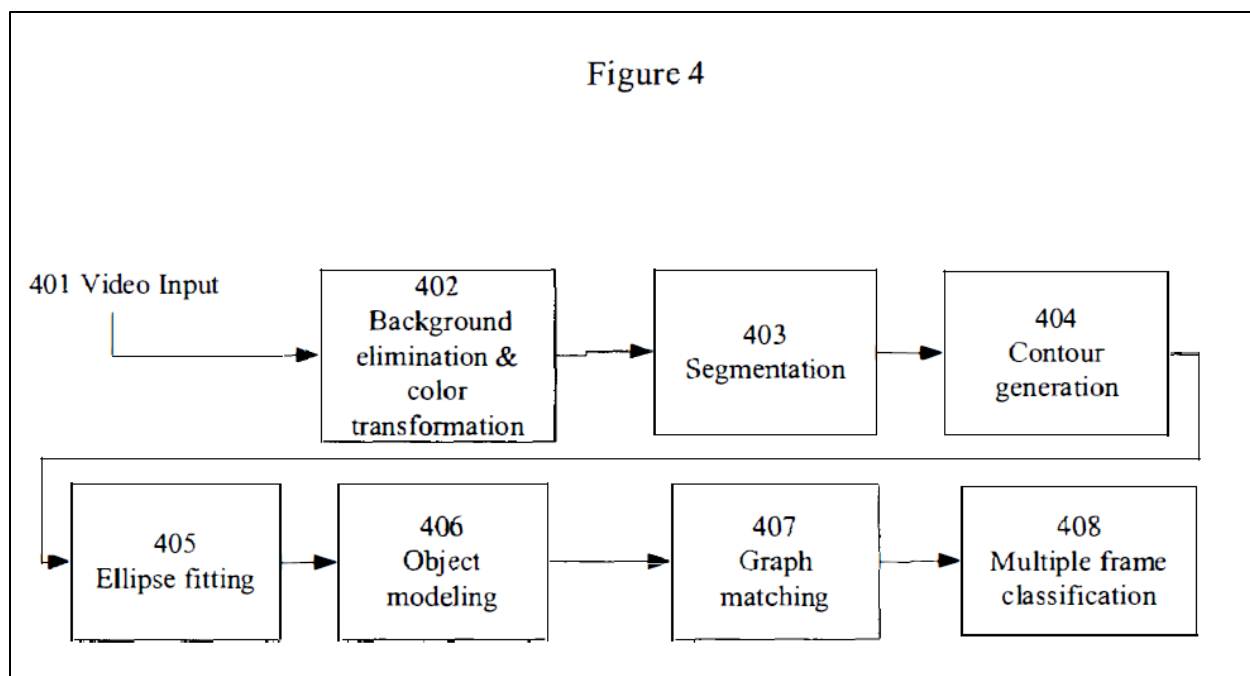
90. Ozer states that it relates to a “method and apparatus to detect the presence of articulated objects, e.g., human body, and rigid objects and to identify their activities.” EX1008 at 1:15-21. Ozer says that it seeks to provide a “system” that can “detect non-rigid (e.g., human body) and rigid object parts and recognize their activities.” EX1008 at 3:55-57. According to Ozer, its technique uses “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 at 3:55-64.

91. According to Ozer, “[o]bject detection is achieved by matching the relational graphs of objects with the reference model.” EX1008 at 4:24-34. Ozer states that “the system can identify each object part separately (in this case hands 100, head 101, and torso 102) after comparing the object attributes with the model database via a graph-matching algorithm.” EX1008 at 5:66-6:3. Ozer states that “this system builds a model of the object of interest on each video frame” and “tracks only objects that fit the model of the user defined subject.” EX1008 at 6:21-24. Ozer’s detection process includes:

- “Background elimination and color transformation,” where “[t]he foreground-background separation is achieved by comparing the DC 50 coefficients of the foreground object with the DC coefficients of the background object via a statistical method.” EX1008 at 6:44-51.
- “Segmentation,” where “the object of interest is segmented hierarchically into its smaller unique parts.” EX1008 at 6:52-58.
- “Contour following,” where “the frame given to the contour following algorithm 501 results in the output frame with the contour 502.” EX1008 at 6:59-63.
- “Ellipse fitting,” that “fits ellipses to the contours.” EX1008 at 6:64-7:3.

- “Object modeling by invariant shape attributes” based on “part attributes” used for “object detection.” EX1008 at 7:4-7.
- “Graph matching,” where “we compare the object model with a set of stored models,” and “[e]ach object part and meaningful combinations represent a class w,” and “[i]f the graph-matching algorithm cannot find a meaningful correspondence of the combined segments in the reference model, the combination will be rejected and a new combination will be generated.” EX1008 at 6:64-7:3. “[A] piecewise quadratic Bayesian classifier with discriminant function $g(X)$ is used” for determining the class of each object part or combination. EX1008 at 7:4-31. The “[o]utput of the graph-matching algorithm is the classified object parts.” EX1008 at 7:34-35.

92. “After the detection of the object parts, the system is ready to recognize the activities of each object part and the overall activity of the object.” EX1008 at 4:35-37.



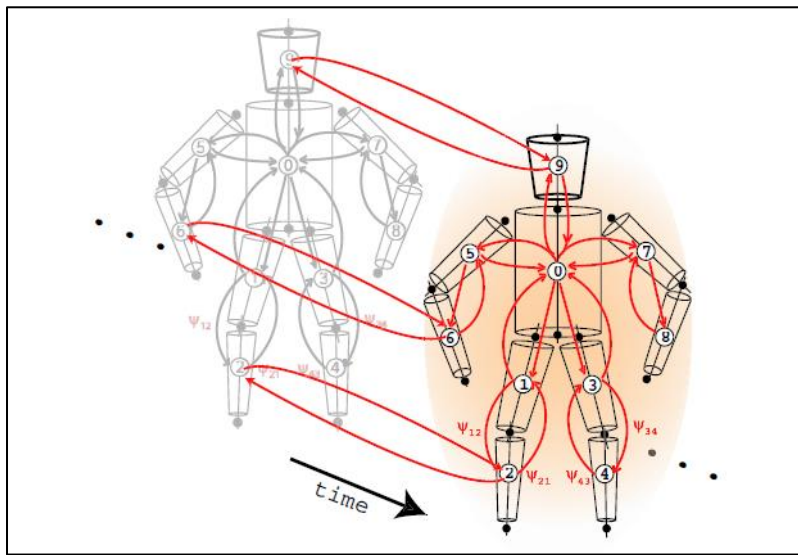
EX1008, Figure 4

93. Ozer states that the system can “be set up to ... recognize a wide variety of activities, such as gestures ranging from pointing a gun to waving arms, leaving unattended objects, entering a prohibited area, tailgating at security doors, spending too much or too little time in an area, etc.” EX1008 at 6:4-14. “The system checks direction of the movements of the object parts for a number of frames and calculates the probabilities of the activities with the known activities by using Hidden Markov Models and chooses the pattern with the highest probability as the recognized activity in these frames (408).” EX1008 at 7:34-45.

D. TLP (EX1009)

1. TLP's body part detection technique

94. According to TLP, it describes a “method for tracking human bodies” using “a loose-limbed body model . . . in which limbs are connected via learned probabilistic constraints.” EX1009 at 1. “Each node in the graphical model represents . . . a limb (Figure 1).” EX1009 at 1.



EX1009 at 1

95. 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 at 1. However, according to TLP, “[w]hile the full body pose is hard to recover directly, the location and pose of individual **limbs** is much easier to compute.” EX1009 at 1 (emphasis added). To that end, TLP teaches that “the body is represented by a

VL Collective Exhibit 2004

IPR2023-00924

Meta Platforms, Inc. v. VL Collective IP LLC

Page 71 of 113

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 at 2 (emphasis added).

96. TLP describes “a loose-limbed body model in which **limbs** are connected via learned probabilistic constraints.” EX1009 at 1 (emphasis added). According to TLP, “[t]he **tracking** problem is formulated as one of inference in a graphical model and belief propagation is used to estimate the pose of the body at each time-step.” EX1009 at 1. Further, “[e]ach node in the graphical model represents the 3D position and orientation of a **limb** (Figure 1),” where “[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 at 1.

97. TLP states that “each node,” *i.e.*, “**each limb**,” “has an associated likelihood defined over a rich set of image cues using a learned Gibbs model.” EX1009 at 1-2 (emphasis added). 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 at 4 (emphasis added).

98. TLP states that “[i]nferring the body **pose** in our framework is defined as estimating belief in the graphical model” using “a form of non-parametric belief propagation,” where “the ‘message’ used in standard belief propagation is approximated with a particle set, and the conditional distribution used in standard particle filtering is replaced by a product of incoming message sets.” EX1009 at 4. TLP states that “we use a **limb** proposal distribution based on local image measurements” using “simple eigen-template detectors for the head, upper arms, and lower legs.” EX1009 at 5. 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.

IX. DR. BAJAJ DOES NOT SHOW THAT ANY CHALLENGED CLAIM WOULD HAVE BEEN OBVIOUS

A. Dr. Bajaj does not demonstrate that Zhao (Ground I) discloses or renders obvious any challenged claim.

1. Claim 1[e]/9[e] “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.”

99. In my opinion, Dr. Bajaj has not shown that a POSITA would have understood Zhao to disclose or render obvious this claim element.

100. Dr. Bajaj fails to show that Zhao discloses “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 stated 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 at 2:49-50, 7:6-22; See analysis in § VII above. A 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 analysis in § VII above. Yet, as explained below, Zhao’s human detection method 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.

a. Dr. Bajaj fails to address every limitation of element 1[e]/9[e] and therefore fails to show it was disclosed or rendered obvious by Zhao.

101. Dr. Bajaj fails to show that Zhao discloses or renders obvious every limitation in this element of “detecting the object **in any image** upon comparing the probabilities **for each image** to a threshold for detecting the object.”

102. Dr. Bajaj’s analysis of element 1[e] is based on his analysis of element 1[d]. See EX1004 ¶98. Element 1[d] requires, “determining a probability that the object is in each image.” EX1001, claim 1. Dr. Bajaj opines 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.” EX1004 ¶92. Dr. Bajaj point to Zhao’s “‘simple decision rule to perform human detection’ where the probability the ‘contour C corresponds to a person’ in each image is calculated [as] $P(\text{person}|C) > \text{threshold}$ ” and its reformulation as $P(C|H^*, \text{person})P(\text{person}|H^*)$. EX1004 ¶93. Dr. Bajaj equates “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.” EX1004 ¶93.

103. Dr. Bajaj relies on the same determination of the probability that a contour C corresponds to a person in Zhao for his element 1[e] opinions. Specifically, in attempting to show that “Zhao compares the probability to a threshold to perform object detection,” Dr. Bajaj points to both “ $P(\text{person}|C) > \text{threshold}$ ” and “ $P(C|H^*, \text{person})P(\text{person}|H^*) > \text{threshold}$.” EX1004 ¶98.

104. But Dr. Bajaj never explains, in either his element 1[d] or element 1[e] analysis, 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.” That is, Dr. Bajaj does not explain how Zhao discloses or renders obvious this element as a whole.

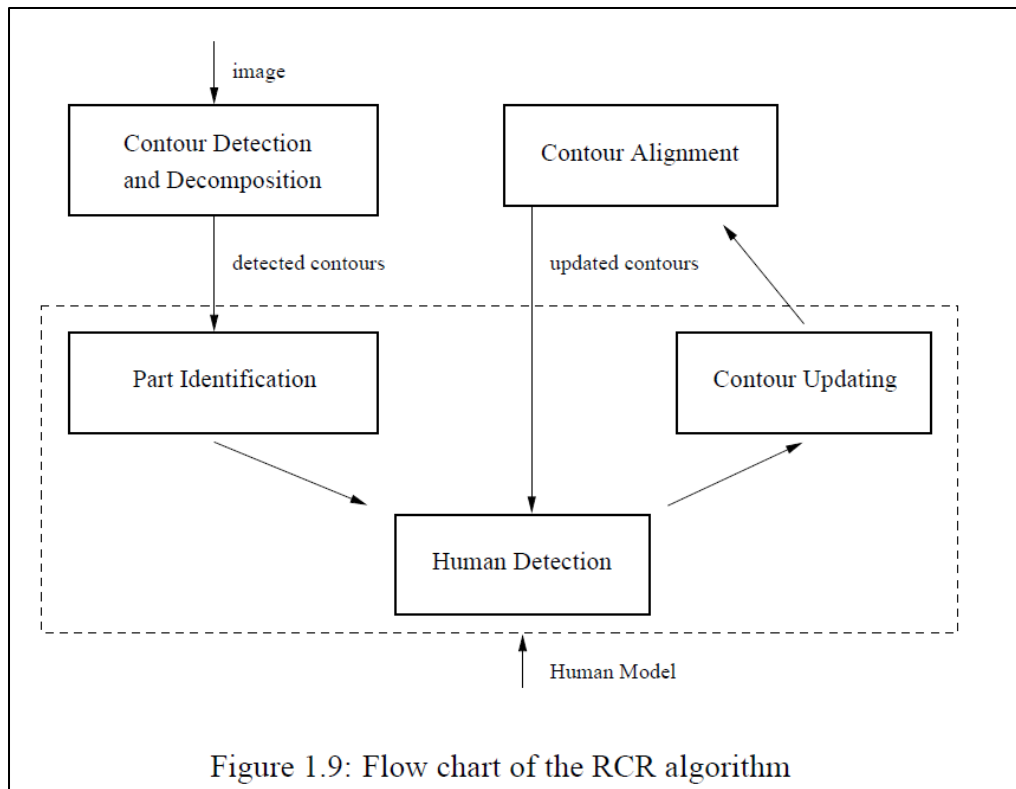
105. Instead, Dr. Bajaj purports to show that Zhao “determine[s] the probability that the object is in each image” and that it also “compares the probability to a threshold to perform object detection.” EX1004 ¶98. But the claim requires more. Dr. Bajaj never explains how Zhao’s RCR algorithm (including 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.” Nor does Dr. Bajaj show that it would have been obvious to modify Zhao to arrive at the claimed invention as a whole.

106. Dr. Bajaj’s further analysis for element 1[e] does not resolve this deficiency. Dr. Bajaj opines that “employing a threshold in a probability measure was known to a 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 a POSITA would have found it obvious to modify Zhao to arrive at the specific limitations of the claimed invention as a whole. That the use of thresholds was known does not indicate to a POSITA that 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.” Dr. Bajaj’s opinion that it does is based on an improper interpretation that ignores these limitations and is skewed toward finding the claims obvious, as evidenced by his opinion that claims 1/9 cover the “w=1” scenario that the patent clearly distinguishes from the claimed invention. See analysis in § VII above; see EX2006 at 18:8-14, 97:22-99:15.

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

107. A POSITA would understand that Zhao’s RCR algorithm performs a human detection analysis for a single-image that is not based on the probabilities of the human being in other images (or comparing those probabilities to a detection threshold). Zhao expressly acknowledges that “without prior knowledge of the identification of the object, this goal [of shape decomposition] is hard to achieve.” EX1006 at 23. But Zhao does not disclose a way of incorporating prior (or later) knowledge of the identification of the object into a detection framework—instead, 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 at 26, 69. Instead, Zhao developed “a recursive context reasoning (RCR) algorithm (see Fig. 1.9) to solve the above dilemma.” EX1006 at 26.



EX1006 at 26

108. As shown in Figure 1.9, Zhao’s RCR algorithm operates on a single image, shown as an input in the upper left corner. See EX2006 at 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 analysis in § VII.A above.

109. 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.

VL Collective Exhibit 2004

IPR2023-00924

Meta Platforms, Inc. v. VL Collective IP LLC

Page 78 of 113

In contrast, 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 at 26 (emphasis added). 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 at 26, 64-65 (“Because the recursive context reasoning algorithm explained in Chapter 4 will improve the accuracy of contour extraction iteratively, the above approximation [in the BSM] will get more accurate accordingly.”). A POSITA would have understood that this process of updating contours and re-running the BSM evaluation on the same image does not disclose or render obvious detecting an object in a given image (in Zhao’s case, a person) based on the probability of the object (the person) being detected in another image or comparing that probability to a detection threshold.

110. Nor does the RCR algorithm involve determining or considering the probability of the object (a person) being present in another image or comparing that probability to a detection threshold. 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. Knowledge of whether the object (person) is in other images is not involved. *See* EX1006 at 23,69.

111. The contour updating and alignment operations of the RCR algorithm (Steps 5 and 6) operate on the same image as steps 1-4 (not different images) and likewise do not involve the probability of the object (the person) being present in another image or comparing that probability to a detection threshold. 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 at 72. But Zhao teaches performing those steps 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 at 71, 91; *see also* EX1006 at 72 (using vector (a, l, x, y, θ)), 74 (using vector $(\hat{a}_j, \hat{l}_j, \hat{x}_j, \hat{y}_j)$).

112. While Zhao states that “[m]otion information obtained from previous frames can be used to predict the orientations of the limbs and to solve the ambiguity” when limbs are misaligned or confused, a 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 at 91. Indeed, 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 at 71, 91, 72, 74. Zhao’s example of considering “how the left knee angle changes with time” confirms to a POSITA that Zhao’s consideration of “motion information” from “previous frames” is based on geometrical and spatial relationships, not the probability that the object (person) is present in another image or comparing that probability to a detection threshold.

2. Claim 4/12

113. Claim 4 reads: “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.” EX1001, claim 4.

114. Claim 12 reads: “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.”
EX1001, claim 12.

115. I disagree with Dr. Bajaj’s opinion that “a POSITA would have understood Zhao to teach a method of detecting an object in a current image according to measurements of the object as a collection of components from other image frames, including prior **and later image frames.**” EX1004 ¶107 (emphasis added). 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.” Thus, the claims specifically require detecting the object in a current (a 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. Zhao does not disclose such a technique or render it obvious. 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.

116. 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 (emphasis added). I disagree. 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. Dr. Bajaj cites no portion of Zhao to support his assertion. Dr. Bajaj’s interpretation of Zhao is also incorrect.

117. 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. Dr. Bajaj then states that “If 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.” EX1004 ¶108 (emphasis added).

118. Zhao’s description of its MAP estimation, however, does not disclose or suggest the use of information from a later image. EX1006, 74-75. 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. 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. This is not disclosed or suggested by Zhao’s MAP estimation.

119. Zhao’s use of “[m]otion information obtained from **previous** frames ... to predict the orientations of the limbs and to solve the ambiguity” also does not disclose or suggest the use of measurements from “a **later** image relative to the current image.” EX1004 ¶107. Dr. Bajaj identifies no disclosure in Zhao of the use of measurements from a later image. EX1004 ¶¶106-110. And 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 ¶109. Even if the Board were to find it was “widespread and commonly understood” (which Dr. Bajaj has not shown), Dr. Bajaj provides no reason or rationale as to why a POSITA would have found it obvious to modify Zhao to arrive at the specific limitations of the claimed invention as a whole.

3. Claim 8/16

120. Claim 8 reads: “The computer implemented method of claim 1, wherein a joint probability distribution for the spatio-temporal model with N components is:

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

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

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

”

EX1001, claim 8.

121. Claim 16 reads: “The method of claim 9, wherein a joint probability distribution for the spatio-temporal model with N components is:

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

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

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

”

EX1001, claim 16.

122. Claim 8/16’s joint probability distribution includes elements for a spatio-temporal model that do not exist in Zhao’s framework. Dr. Bajaj provides no explanation as to how it would have been obvious to modify Zhao to arrive at the particular joint probability distribution of claims 8/16.

123. The joint probability distribution of claims 8 and 16 is defined in the ’980 patent’s specification at 3:47-4:8. The patent indicates what the terms in the

distribution mean. X_t^O is the state of the object “O” at time t, and X_t^C is the state of the object’s n-th component “ C_n ” at time t, where the object has N components. EX1001 at 3:53-4:8. Accordingly, as a POSITA would have understood, the patent’s spatio-temporal model explicitly models the object (as a distinct entity from its components) and its associated components at multiple points in time. See EX2006 at 70:9-17, 80:15-21.

124. The joint distribution also includes multiple terms that represent spatial compatibility or temporal compatibility between specific nodes types. See analysis in § VI above. 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. *Id.*; EX1001 at 4:49-51, 4:15-17, 2:59-60.

125. I disagree with Dr. Bajaj that Zhao discloses or renders the ’980 patent’s joint distribution of claims 8 and 16 obvious because, as explained above, that distribution models object and component nodes explicitly at multiple points in time, and explicitly defines temporal constraints between object-level nodes, as well as spatial constraints object node-to-component node and component node-to-component node. Dr. Bajaj fails 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 “[i]n this thesis, probability distributions are employed to accommodate shape variation among people and they can also absorb some of the shape variation due to changes of viewpoint.” EX1006 at 44. 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 at 45.

126. A 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 the TRS-invariant model. See analysis of claim 4 above. Accordingly, a POSITA would have understood that Zhao’s probability distribution is not, as Dr. Bajaj suggests, “representable in the form of the joint probability distribution formulated by Claims 8 and 16,” and that it would not have been obvious to formulate it that way. EX1004 ¶121.

127. Dr. Bajaj also has not shown that the joint probability formulation of claims 8/16 would have been well known or that EX1012 discloses that formulation. 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 at 1 (emphasis added), 3, 6. Dr. Bajaj identifies no disclosure in EX1012 of object-level representation in combination with component-level representation. Not does he provide a reason or rationale as to why a POSITA would have found it obvious to modify Zhao to arrive at the specific probability distribution of claims 8/16 as a whole.

B. Dr. Bajaj does not demonstrate that Steffens in view of Zhao (Ground II) discloses or renders obvious any challenged claim.

1. Claim 1[d]/9[d] “determining a probability that the object is in each image.”

128. Dr. Bajaj acknowledges that Steffens’ alleged “‘similarity’ measure” is not formulated as calculating a “probability” that an object is in an image.” EX1004 ¶144. Dr. Bajaj opines that because “probabilistic frameworks” were known, it would have been obvious to combine the teachings of Steffens with Zhao’s Bayesian Similarity Measure. EX1004 ¶¶144-145. I disagree that a POSITA would have found it obvious to combine Steffens and Zhao to achieve this claim element.

129. A POSITA would have understood that Steffens’ wavelet-based approach is fundamentally different from, and thus incompatible, with Zhao’s contour-based approach. 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 at 20, 81-82, 86. Zhao’s Bayesian similarity model for human detection, however, is fundamentally based on contour extraction, *see* EX2006 at 111:21-112:4, which a POSITA would have viewed as incompatible with Steffens’ wavelet-based approach. EX1006 at 64 (“The human detection problem is: given a detected contour *C*, determine whether this contour represents the silhouette of a person or not.”). Thus, a POSITA would not have looked to Zhao’s contour-based Bayesian similarity measure approach to modify Steffens’ wavelet-based bunch graph approach.

130. Furthermore, Zhao teaches away from combining Steffens’ wavelet-based bunch graph approach with Zhao. 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 at 82-83. 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 at 88. 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 (the best fitting jets within the bunch jets are most similar to jets extracted from the image at the current positions of the nodes). EX1007 at 3:42-44, 9:37-41. Thus, a POSITA would not have been motivated to combine Zhao and Steffens and would have been led away from the combination.

2. **Claim 1[e]/9[e] “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.”**

a. **Dr. Bajaj fails to address every limitation of element 1[e]/9[e] and therefore fails to show it was disclosed or rendered obvious by Steffens and Zhao.**

131. Relying on his opinions for claim 1[d]/9[d], Dr. Bajaj opines that “it would have been obvious to a POSITA to use a probabilistic framework,” and further asserts that “as part of a probabilistic framework, it would have been obvious to a POSITA to detect the image by comparing the probabilities for each image to a threshold for detecting the object.” EX1004 ¶146. As with his analysis of Zhao, Dr. Bajaj fails to consider this element as a whole with all of its limitations. Dr. Bajaj never explains 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 Steffens and Zaho. Nor does Dr. Bajaj show that it would have been obvious to combine Steffens and Zhao to arrive at the claimed invention as a whole.

132. Rather than addressing the actual claim language, Dr. Bajaj claims that “employing a threshold in a probability measure was known to a POSITA” because “there are many examples of using thresholds in probabilistic frameworks.” EX1004 ¶147. But Dr. Bajaj does 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. Whether Dr. Bajaj’s vague and generic concept of “employing a threshold in a probability measure” was known in the art does not show that a POSITA would have found the specific requirements of claim 1[e]/9[e] obvious over the combination of Steffens and Zhao. That the use of “a threshold in a probability measure” was known does not indicate to a 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. As explained below, it does not.

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

133. Dr. Bajaj further asserts that a 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. I disagree. For the same reasons explained above with respect to claim 1[d]/9[d], a

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 analysis in § IX.B.1 above.

134. Furthermore, for the reasons I explained above for Ground I, Zhao does not disclose or render obvious claim 1[e]/9[e]. See analysis in § IX.A.1 above. Dr. Bajaj does also not show that Steffens discloses this claim element. Accordingly, a POSITA would not have found claim 1[e]/9[e] disclosed or rendered obvious by the combination of Steffens and Zhao.

3. Claim 4/12

135. Claim 4 reads: "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." EX1001, claim 4.

136. Claim 12 reads: "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." EX1001, claim 12.

137. Dr. Bajaj argues that in view of Steffens' reference to tracking an object "between frames," "a POSITA would have found it disclosed or at least obvious to detect the object according to measurements of the object ... from a 'prior image' and a 'later image' relative to the current image." EX1004 ¶¶156-157. According to Dr. Bajaj, this is because a POSITA would have allegedly found it obvious that "'between frames' would include doing so according to a 'prior image' and a 'later image.'" EX1004 ¶157. I disagree.

138. Dr. Bajaj has not shown that Steffens discloses or renders obvious "**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." Steffens detects objects using a graph matching technique that involves "determining the similarity between [an] image graph and [a] model graph." EX1007 at 7:51-67. Dr. Bajaj agrees. EX1004 ¶125 ("Steffens describes 'identifying the object based on a similarity' between an image and a corresponding model."). But a POSTIA 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. 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 at 7:51-67. A POSITA would therefore have understood that Steffens’ graph matching detection detects an object in a given image by comparing a graph for that image with a given model graph—it does not detect that based on graph data from other images, whether earlier or later.

139. Additionally, I disagree with Dr. Bajaj that “between frames” with respect to tracking in Steffens discloses detecting an object in a current image according to measurements of the object from later images or renders it obvious. Steffens confirms to a POSITA that Steffens tracking technique operates in only the forward direction. 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 at 10:46-61. Steffens does not disclose or suggests tracking

backwards in time or detecting an object in a current frame according to graph information for the object from a later frame.

140. 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.

4. Claim 8/16

141. Claim 8 reads: “The computer implemented method of claim 1, wherein a joint probability distribution for the spatio-temporal model with N components is:

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

EX1001, claim 8.

142. Claim 16 reads: “The method of claim 9, wherein a joint probability distribution for the spatio-temporal model with N components is:

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

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

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

”

EX1001, claim 16.

143. Dr. Bajaj relies on his opinions for combining Steffens and Zhao with respect to elements 1/9 (d-e). For the reasons explained above with respect to claims 1[d]/9[d] and 1[e]/9[e], a POSITA would not have found it obvious to combine Steffens and Zhao to achieve those elements. Further, for the reasons explained above under Ground I, Zhao does not disclose or render obvious the joint probability distribution of claims 8 and 16. See analysis in § IX.A.3 above.

- C. Dr. Bajaj does not demonstrate that Ozer in view of Zhao (Ground III) discloses or renders obvious any challenged claim.**
- 1. Claim 1[e]/9[e] “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.”**

144. For this claim element, Dr. Bajaj relies on his Ground I opinions that this element was allegedly “disclosed or at least rendered obvious in view of the teaches of Zhao.” EX1004 ¶187. For the reasons explained above with respect to claim 1[e]/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 analysis in § IX.A.1 above. Also, Dr. Bajaj does not argue that Ozer discloses or renders obvious the limitations of this element. Accordingly, Dr. Bajaj has not shown that this element was disclosed or rendered obvious to a POSITA by the combination of Ozer and Zhao.

2. Claim 4/12

145. Claim 4 reads: “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.” EX1001, claim 4.

146. Claim 12 reads: “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.” EX1001, claim 12.

147. Dr. Bajaj argues that in view of Ozer’s method of “Classifying Over Multiple Frames,” a POSITA would have allegedly “found it disclosed or at least obvious to detect an object 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.” EX1004 ¶193. I disagree.

148. Dr. Bajaj has not shown that Ozer discloses or renders obvious “*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.” 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 at 3:55-64.

149. A POSITA would have understood that the object detection is based on information in a single image, not prior or later images. 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 at 5:10-20 (emphasis added). 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 at 6:52-58. 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 at 3:65-67 (emphasis added). Furthermore, Ozer states that “this system builds a model of the object of interest **on each video frame.**” EX1008 at 6:19-24. These teachings of Ozer confirm to a POSITA that Ozer’s detection processes 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. 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.

150. Dr. Bajaj’s 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 at 3:55-64, 4:44-53 (“The high-level part ... classifies activity patterns” and detects “specific gestures and activities.”). Ozer states that 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 at 4:35-37. From this a POSITA would understand that Ozer’s object

classification step is for detecting an object, *see* EX2006 at 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. 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 at 7:34-45.

151. 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 ¶ 194.

3. Claim 8/16

152. Claim 8 reads: “The computer implemented method of claim 1, wherein a joint probability distribution for the spatio-temporal model with N components is:

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

EX1001, claim 8.

153. Claim 16 reads: “The method of claim 9, wherein a joint probability distribution for the spatio-temporal model with N components is:

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

”

EX1001, claim 16.

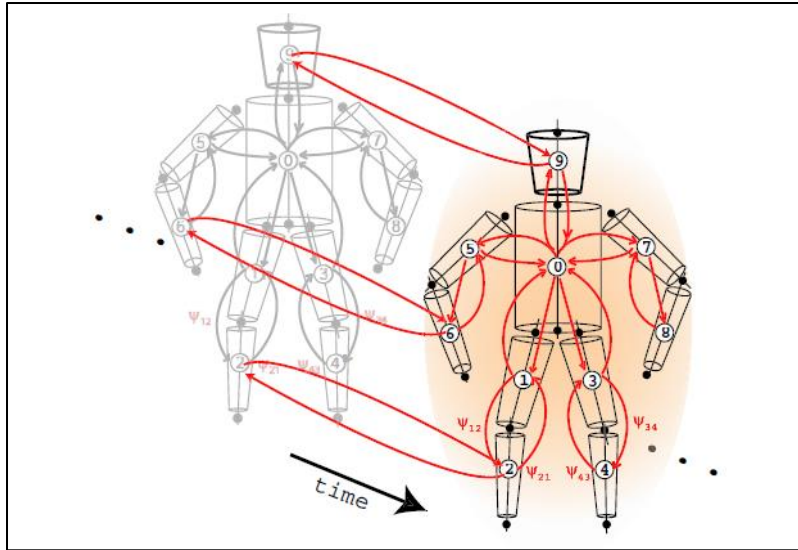
154. Dr. Bajaj relies on his opinions for claims 8/16 under Ground I with respect to Zhao as the basis for his opinion that claims 8/16 would have been allegedly obvious over the combination of Ozer and Zhao. For the reasons explained above under Ground I, Zhao does not disclose or render obvious the joint probability distribution of claims 8 and 16. See analysis in § IX.A.3 above. Accordingly, a POSITA would not have found claims 8/16 disclosed or rendered obvious by the combination of Ozer and Zhao.

D. Dr. Bajaj does not demonstrate that that TLP (Ground IV) discloses or renders obvious any challenged claim.

1. Claim 1[d]/9[d] “determining a probability that the object is in each image.”

155. Dr. Bajaj argues that TLP “disclosed or at least rendered this limitation 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 .” EX1004 ¶212. Dr. Bajaj argues that this is “equivalent to determining a probability an object is in each image.” EX1004 ¶212. I disagree. Dr. Bajaj provides 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. EX1009 at 4.

156. TLP makes clear to a POSITA that in its model, “[e]ach node in the graphical model represents ... a **limb** (Figure 1),” *i.e.*, a component. EX1009 at 1 (emphasis added); see EX2006 at 119:11-17, 120:9-121:3.



EX1009 at 1

A POSITA would have understood that TLP’s model does not include nodes that represent the object (*i.e.*, the human, *see* EX1004 ¶107), only components referred to as “limbs” or “parts.” 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 at 2 (emphasis added). However, there is no “body” node. 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 at 1. Accordingly, when TLP states that “each node has an associated likelihood defined over a rich set of image cues using a

learned Gibbs model,” a POSITA would have understood TLP as referring to each limb or part (not the human) having an associated likelihood. EX1009 at 1-2. This would have been clear to a 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 at 4 (emphasis added).

2. **Claim 1[e]/9[e] “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.”**
 - a. **Dr. Bajaj fails to address every limitation of element 1[e]/9[e] and therefore fails to show it was disclosed or rendered obvious by TLP.**

157. Relying on his opinions for claim 1[d]/9[d], Dr. Bajaj argues that “a POSITA would have known to further compare the probabilities to a ‘threshold’ to detect the object.” EX1004 ¶214. Dr. Bajaj argues that “[t]o infer belief in a model, a POSITA would have known to compare a determined probability against a threshold.” EX1004 ¶214. As with his analysis of Zhao, Steffens, and Ozer, however, Dr. Bajaj fails to consider this element as a whole with all of its limitations. Dr. Bajaj never explains 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. Nor does he show

that it would have been obvious to modify TLP to arrive at the claimed invention as a whole.

158. Rather than addressing the actual claim language, Dr. Bajaj claims that “a POSITA would have known to compare a determined probability against a threshold.” EX104 ¶214. But whether Dr. Bajaj’s vague and generic concept of “compar[ing] a determined probability against a threshold” was known in the art does not show that a POSITA would have found the specific requirements of claim 1[e]/9[e] obvious over TLP. It does not show that a 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.

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

159. Dr. Bajaj has not shown that TLP discloses or renders obvious this element. 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**.” As explained above, Zhao’s model includes nodes only for **limbs** or body **parts**—it has no object-level nodes. See analysis of claim 1[d] above. Alternatively,

if these limbs or parts are considered objects, then it has no components. Accordingly, the “image likelihood model” that Dr. Bajaj relies on only determines the likelihood that a body **part** or **limb** is present in an image. 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 at 5. A POSITA would therefore have understood that TLP’s model has no object level nodes, likelihood model, or 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.

160. While TLP uses non-parametric belief propagation with message passing, a POSITA 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). EX1009 at 4-5. 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. A 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 at 4. Furthermore, 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 at 4-5.

161. A POSITA would have further understood that TLP leads away from the claimed invention. 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 at 1. Consistent with that teaching, TLP teaches a framework based on detecting and assembling individual limbs, as explained above, and therefore lead a POSITA away from the claimed invention towards its limb-based detection framework.

3. Claim 3/11

162. Claim 3 reads: “The computer implemented method of claim 1, wherein the spatio-temporal model is a graphical model comprising nodes corresponding to each of the collection of components and to the object.” EX1001, claim 3.

163. Claim 11 reads: “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 .” EX1001, claim 11.

164. Claims 3 and 11 explicitly require “nodes corresponding to each of the collection of components **and** to the **object**.” Dr. Bajaj acknowledges that in TLP’s model “each graph node corresponds to a body part (upper leg, torso, etc.),” but argues that “the nodes correspond to each of the collection of components **and the object**.” EX1004 ¶¶204, 219. However, Dr. Bajaj does not identify an object node in TLP—he identifies only TLP’s part nodes and limb nodes, which demonstrates the absence of (1) object nodes in TLP and (2) the presence of both component and object nodes in TLP. EX1004 ¶¶204, 219. Furthermore, for the reasons I explained above with respect to elements 1[d]/9[d] and 1[e]/9[e] and below with respect to claims 8/16, TLP does not disclose or render obvious object nodes.

4. Claim 8/16

165. Claim 8 reads: “The computer implemented method of claim 1, wherein a joint probability distribution for the spatio-temporal model with N components is:

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

EX1001, claim 8.

166. Claim 16 reads: “The method of claim 9, wherein a joint probability distribution for the spatio-temporal model with N components is:

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

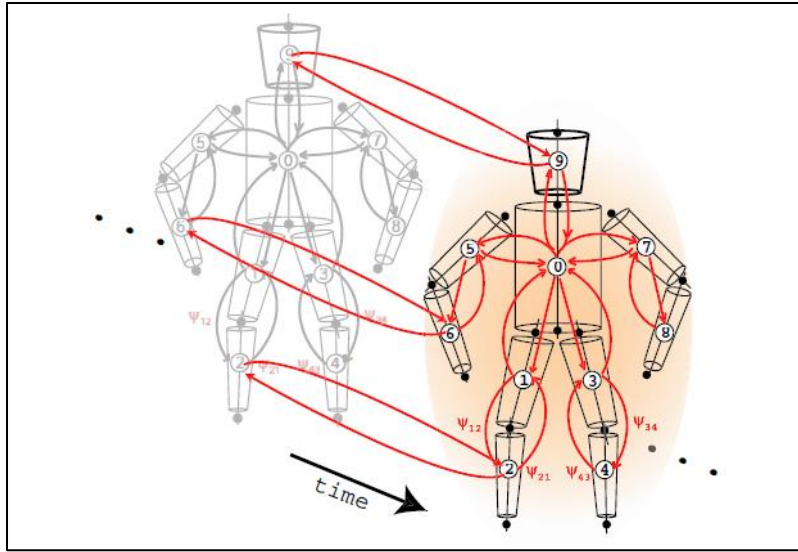
EX1001, claim 16.

167. Dr. Bajaj does not argue that TLP discloses claims 8 and 16. EX1004 ¶¶230-231. Dr. Bajaj argues that “TLP rendered Claims 8/16 obvious,” asserting that the joint probability distribution formulated in those claims “would have been well known to a POSITA” and is purportedly described in an article cited by TLP for its concept of non-parametric belief propagation. EX1004 ¶231. I disagree for multiple reasons.

168. First, TLP does not disclose or render obvious a model with object-level nodes or object-level constraints, which are required by the probability distribution of claims 8/16. See EX2006 at 70:9-17, 80:15-21. The joint probability distribution of claims 8 and 16 is defined in the ’980 patent’s specification at 3:47-4:8. The patent indicates what the terms in the distribution mean. X_t^O is the state of

the object “O” at time t, and X_t^C is the state of the object’s n-th component “C_n” at time t, where the object has N components. EX1001 at 3:53-4:8. Accordingly, as a POSITA would have understood, the patent’s spatio-temporal model explicitly models the object itself **and** its associated components at multiple points in time. As explained with respect to elements 1[d]/9[d] and 1[e]/9[e], TLP does not disclose or render obvious a model with object-level nodes.

169. The joint distribution of claims 8/16 also includes multiple terms that represent spatial compatibility between nodes (*i.e.*, for nodes in a given image) or temporal compatibility between nodes (*i.e.*, between nodes in different frames) that correspond to directed edges representing spatial or temporal probabilistic constraints. See analysis in § VI above. Those terms include $\psi_{ij}(X_i^O, X_j^O)$ (“the temporal compatibility of object state between frames i and j”); $\psi_{ik}(X_i^O, X_{i-k}^C)$ (“the spatial compatibility of the object and it’s components at frame i”); and $\psi_{ik}(X_{i-k}^C, X_i^C)$ (“the spatial compatibility between object components at frame i”). *Id.*; EX1001 at 4:49-51, 4:15-17, 2:59-60. As explained with respect to elements 1[d]/9[d] and 1[e]/9[e], TLP does not disclose or render obvious a model with object-level nodes. TLP’s Figure 1 confirms the absence of object-level nodes and the absence of object-level temporal constraints (directed edges).



EX1009 at 1

170. Furthermore, Dr. Bajaj has not shown that the joint probability formulation of claims 8/16 “would have been well known” or that EX1012 discloses that formulation. 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 at 1 (emphasis added), 3, 6 (“The proposed model extends the approaches of [6, 14, 22] to estimate not only the location, but also the appearance of occluded **parts**.” (emphasis added)). Dr. Bajaj identifies no disclosure in EX1012 of object-level representation in combination with component-level representation.

171. I declare that all statements made herein of my knowledge are true, and that all statements made on information and belief are believed to be true, and that

these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001, of Title 18 of the United States Code.

Executed on this 13th day of March 2024, in Webster, New York.

A handwritten signature in blue ink, appearing to read 'Eli Saber', is written over a horizontal line.

Eli Saber

APPENDIX A

Materials Considered

- U.S. Patent No. 7,436,980 (EX1001)
- File History of U.S. Patent No. 7,436,980 (EX1002)
- Provisional Application No. 60/574,799 (EX1003)
- Declaration of Chandrajit Bajaj, Ph.D. in Support of Petition for *Inter Partes* Review of U.S. Patent No. 7,436,980 (EX1004)
- The exhibits cited in Dr. Bajaj's declaration.
- Patent Owner's Preliminary Response (Paper 7)
- Decision Granting Institution of Inter Partes Review (Paper 12)
- Patent Owner's Request for Rehearing (Paper 13)
- Decision Denying Patent Owner's Request on Rehearing (Paper 19)
- Deposition of Dr. Chandrajit Bajaj (EX2006)