

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

U.S. PATENT NO. 7,436,980

Case IPR2023-00924

**PETITION FOR *INTER PARTES* REVIEW
UNDER 35 U.S.C. § 312 AND 37 C.F.R. § 42.104**

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	MANDATORY NOTICES	2
	A. Real Party-In-Interest	2
	B. Related Matters.....	2
	C. Counsel and Service Information.....	3
	D. 37 C.F.R. §42.8(b)(4): Service Information.....	3
III.	PAYMENT OF FEES	3
IV.	CERTIFICATION	3
V.	OVERVIEW OF CHALLENGE AND RELIEF REQUESTED.....	5
	A. Prior Art.....	5
	B. Relief Requested.....	6
VI.	NO DISCRETIONARY DENIAL	7
	A. No Prior Petition.....	7
	B. The Presented Grounds Are Dissimilar to the Previous Art and Arguments	7
	C. Efficiency, Fairness, and the Merits Support Institution	9
VII.	THE '980 PATENT.....	11
VIII.	PERSON OF ORDINARY SKILL IN THE ART	11
IX.	CLAIM CONSTRUCTION	12
X.	SPECIFIC GROUNDS FOR PETITION.....	12
	A. <u>Ground I</u> : Zhao Renders Claims 1-16 Obvious	13
	B. <u>Ground II</u> : Steffens In View Of Zhao Renders Claims 1-16 Obvious	35

C.	<u>Ground III</u> : Ozer In View Of Zhao Renders Claims 1-16 Obvious	53
D.	<u>Ground IV</u> : TLP Renders Claims 1-16 Obvious.....	65
XI.	SECONDARY CONSIDERATIONS	76
XII.	CONCLUSION.....	76

TABLE OF AUTHORITIES

	Page(s)
Cases	
<i>Apple Inc. v. Aire Tech. Ltd.</i> , No. IPR2022-01136, Paper 11 (P.T.A.B. January 4, 2023)	10
<i>Google LLC v. EcoFactor, Inc.</i> , IPR2021-01578, Paper 9 (P.T.A.B. Mar. 18, 2022)	10
<i>Sand Revolution II, LLC v. Continental Intermodal Grp. Trucking LLC</i> , No. IPR2019-01393, Paper 24 (P.T.A.B. June 16, 2020)	10, 11, 13, 22
<i>Wyers v. Master Lock Co.</i> , 616 F.3d 1231 (Fed. Cir. 2010)	76
Statutes	
35 U.S.C. §314.....	7, 9, 10, 11
35 U.S.C. §325	7, 8
Other Authorities	
37 C.F.R. §42	7

LIST OF EXHIBITS

Exhibit No.	Description
1001	U.S. Patent No. 7,436,980 ('980 Patent)
1002	File History of U.S. Patent No. 7,436,980
1003	Provisional Application No. 60/574,799
1004	Declaration of Dr. Chandrajit Bajaj
1005	Curriculum Vitae of Dr. Chandrajit Bajaj
1006	Liang Zhao, Dressed Human Modeling, Detection, and Parts Localization, The Robotics Institute, Carnegie Mellon University (2001) ("Zhao")
1007	U.S. Patent No. 6,301,370 ("Steffens")
1008	U.S. Patent No. 7,200,266 ("Ozer")
1009	Leonid Sigal, Sidharth Bhatia, Stefan Roth, Michael J. Black & Michael Isard, <i>Tracking Loose-limbed People</i> , IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2004) ("TLP")
1010	Program Documents for the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2004), http://tab.computer.org/pamitc/archive/cvpr2004/index.html
1011	Adam Baumberg & David Hogg, <i>Generating Spatiotemporal Models from Examples</i> , School of Computer Studies, University of Leeds, http://www.bmva.org/bmvc/1995/bmvc-95-040.pdf ("Baumberg")
1012	Erik B. Sudderth, Alexander T. Ihler, William T. Freeman & Alan S. Willsky, <i>Non-parametric Belief Propagation</i> , IEEE Conference on Computer Vision and Pattern Recognition, (June 2003), https://www.ics.uci.edu/~ihler/papers/tr2551.pdf
1013	Declaration of Jessica Benner regarding "Zhao", Carnegie Mellon University (April 4, 2023)

Exhibit No.	Description
1014	Declaration of Carl Megeski regarding “Zhao,” ProQuest (March 2, 2023)
1015	Leonid Sigal, Michael Isard, Benjamin H. Sigelman & Michael J. Black, <i>Attractive people: Assembling loose-limbed models using non-parametric belief propagation</i> , Advances in Neural Information Processing Systems 16, NIPS 1539–46 (MIT Press 2003)
1016	Hedvig Sidenbladh, Michael J. Black & David Fleet, <i>Stochastic tracking of 3D human figures using 2D image motion</i> , European Conference on Computer Vision, 702–18, LNCS 1843 (Springer-Verlag 2000)
1017	Noel A. C. Cressie, <i>Statistics for Spatial Data</i> , Chapters 1, 9, John Wiley and Sons (1993)
1018	Yaakov Bar-Shalom, X. Rong Li & Thiagalingam Kirubarajan, <i>Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software</i> , Chapters 1–2, 11, John Wiley and Sons, (2001)
1019	M. Isard, <i>PAMPAS: Real-valued graphical models for computer vision</i> , CVPR, vol. 1, at 613–20 (2003)
1020	D. L. Alspach & H. W. Sorenson. <i>Nonlinear Bayesian estimation using Gaussian sum approximations</i> , IEEE Trans. AC, 17(4): at 439–48, (August 1972)
1021	James M. Coughlan & Sabino J. Ferreira, <i>Finding deformable shapes using loopy belief propagation</i> , ECCV 453–68 (2002)
1022	F. De la Torre & M. J. Black., <i>Robust parameterized component analysis: Theory and applications to 2D facial modeling</i> , ECCV 653–69 (2002)
1023	S. L. Lauritzen, <i>Graphical Models</i> , Chapters 1, 3, Oxford University Press, Oxford (1996)

Exhibit No.	Description
1024	A. Doucet, N. de Freitas, & N. Gordon, editors. <i>Sequential Monte Carlo Methods in Practice</i> , Chapters 5, 16, Springer-Verlag, New York (2001)
1025	Vladimir Pavlovic, James M. Rehg, Tat-Jen Cham & Kevin Murphy, <i>A dynamic Bayesian network approach to figure tracking using learned dynamic models</i> , Int’l. Conf. on Computer Vision at 94–101 (1999)
1026	Y. Wu, G. Hua & T. Yu, <i>Tracking articulated body by dynamic Markov network</i> , Int’l. Conf. on Computer Vision at 1094–101 (2003)
1027	Christopher Bregler & Jitendra Malik, <i>Tracking people with twists and exponential maps</i> , CVPR, at 8–15 (1998)
1028	J. Deutscher, A. Davison & I. Reid, <i>Automatic partitioning of high dimensional search spaces associated with articulated body motion capture</i> , IEEE Computer Society Conference on Computer Vision and Pattern Recognition, at 669–76 (2001)
1029	Cristian Sminchisescu & Bill Triggs, <i>Covariance scaled sampling for monocular 3D body tracking</i> , IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, at 447–54 (2001)
1030	John MacCormick & Michael Isard, <i>Partitioned sampling, articulated objects, and interface-quality hand tracking</i> , European Conference on Computer Vision (2), at 3–19 (2000)
1031	Yacov Hel-Or & Michael Werman, <i>Constraint Fusion for Recognition and Localization of Articulated Objects</i> , International Journal of Computer Vision 19(1), at 5–28 (1996)

Exhibit No.	Description
1032	Y. Weiss & W. T. Freeman, <i>Correctness of belief propagation in Gaussian graphical models of arbitrary topology</i> , Neural Comp., 13:2173–2200 (2001)
1033	M. Fischler & R. Elschlager, <i>The representation and matching of pictorial structures</i> , IEEE. Trans. Computers, 22(1):67–92 (1973)
1034	M. Burl, M. Weber & P. Perona, <i>A probabilistic approach to object recognition using local photometry and global geometry</i> , European Conference on Computer Vision at 628–41 (1998)
1035	P. Felzenszwalb & D. Huttenlocher, <i>Efficient matching of pictorial structures</i> , IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, at 66–73 (2000)
1036	D. Ramanan & D. Forsyth, <i>Finding and tracking people from the bottom up</i> , IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. II, at 467–716 (2003)
1037	Andrew Blake & Michael Isard, <i>Active Contours</i> , Chapters 7, 8, Springer-Verlag (1998)
1038	D. L. Alspach, <i>Gaussian Sum Approximations in Nonlinear Filtering and Control</i> , Information Sciences 7,271–290 (1974)
1039	A. R. Tilley, <i>The Measure of Man and Woman: Human Factors in Design</i> H.D. Associates (1993)
1040	Paul Viola & Michael Jones, <i>Rapid Object Detection using a Boosted Cascade of Simple Features</i> , Conference on Computer Vision and Pattern Recognition (2001)
1041	Hedvig Sidenbladh & Michael J. Black, <i>Learning Statistics Learning the Statistics of People in Images and Video</i> ,

Petition for *Inter Partes* Review of U.S. Patent No. 7,436,980

Exhibit No.	Description
	International Journal of Computer Vision 54 (1/2/3), at 183–209 (2003)
1042	Jonathan S. Yedidia, William T. Freeman & Yair Weiss, <i>Generalized Belief Propagation</i> , Advances in Neural Info. Proc. Sys. 13, at 689–95 (2000)
1043	<i>VideoLabs, Inc. et al v. Meta Platforms, Inc. et al</i> , Case No. 1-22-cv-00680, Dkt. 15 (D. Del. Aug. 29, 2022)

Unless otherwise indicated, all emphasis added.

I. INTRODUCTION

Meta Platforms, Inc.; Instagram, Inc.; WhatsApp LLC; Meta Platforms Technologies, LLC; and GIPHY, Inc. (“Petitioners”) request *inter partes* review of Claims 1-16 (“Challenged Claims”) of U.S. Patent No. 7,436,980 (EX1001, the “’980 Patent”).

The Challenged Claims relate to the image processing field, and more specifically, to a computer implemented method for the “detection and tracking of objects in images,” such as video images. EX1001, 1:10-13. The ’980 Patent purportedly addresses a need for “a system and method of detecting and tracking an object, implementing component detection and performing inference over space and time.” EX1001, 1:22-24. But the claimed invention was already known in the prior art. *See, e.g.*, EX1004 (Sections VI, VIII). For example, the prior art included the use of “spatio-temporal models” for object detection, including in the form of graphical models. The prior art also disclosed “component detection” that measured objects in video images as a collection of components and probabilistic frameworks to detect and track objects. Accordingly, the Challenged Claims should be canceled as obvious and unpatentable.

II. MANDATORY NOTICES

A. Real Party-In-Interest

Meta Platforms, Inc., Instagram, Inc., WhatsApp LLC, Meta Platforms Technologies, LLC, and Giphy, Inc. are real parties-in-interest.

B. Related Matters

The '980 Patent is asserted against Petitioners in litigation in the District of Delaware, *VideoLabs, Inc., et al. v. Meta Platforms, Inc. et al.*, Civil Action No. 1-22-cv-00680 (D. Del.), filed May 24, 2022 (the "Delaware Action"). The Delaware Action may affect, or be affected by, decisions in this proceeding. Nothing herein is any endorsement of VideoLabs' or VL Collective IP LLC's position in any litigation.

C. Counsel and Service Information

Lead Counsel	Back-Up Counsel
W. Todd Baker (No. 45,265) todd.baker@kirkland.com Postal and Hand-Delivery Address: KIRKLAND & ELLIS LLP 1301 Pennsylvania Avenue, N.W. Washington, D.C. 20004 Telephone: (202) 389-3135 Facsimile: (202) 389-5200	Ellisen Shelton Turner (No. 54,503) ellisen.turner@kirkland.com Postal and Hand-Delivery Address: KIRKLAND & ELLIS LLP 2049 Century Park East, 37th Floor Los Angeles, CA 90067 Telephone: (310) 552-4200 Facsimile: (310) 552-5900 Jonathan D. Brit (No. 75,646) jonathan.brit@kirkland.com Postal and Hand-Delivery Address: KIRKLAND & ELLIS LLP 601 Lexington Avenue New York, NY 10022 Telephone: (212) 390-4209 Facsimile: (212) 446-4900

D. 37 C.F.R. §42.8(b)(4): Service Information

Petitioners concurrently submit a Power of Attorney, 37 C.F.R. §42.10(b), and consents to electronic service to Meta_VideoLabs@kirkland.com.

III. PAYMENT OF FEES

The undersigned authorizes the Office to charge to Deposit Account No. 506092 the 37 C.F.R. §42.15(a)(1) fee and any additional fees due for this Petition.

IV. CERTIFICATION

Petitioners certify under Rule 42.104(a) that the '980 Patent is available for IPR and that Petitioners are not barred or estopped from requesting an IPR of the

Petition for *Inter Partes* Review of U.S. Patent No. 7,436,980

Challenged Claims on the grounds identified. Petitioners certify: (1) Petitioners do not own the '980 Patent; (2) Petitioners (or any real party-in-interest) have not filed a civil action challenging the validity of any '980 Patent claim; (3) Petitioners filed this Petition within one year of the date they were served with a complaint asserting infringement of the '980 Patent; (4) estoppel provisions of 35 U.S.C. §315(e)(1) do not prohibit this IPR; and (5) this Petition is filed after the '980 Patent was granted.

V. OVERVIEW OF CHALLENGE AND RELIEF REQUESTED

A. Prior Art

The earliest potential effective filing date of the '980 Patent's claims is May 27, 2004.¹ Petitioners' challenge is based on the following prior art, for which a detailed overview is provided in the Declaration of Dr. Bajaj, which is incorporated by reference. EX1004, Section VIII (Overview of the Prior Art).

1. **“Dressed Human Modeling, Detection, and Parts Localization” by Liang Zhao (“Zhao”) (EX1006)**, is a dissertation submitted to The Robotics Institute of Carnegie Mellon University. *See* EX1006, 1. Zhao is dated July 26, 2001, and was publicly available no later than September 23, 2002. *See* EX1013, 3 (“the Zhao Dissertation would have been accessible to the public at the Engineering & Science Library by September 23, 2002”). Zhao was also stored in the ProQuest repository and publicly available at least as early as June 25, 2002. *See* EX1014, 3 (“Date first made available for sale (print or electronic)[:] 06/25/2002”). Thus, Zhao is prior art under at least pre-AIA 35 U.S.C. §102(a) and (b).

2. **U.S. Patent No. 6,301,370 (“Steffens”) (EX1007)**, filed on December 4, 1998, and granted on October 9, 2001, is prior art under at least pre-AIA 35 U.S.C. §102(e).

¹ Petitioner reserves the right to challenge the May 27, 2004 priority date or any evidence of prior invention.

3. **U.S. Patent No. 7,200,266 (“Ozer”) (EX1008)**, filed on August 27, 2003, and granted on April 3, 2007, is prior art under at least pre-AIA 35 U.S.C. §102(e).

4. **“Tracking Loose-limbed People” by Leonid Sigal, Sidharth Bhatia, Stefan Roth, Michael J. Black, and Michael Isard (“TLP”) (EX1009)**, was widely disseminated and made publicly available for the IEEE Computer Society Conference on Computer Vision and Pattern Recognition by at least April 9, 2004. EX1004, ¶¶65-66; *see also* EX1010, 1 (“Deadline for Manuscript Submission: 9th April, 2004”). Further, TLP was disclosed during prosecution (EX1002, 37-48) and is cited as prior art on the face of the ’980 Patent and its provisional application. EX1001, 1 (“OTHER PUBLICATIONS...Tracking loose-limbed people, CVPR, 2004”); EX1003, 12 (“References...Tracking Loose-limbed People, CVPR, 2004.”). Accordingly, TLP is prior art under at least pre-AIA 35 U.S.C. §102(a).

B. Relief Requested

Petitioners request cancellation of the Challenged Claims as follows:

Ground	Claims	Proposed Statutory Rejection
I	1-16	Obvious under §103 over Zhao
II	1-16	Obvious under §103 over Steffens in view of Zhao
III	1-16	Obvious under §103 over Ozer in view of Zhao
IV	1-16	Obvious under §103 over TLP

VI. NO DISCRETIONARY DENIAL

A. No Prior Petition

Neither the Petitioners nor anyone else have filed any IPR or PGR petitions against the '980 Patent. Thus, there is no basis for the Board to exercise its discretion under 35 U.S.C. §314(a) and 37 C.F.R. §42.108(a).

B. The Presented Grounds Are Dissimilar to the Previous Art and Arguments

The four grounds in this Petition have not been previously applied against the Challenged Claims, and each ground favors institution under the factors of 35 U.S.C. §325(d).

Ground I: Ground I favors institution because it presents new art and argument not previously considered by the examiner. Indeed, Ground I is a single reference ground based on Zhao, which was not cited, let alone applied, during prosecution. Further, the obviousness arguments based on Zhao are substantially different than what the Patent Office considered. As described below, Zhao alone renders the Challenged Claims obvious based on its teachings of both a spatio-temporal model and its probabilistic framework for detecting an object, whereas the Patent Office only considered references that each disclosed some but purportedly not all aspects of the Challenged Claims.

Ground II: Ground II, which combines Steffens with Zhao, further favors institution. Like Zhao, Steffens was not considered or cited during prosecution, and

the combined teachings of Steffens and Zhao presents arguments different than the art at issue during prosecution. Here, each reference describes the use of a spatio-temporal model and teaches the use of similarity measures to perform object detection, which a POSITA would have been motivated to combine. *See* Section X.B.1 (Motivation to Combine). By contrast, the Patent Office only considered art where “the object detection algorithms in [the] two references are vastly different” such that a POSITA would not have been motivated to combine them. EX1002, 163.

Ground III: Ground III, which combines Ozer and Zhao, also favors institution. During prosecution, the Patent Office only considered Ozer in view of Cham, determining that “the object detection algorithms in [the] two references are vastly different” such that they would not have been combined. *Id.* By contrast, Ozer and Zhao each describe closely related object detection algorithms, with both teaching spatio-temporal models and similarity measures to perform object detection, which a POSITA would have been motivated to combine. Institution is also favored because when allowing the Challenged Claims, the Patent Office did not have the benefit of the new references raised in this Petition (Zhao), as well as the aid of expert testimony that Petitioner provides from Dr. Bajaj.

Ground IV: Ground IV, based off TLP, favors institution. While TLP is cited in an IDS, the examiner did not discuss TLP and failed to apply it. Indeed, as described below in Section X, TLP renders each of the Challenged Claims obvious

and is substantially different than the Ozer and Cham combination discussed by the examiner. Moreover, to the extent the examiner is believed to have considered TLP when allowing the Challenged Claims, the examiner did not have the benefit of any expert testimony, which Petitioner now provides from Dr. Bajaj.

C. Efficiency, Fairness, and the Merits Support Institution

Pursuant to the Interim Procedure for Discretionary Denials in AIA Post-Grant Proceedings with Parallel District Court Litigation (June 21, 2022) (“Vidal Memo”), the Board should not exercise its discretion under §314(a) considering the Delaware Action, since the *Fintiv* factors support institution. Further, the Petition presents compelling evidence of unpatentability.

Factor 1: Institution will enable the Board to resolve patentability and if the Challenged Claims are deemed unpatentable, it would relieve the Court of the need to continue with that issue.

Factor 2: The Board will very likely issue its Final Written Decision before a trial occurs in the Delaware Action, where no answer has been filed, no discovery has taken place, no claim construction proceedings have occurred, and no scheduling order or trial date have been set.

Factor 3: All major case milestones in the Delaware Action are yet to occur, and deadlines for them have not been set, such that substantial work remains in the

Delaware Action. Indeed, the Delaware Action remains in its infancy at the pleading stage.

Factor 4: Petitioners stipulate in the Delaware Action that, if this IPR is instituted, Petitioners will not pursue invalidity of the Challenged Claims on the actual grounds raised in this IPR. Although not as broad as a *Sotera* type stipulation, the Board has concluded—subsequent to the Vidal Memo—that a *Sand Revolution* type stipulation effectively addresses the risk of duplicative efforts. *Apple Inc. v. Aire Tech. Ltd.*, No. IPR2022-01136, Paper 11 (P.T.A.B. January 4, 2023) (citing *Sand Revolution II, LLC v. Continental Intermodal Grp. Trucking LLC*, No. IPR2019-01393, Paper 24 at 11-12 (P.T.A.B. June 16, 2020) (informative)).

In addition, claims 1-16 are challenged in this Petition, but only claim 1 is asserted in the Delaware Action. The Board should address the validity of the additional claims for the public’s benefit. *See Google LLC v. EcoFactor, Inc.*, IPR2021-01578, Paper 9 (P.T.A.B. Mar. 18, 2022).

Factor 5: The parties have no similarity, but in view of the other *Fintiv* factors—which heavily weigh against the Board’s exercise of §314(a) discretion—the parties’ similarity is of limited weight.

Factor 6: This Petition’s merits are strong, which weighs against the Board exercising its discretion under §314(a). *Sand Revolution II, LLC*, IPR2019-01393, Paper 24 at 13. Further, the petition presents compelling evidence of unpatentability

such that “the evidence, if un rebutted in trial, would plainly lead to a conclusion that one or more claims are unpatentable by a preponderance of the evidence.” Vidal Memo at 4. Accordingly, the Board should not deny institution under *Fintiv*.

“Considering the *Fintiv* factors as part of a holistic analysis,” it would run counter to “the interests of the efficiency and integrity of the system” if this Board were “to deny institution of a potentially meritorious Petition.” *Sand Revolution II, LLC*, IPR2019-01393, Paper 24 at 14. Thus, the Board should decline to exercise its discretion under § 314(a).

VII. THE ’980 PATENT

An overview of the ’980 patent, including a summary of the technology, claims, and prosecution history is provided in the Declaration of Dr. Bajaj. EX1004, Section VI.

VIII. Person of Ordinary Skill in the Art

A POSITA at the time of the ’980 Patent would have had a bachelor’s degree in computer science, computer engineering, computer vision or visualization, or an equivalent field, and approximately two years of experience in software development for computer vision applications. EX1004, ¶16. Additional education might compensate for less experience, and vice-versa. *Id.*

IX. CLAIM CONSTRUCTION

For purposes of this petition only, Petitioners present no terms for construction. Although the parties have not reached the claim construction stage in the Delaware Action, it appears that Patent Owner may seek to construe the term “spatio-temporal model.” In Patent Owner’s opposition to a motion to dismiss under Section 101, Patent Owner contended:

The ’980 patent’s spatio-temporal model is a graphical model with nodes that correspond to each of the collection of components and to the object.

EX1043, 6. While Petitioners offer no position at this time on the construction of this term, the Asserted Grounds render the Challenged Claims unpatentable even under Patent Owner’s purported construction.

X. SPECIFIC GROUNDS FOR PETITION

In the ’980 Patent, Claims 1-8 are substantially identical to Claims 9-16. Claim 1 mirrors Claim 9 except for the preamble. The preamble of Claims 1 and 9 are substantively similar, except that Claim 1 recites “[a] computer implemented method for object detection” whereas Claim 9 recites “[a] computer readable medium embodying a computer program to perform method steps for object detection.” Claims 2-8 mirror Claims 10-16, respectively, except Claims 2-8 depend on Claim 1 whereas Claims 10-16 depend on Claim 9. Accordingly, the mirrored claims and elements are addressed together in the discussion of each ground. Further,

when addressing a dependent claim, the analysis from any independent claim that it depends on is incorporated by reference.

A. Ground I: Zhao Renders Claims 1-16 Obvious

1. Claims 1/9

a. Element 1/9[preamble]

Zhao disclosed this element or at least rendered it obvious. EX1004, ¶¶70-73.

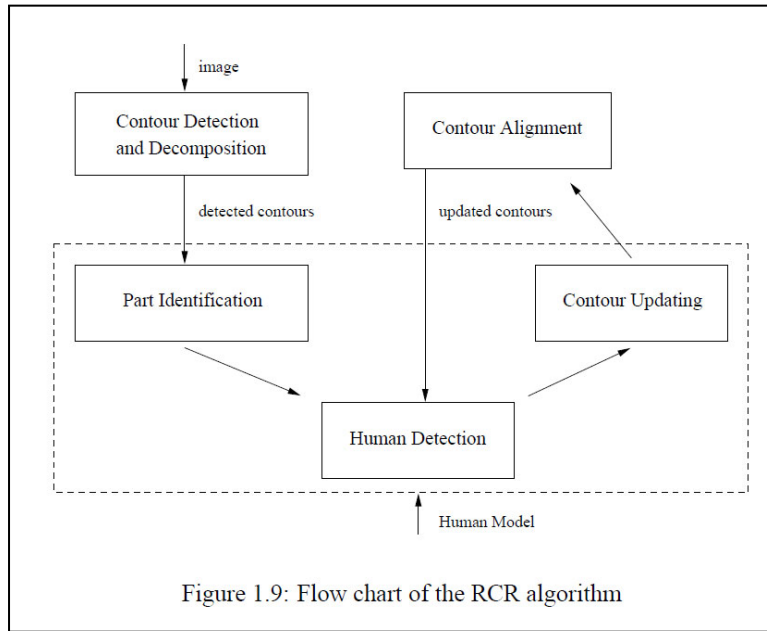
Zhao describes a method for object detection in the field of computer vision. EX1006, 17 (“This dissertation presents an integrated human shape modeling, detection, and body part localization vision system.”); 97 (“This thesis has addressed several computer vision problems...”). As such, a POSITA would have understood Zhao to describe a “computer implemented method” for objection detection. *Id.*, 19 (“Because humans are a class of objects with articulated parts and deformable shapes, research of human detection and body part localization will...shed light on general object detection and shape analysis.”). A POSITA would have understood this to further include “a computer readable medium embodying a computer program” to perform its method for object detection. EX1004, ¶¶72.

Accordingly, Zhao disclosed this element or at least rendered it obvious. EX1004, ¶¶70-73.

b. Element 1/9[a]

Zhao disclosed this element or at least rendered it obvious. EX1004, ¶¶74-86. To detect objects, Zhao uses an object model that takes the form of a “human model”

for a human to be detected. EX1006, 17-18, 31 (“Human modeling is an essential part of model-based human detection.”). The use of this model is illustrated in the flow chart of Figure 1.9:



EX1006, 26. The human object model described in Zhao—which includes spatial and temporal elements—disclosed or at least rendered the claimed “spatio-temporal model” obvious. EX1004, ¶75.

With respect to the “spatial” element, the ’980 Patent itself does not contend that incorporating spatial elements into an object model is a point of novelty, as it was “common to represent objects as collections of features with distinctive appearance, spatial extent, and position.” EX1001, 1:14-19. Indeed, a POSITA would have found it obvious for a model for an object to be detected to include spatial elements. EX1004, ¶76. Further, Zhao discloses that its “model is designed

to encode the shapes of the body parts and the context information—the size and *spatial relationships* between body parts.” EX1006, 69. Specifically, Zhao uses the spatial relationships between body parts as part of its model for a human to be detected in the form of four model matrices: (1) an aspect ratio vector, (2) a length ratio matrix, (3) a relative position vector, and (4) an orientation or posture vector:

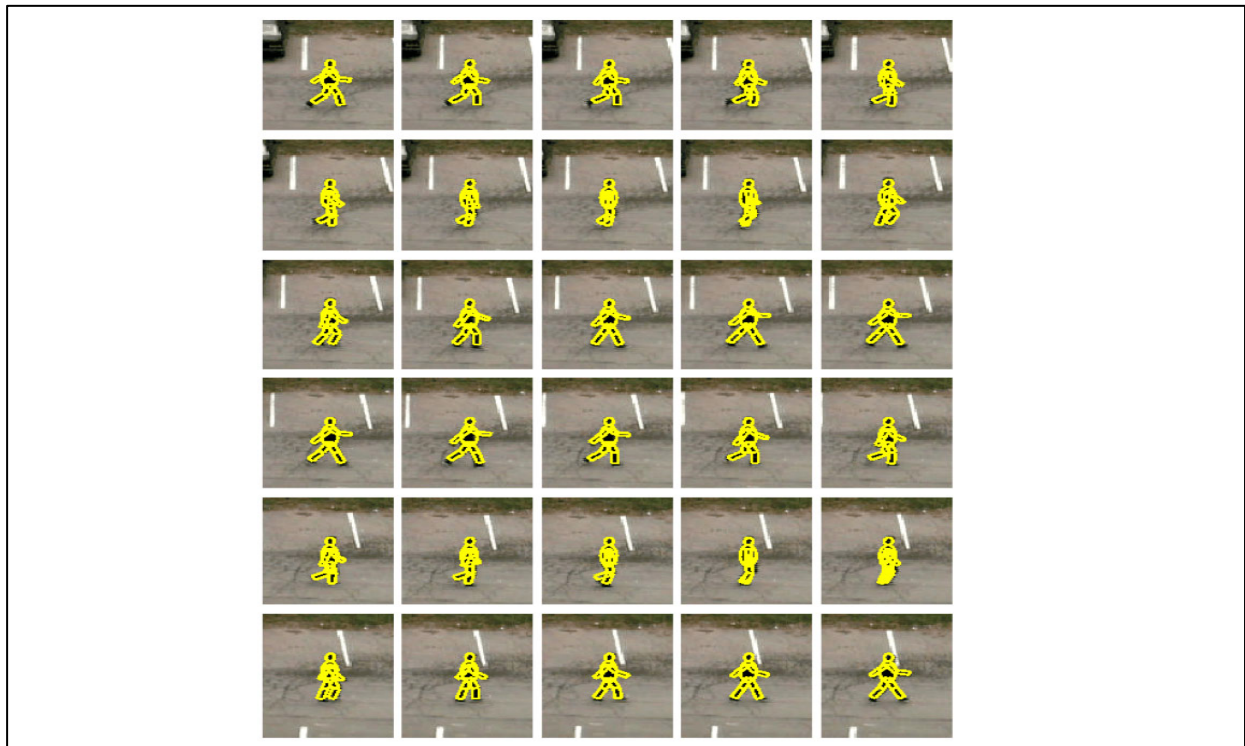
Assuming that the human body model consists of m body parts that are parameterized with vectors $f_1, f_2, \dots, f_m$, where $f_i = (a_i, l_i, x_i, y_i, \theta_i)$. Then the human body model is parameterized with four model matrices: the aspect ratio vector $A = \{a_1, \dots, a_m\}$, the length ratio 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 $\Theta = \{\theta_1, \dots, \theta_m\}$.

EX1006, 43. Accordingly, the “spatial” element of the claimed “spatio-temporal model” was disclosed or at least rendered obvious to a POSITA. EX1004, ¶77.

Zhao also disclosed or at least rendered obvious the claimed “spatio-*temporal* model” by incorporating temporal elements into its object model. As Zhao recognizes, numerous prior art methods include “motion-based” recognition procedures using “a sequence of images” for identification. EX1006, 82; *see also* EX1003, 5 (prior art methods for object detection included “temporal propagation”). In other words, it would have been known to a POSITA to incorporate temporal elements into an object model for an object to be detected by incorporating information measured from multiple image frames. *See* EX1004, ¶¶78-79 (citing EX1011 at 1, 9, “Generating Spatiotemporal Models from Examples” which details an approach for generating spatiotemporal models that “allow[] accurate temporal

extrapolation” and have “uses for tracking, recognition and data compression of deformable or articulated objects undergoing complex motions.”).

Accordingly, Zhao’s “RCR algorithm is useful for vision tasks such as object tracking as well by incorporating motion information into the model.” EX1006, 99. When detecting humans, Zhao explains that overlapping body parts can create ambiguity and be confused with one another. As such, “[m]otion information obtained from previous frames can be used to predict the orientations of the limbs and to solve the ambiguity.” EX1006, 91. Figure 5.6 presents a full cycle of a walking person:



EX1006, 94. According to Zhao, no motion information is used in the first half of the cycle, “while in the second half of the cycle, the prediction from previous frames

(constant angular velocity is assumed) is used to get better results of body part localization.” EX1006, 92.

Figure 5.5 illustrates the effect of incorporating temporal motion information in Zhao’s system and demonstrates the change of angle at a subject’s knee with time:

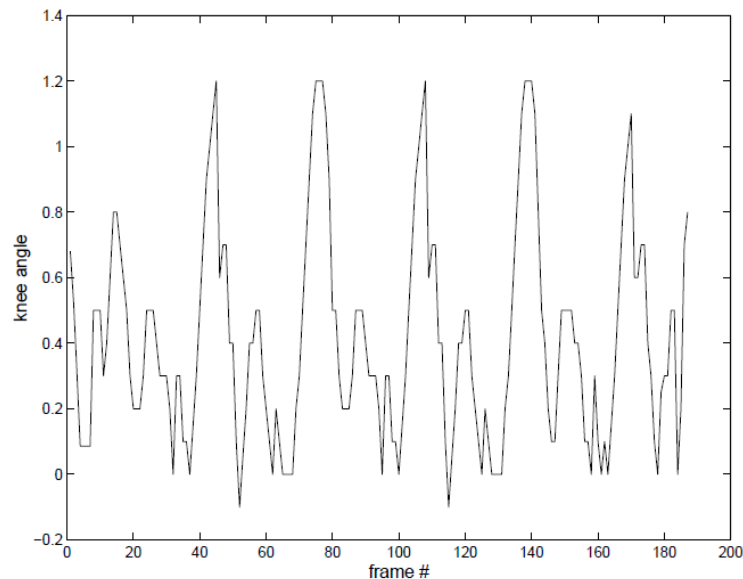
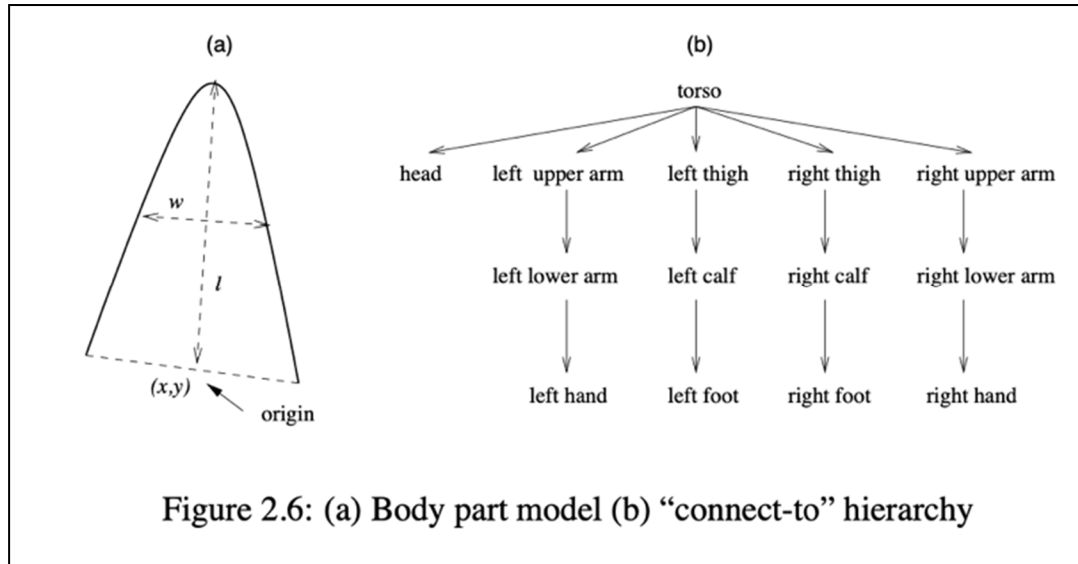


Figure 5.5: The change of angle at the knee with time

EX1006, 93. In the beginning (*i.e.*, frames 1-30), the left and right legs are sometimes confused so the knee angle is not well observed, but by incorporating motion information into the model (*i.e.*, frames 31-190), “an obvious pattern of walking cycle can be observed.” EX1006, 92. (“[M]otion information is not used from frames 1 to 30, but is from frames 31 to 190.”). Accordingly, a POSITA would have found the claimed “spatio-temporal model for an object to be detected” to be disclosed or at least rendered obvious in view of Zhao, as its model incorporates temporal elements for an object to be detected. EX1004, ¶82.

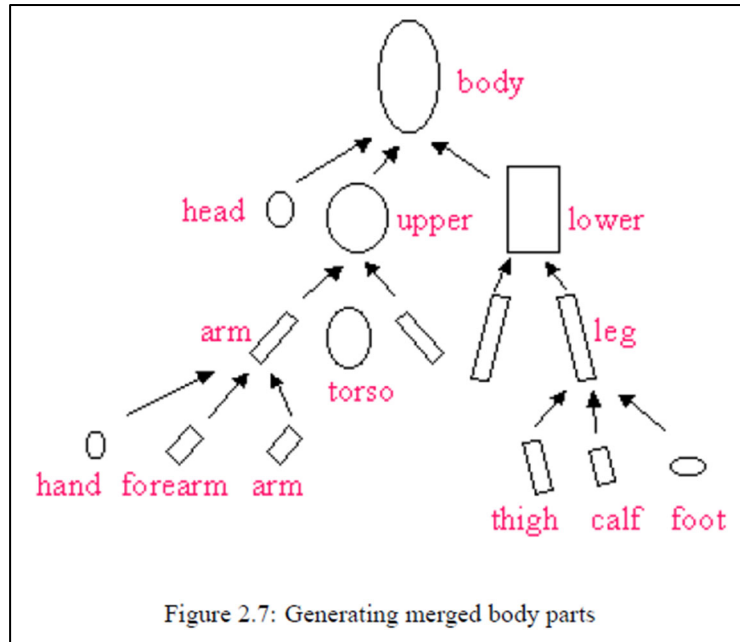
Further, to the extent the term “spatio-temporal model” is construed as “a graphical model comprising nodes corresponding to each of the collection of components and to the object,” Zhao still disclosed or at least rendered this element obvious. EX1004, ¶83. As an initial matter, a POSITA would have understood that graphical models for object detection were known in the prior art long before the ’980 Patent. *See, e.g.*, EX1006, 52 (“Pope and Lowe [prior art] modeled an object with a graph whose nodes represent the feature values and whose edges represent the spatial arrangement (symmetric, parallel) of the features.”); EX1004, ¶83. Indeed, the ’980 Patent admits to not having invented graphical models. EX1003, 5 (“We should note that the use of graphical models and belief propagation is not novel for object detection.”).

Further, a POSITA would have found this element disclosed or at least obvious in view of Zhao, as Zhao’s object model is also described as a graphical model with nodes corresponding to an object and a collection of its components. EX1004, ¶84. As Zhao explains with respect to human models, “[e]ach 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. The body parts are constrained to connect to each other at the joints.” EX1006, 42. The body model is defined as a graphical model that’s comprised of “ribbons” for these body parts, as shown in Figure 2.6(a):

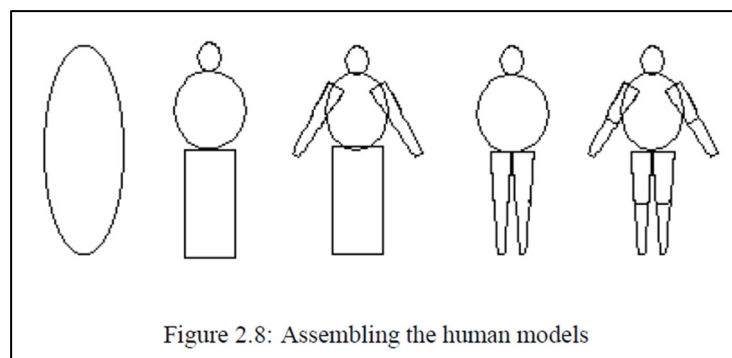


EX1006, 42 ("The body parts are modeled with ribbons as shown in Fig. 2.6(a)") (citations omitted). The ribbons are defined as graphical representations of a human's body parts and act as "nodes" corresponding to each of the collection of components and to the object.

Further, the Zhao object model incorporates a "coarse-to-fine modeling of human shapes" that merges multiple body parts. These "merged parts are also modeled with ribbons." EX1006 at 46. Figure 2.7 of Zhao illustrates this procedure:



EX1006, 45. In doing so, Zhao’s object model can handle the challenging problem of “shape variation from clothing,” which “may drastically change the appearance of a person.” EX1006, 45-46. Accordingly, Zhao’s object model, which is a “coarse-to-fine” model, includes “levels between the whole body [*i.e.*, the object] and the major body parts to include the merged body parts [*i.e.*, the collection of components],” as illustrated in Figure 2.8 of Zhao:

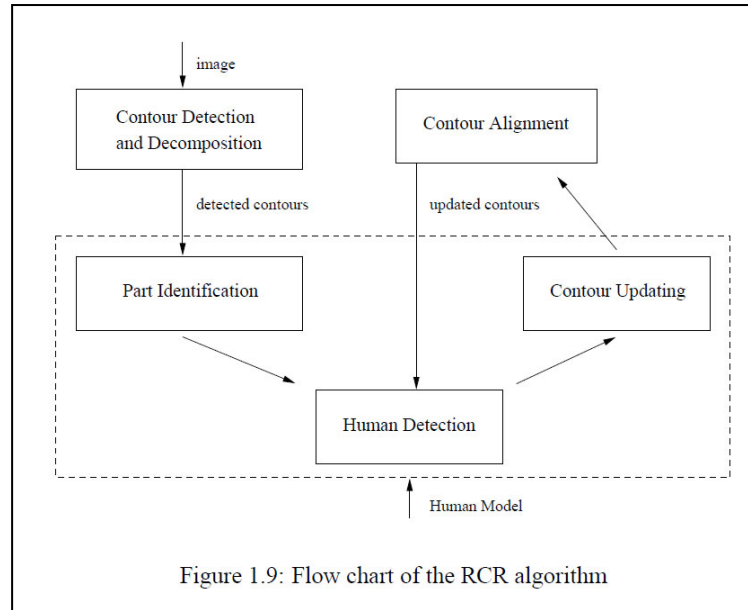


EX1006, 46; 48 (“all parts are organized in a hierarchy to facilitate coarse-to-fine shape decomposition and classification”).

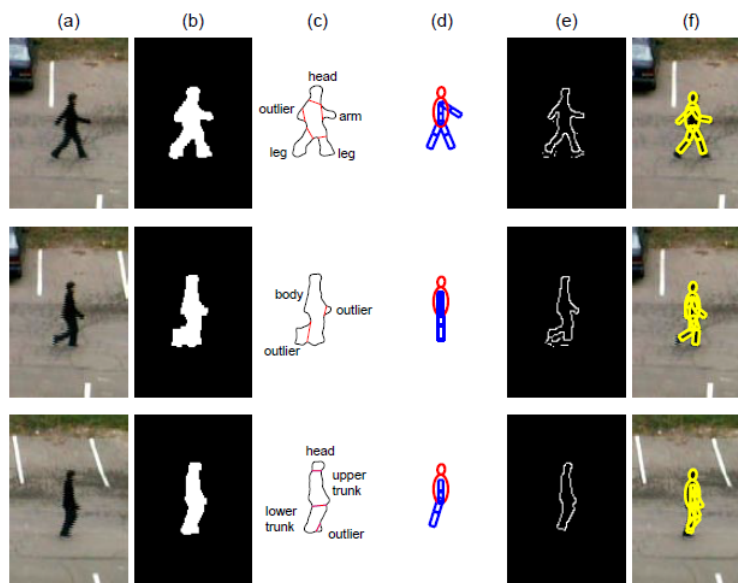
A POSITA would have found the purported spatio-temporal model to be disclosed or at least rendered obvious in view of Zhao, as the “ribbons” serve as “nodes” that correspond to each of the collection of components (*i.e.*, ribbons corresponding to component body parts) and the object (*i.e.*, ribbons corresponding to merged parts). EX1004, ¶86.

c. Element 1[b]

Zhao disclosed this element or at least rendered it obvious. EX1004, ¶87. Zhao describes a “human shape modeling, detection, and body part localization vision system,” which is used on image data including in videos. EX1006, 17, 41 (“the goal of this thesis is to detect people in an image”), 62 (depicting people and body parts detected in a “raw image.”). In one embodiment, Zhao applies its system as part of a “pedestrian detection system” which “has been tested on the videos of urban areas obtained from a stereo system mounted on the top of a minivan.” EX1006, 84-85. A POSITA would have understood a video to comprise a plurality of images. EX1004, ¶87. Figure 1.9 illustrates providing this video image data (*i.e.*, comprising a plurality of images) that includes the object:



EX1006, 26. Further, Figure 5.4 illustrates Zhao’s detection method used on a series of image frames (*i.e.*, video) of a person walking:



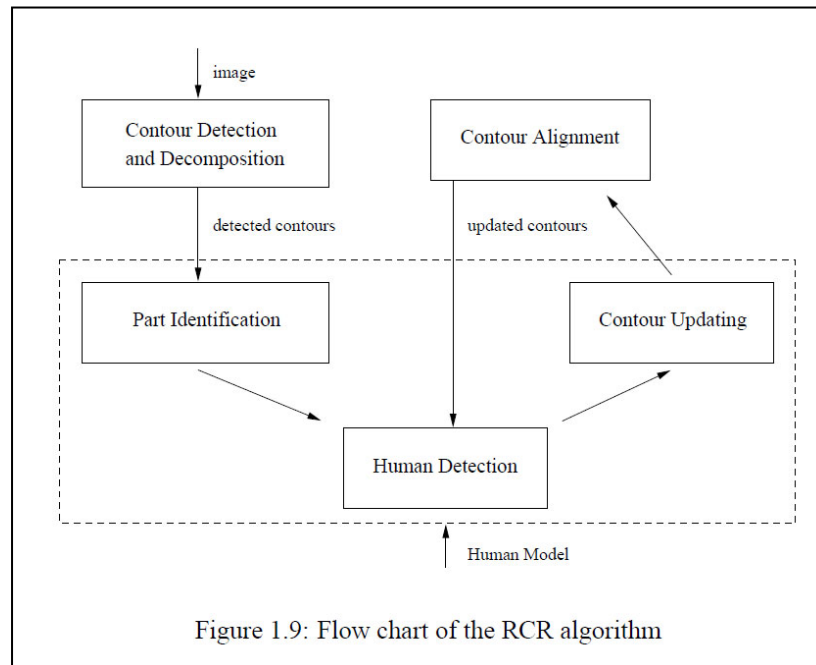
Id., 92. Accordingly, Zhao disclosed or at least rendered this limitation obvious.

EX1004, ¶87.

d. Element 1[c]

Zhao disclosed this element or at least rendered it obvious. EX1004, ¶¶88-90.

To measure the object in each image, Zhao begins with the step of “contour detection and decomposition” as illustrated in the flow chart of Figure 1.9:



EX1006, 26. Contour detection is the process of extracting silhouettes of objects. EX1004, ¶88. As Zhao recognizes, contour detection is “commonly used to overcome variable texture” and “many contour extraction methods have been proposed” in the prior art. EX1006, 20. After extraction, Zhao decomposes the contours or silhouettes into their component parts, as illustrated in Figure 2.4:

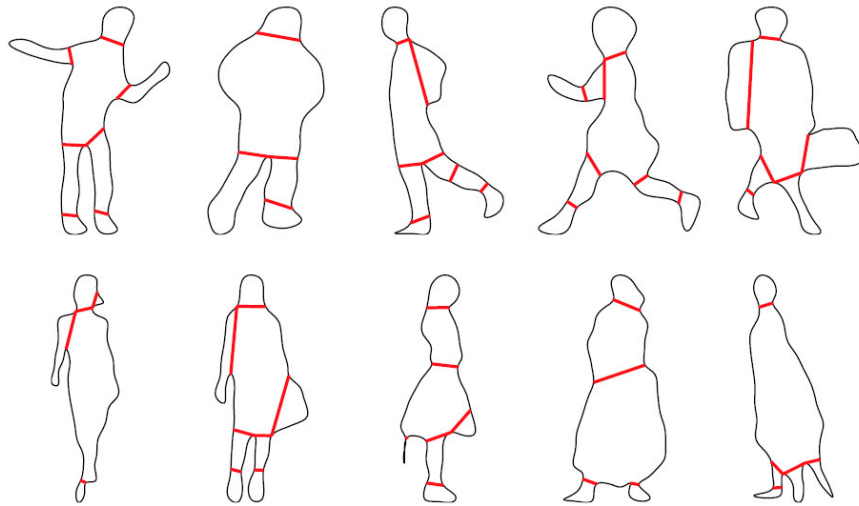


Figure 2.4: Example results of natural shape decomposition

EX1006, 40. Zhao then performs part identification and human detection. EX1006, 25-26. Accordingly, Zhao measures the object as “a collection of components.” EX1004, ¶88. Indeed, Zhao describes a “coarse-to-fine search procedure to identify and locate the body parts” (EX1006, 57) where components have an associated state vector defining the component’s position and scale within an image. EX1006, 43 (“In summary, 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.”). This matches the specification of the ’980 Patent, which explains that nodes representing an object’s components “have an associated state vector $X^T=(x,y,s)$ defining the component’s real-valued position and scale within an image.” EX1001, 3:50-52.

Accordingly, this element was disclosed or at least rendered obvious. EX1004, ¶¶89-90.

e. Element 1[d]

Zhao disclosed this element or at least rendered it obvious. EX1004, ¶¶91-97. After measuring an object as a collection of components, Zhao determines the probability that the object is in each image. To do so, Zhao employs a “Bayesian Similarity Measure,” which is a “part-based similarity measure [that] evaluates the resemblance between a contour and a model based on the best match between their body parts.” EX1006, 54; EX1004, ¶¶91-92.

Zhao begins with “[a] simple decision rule to perform human detection” that calculates the probability the “contour *C* corresponds to a person” in each image:

$$P(\text{person}|C) > \text{threshold}.$$

EX1006, 64; EX1004, ¶93. By applying Bayes’ theorem and conditioning the hypothesis for body part identification, determining the probability that a contour corresponds to a person (*i.e.*, determining a probability that the object is in each image) is re-formulated as:

$$P(C|H^*, \text{person}) P(\text{person}|H^*)$$

where H^* is the optimal or best hypothesis that a contour matches the human model. EX1006, 64-65; EX1004, ¶93.

$P(C|H, person)$ represents “the degree of resemblance between the matched pairs.” EX1006, 55. “The likelihood $P(C|H, person)$ derived above constrains the best match between the set of decomposed ribbons and the model body parts to be of similar relative sizes and positions as well as similar aspect ratios.” EX1006, 56. $P(person|H)$ is “proportional to the number of identified body parts—the more body parts being identified, the more likely the extracted contour is a person.” EX1006, 55. “The conditional probability $P(person|H)$ is estimated based on the number and types of the identified body parts. The more body parts being identified, the more likely the extracted contour is a person. And the presence of a fine level body part indicates a higher likelihood of the presence of a person than that of a coarse level body part does.” EX1006, 56. “[T]he following formula is designed to estimate $P(person|H)$ ”:

$$P(person|H) = \sum d_i w_i \quad (3.7)$$

EX1006, 56.

Accordingly, a POSITA would have recognized that Zhao disclosed this element or at least rendered it obvious. EX1004 at ¶¶91-97.

f. Element 1[e]

Zhao disclosed this element or at least rendered it obvious. EX1004, ¶¶98-100. After determining the probability that the object is in each image, Zhao compares the probability to a threshold to detect the object (*i.e.*, a human). Zhao

explains the “human detection problem” as “given a detected contour C , determine whether this contour represents the silhouette of a person or not.” EX1006, 64. Accordingly, Zhao begins with a simple decision rule to detect a human that compares the probability of an extracted contour matching a person with a threshold: “the contour C corresponds to a person if

$$P(\text{person}|C) > \text{threshold}.$$

Otherwise, C is not a person.” EX1006, 64. By applying Bayes’ formula and conditioning the hypothesis for body part identification, the decision rule is reformulated as: “the contour C corresponds to a person if

$$P(C|H^*, \text{person})P(\text{person}|H^*) > \text{threshold}$$

EX1006, 65; EX1004, ¶¶98-99.

Accordingly, a POSITA would have understood that Zhao disclosed or at least rendered this element obvious. EX1004, ¶¶98-100.

2. Claims 2/10

Zhao disclosed or at least rendered obvious Claims 2/10. EX1004, ¶¶101-103.

As described above for element 1/9[a], Zhao discloses a “spatio-temporal model” for an object to be detected. Zhao further teaches the use of detectors for each of the collection of components. For example, to detect body parts (*i.e.*, the collection of components), Zhao explains that its method employs detection algorithms that match decomposed body parts against a “set of model body parts.”

EX1006, 54. Further, “ $P(C|H, person)$ evaluates the degree of resemblance between the matched pairs, while $P(person|H)$ is proportional to the number of identified body parts—the more body parts being identified, the more likely the extracted contour is a person.” EX1006, 55. Accordingly, it was known to a POSITA for a spatio-temporal model for an object to be detected to be comprised of “detectors” where each of the collection of components (*i.e.*, body parts) is matched with a “set of model body parts.” EX1004, ¶102. As such, Claims 2/10 were disclosed or at least rendered obvious.

3. Claims 3/11

Zhao disclosed or at least rendered obvious Claims 3/11. EX1004, ¶¶104-105.

Further, the additional requirement of Claim 3/11—“wherein the spatio-temporal model is a graphical model comprising nodes corresponding to each of the collection of components and to the object”—matches the Patent Owner’s purported construction of the term “spatio-temporal model.” Accordingly, Claims 3/11 were disclosed or at least rendered obvious for the same reasons discussed for element 1/9[a] and said purported construction. As described above, a POSITA would have found the purported spatio-temporal model to be disclosed or at least rendered obvious in view of Zhao as it describes a graphical model where “ribbons” serve as “nodes” that correspond to each of the collection of components (*i.e.*, the component body parts) and the object (*i.e.*, the merged parts that form a human). EX1004, ¶105.

4. Claims 4/12

Zhao rendered Claims 4/12 obvious. EX1004, ¶¶106-110.

As described above in Claim 1, Zhao determines the probability that an object is in each image based on identified body parts. EX1004, ¶107. But Zhao states that in detecting objects, overlapping body parts can create ambiguity and be confused with one another. So, “[m]otion information obtained from previous frames can be used to predict the orientations of the limbs and to solve the ambiguity.” EX1006, 91. Accordingly, a POSITA would have understood that Zhao describes 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. EX1004, ¶¶107-108. Notably, analyzing data in image frames from video by using measurements from adjacent frames, including previous and later frames, was widespread and commonly understood in image processing. EX1004, ¶109. Accordingly, while Zhao explicitly describes this principle by detecting an object according to measurements of the object as a collection of components determined from a “prior image,” a POSITA would have further understood that doing so with a “later image” too would have been obvious. EX1004, ¶¶106-110.

As such, Claims 4/12 would have been obvious.

5. Claims 5/13

Zhao disclosed or at least rendered obvious Claims 5/13. EX1004 at ¶¶111-115.

As described above for element 1/9[a], and for Claim 4/12, Zhao describes a spatio-temporal model for an object to be detected and tracked. Zhao states that in detecting objects, overlapping body parts can create ambiguity and be confused with one another. So, “[m]otion information obtained from previous frames can be used to predict the orientations of the limbs and to solve the ambiguity.” EX1006, 91. Accordingly, a POSITA would have known to define a “temporal window” in view of Zhao’s teachings to look to measurements from “previous frames.” EX1004, ¶112. Indeed, a POSITA would have known that a widespread and common technique in image processing was to define a temporal window of image frames to analyze data. *Id.* Using measurements in that temporal window for analyzing the current image was also known. *Id.*

Further, a POSITA would have understood that measurements of the object’s components within that temporal window (*i.e.*, a window of the “previous frames”) are passed to components detected in the current image. EX1004, ¶113. Figure 5.5 illustrates the effect of passing measurements of components detected in the temporal window (*i.e.*, the window of previous frames) in Zhao’s system and demonstrates the change of angle at a subject’s knee with time:

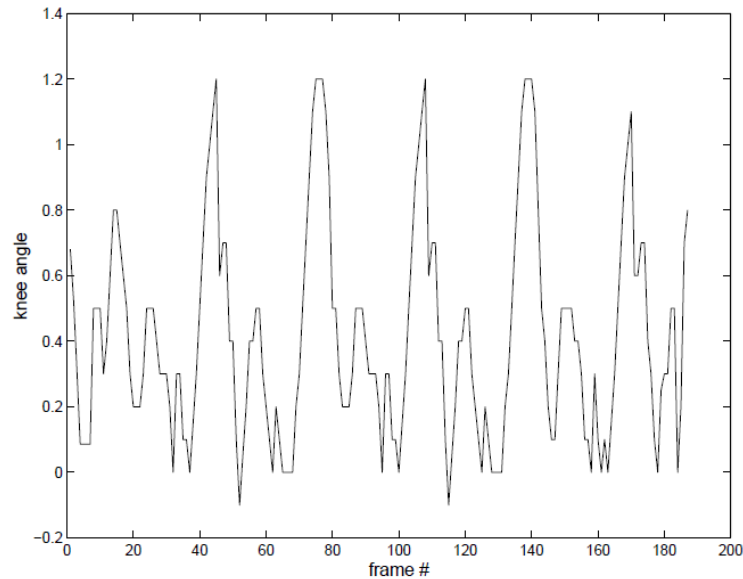
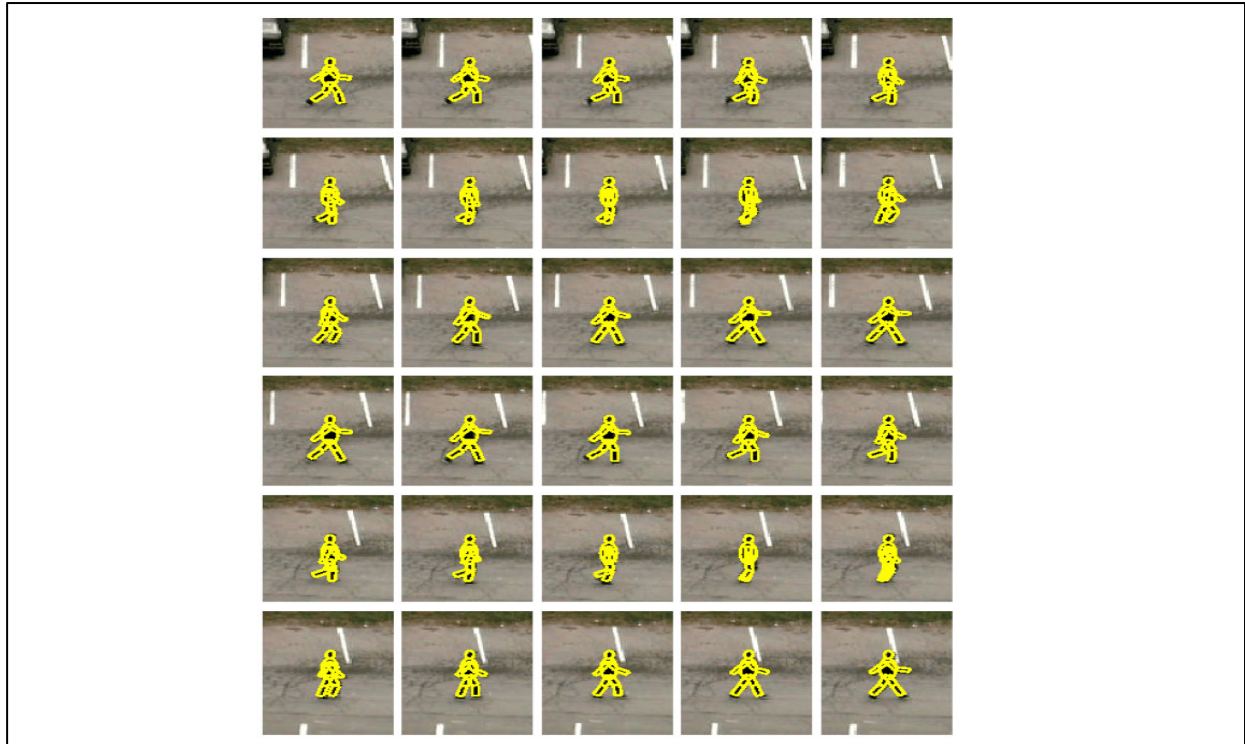


Figure 5.5: The change of angle at the knee with time

EX1006, 93.

In the beginning (*i.e.*, frames 1-30), the left and right legs are sometimes confused so the knee angle is not well observed, but by incorporating motion information into the model (*i.e.*, frames 31-190), “an obvious pattern of walking cycle can be observed.” EX1006 at 91-92 (“[M]otion information is not used from frames 1 to 30, but is from frames 31 to 190.”). As another example, Figure 5.6 presents a full cycle of a walking person:



EX1006 at 94. According to Zhao, no motion information is used in the first half of the cycle, “while in the second half of the cycle, the prediction from previous frames (constant angular velocity is assumed) is used to get better results of body part localization.” EX1006, 92. As such, measurements of the body parts from previous frames (*i.e.*, the “prediction from previous frames”) are passed to the components detected in the current frame. *Id.*

Accordingly, Claims 5/13 were disclosed or at least rendered obvious. EX1004, ¶¶111-115.

6. Claims 6/14

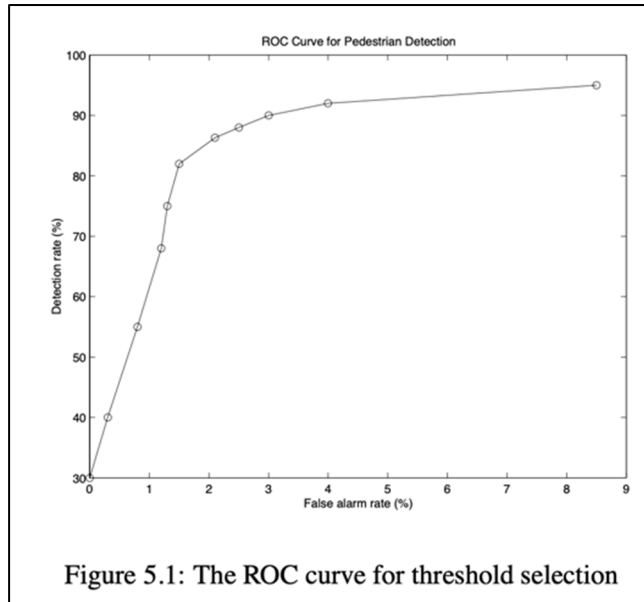
Zhao disclosed or at least rendered obvious Claims 6/14. EX1004 at ¶¶116-117.

As described above for element 1/9[d], Zhao describes determining the probability that an object is in each image. In determining said probability, Zhao states, “[t]he above probability distributions provide metrics to evaluate the ***shape, size relationship, and configuration*** similarities between the detected contour and the human body model.” EX1006 at 45. A POSITA would have understood that the metrics of Zhao used to evaluate the shape, size relationship, and configuration are equivalent to the determination of a probability for a position and a size of the object in each image as required by claims 6/14. EX1004 at ¶117.

7. Claims 7/15

Zhao disclosed or at least rendered obvious Claims 7/15. EX1004 at ¶¶118-119.

As described above for element 1/9[e], Zhao uses a threshold as part of its probabilistic framework for detecting an object. Zhao additionally explains that this threshold can be selected empirically based on an “ROC curve” and the rate of false detections “generated by testing the RCR algorithm on 20 sequences (100 frames/sequences) of urban street scenes,” as illustrated in Figure 5.1:



EX1006, 84-85. By basing the threshold off the rate of false detections, which was measured by applying and testing the system, a POSITA would have understood that the threshold is determined “empirically.” A POSITA would have understood that “empirical” results are those that are determined through experimentation or observation, as opposed to theoretical or arbitrary measures. EX1004, ¶119. Accordingly, a POSITA would have understood that Zhao describes that the threshold may be determined empirically. *Id.*

8. Claims 8/16

Zhao rendered Claims 8/16 obvious. EX1004 at ¶¶120-122.

As described in Section X.A.1. (elements 1/9[d]-[e]), Zhao presents a probabilistic framework for a spatio-temporal model. Accordingly, the joint probability used in Zhao, and known to a POSITA, is easily representable in the form of the joint probability distribution formulated by Claim 8. *See* EX1004, ¶121

(“The equation represents a typical factorized form of the joint probability distribution $P(\dots)$ for a graphical model G of an object O with state X_i^O at time frame i of the video and $\{k = 1..N\}$ components state X_i^{C-k} .”). This form of the joint probability would have been known to a POSITA as a pairwise Markov field representation of a spatio-temporal graphical model. *Id.* (citing EX1012 Sudderth, Sec 2.1).

B. Ground II: Steffens In View Of Zhao Renders Claims 1-16 Obvious

1. Motivation to Combine

Steffens teaches a method for object detection in video images. EX1007, 1:14-17. Zhao similarly concerns object detection in the field of computer vision. EX1006, 17 (“This dissertation presents an integrated human shape modeling, detection, and body part localization vision system.”); 97 (“This thesis has addressed several computer vision problems...”). Because Steffens and Zhao each concern object detection, a POSITA would have naturally looked to and combined their complementary teachings. EX1004, ¶¶123-124. Indeed, the similarities between Steffens and Zhao are substantial, such that a POSITA would have had reasonable expectation of success in combining Steffens and Zhao, as they both describe similar methods and algorithms for object detection, including through the use of spatio-temporal models. *Id.* Further, Zhao explains that its object detection “will also shed light on general object detection and shape analysis,” such that a POSITA would

have had a reasonable expectation of success in applying its teachings to frameworks for detecting other types of objects, such as faces in Steffens. *Id.*

As part of its method for object detection, Steffens describes “identifying the object based on a similarity” between an image and a corresponding model. EX1007, Claim 13; *see also* 7:63-67 (“The search for 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.”). But Steffens does not precisely explain how an object is identified based on a similarity measure; just that “confidence values”...“indicate the reliability of face detection and which are generated during bunch graph matching.” EX1007, 5:12-13. As a POSITA would have known, similarity measures may include probabilistic frameworks for identifying objects. EX1004, ¶125. One such “similarity measure” is expressly described in Zhao as a “Bayesian Similarity Measure,” which “combines the local and global relationship constraints into a single equation to evaluate the degree of resemblance.” EX1006, 25. Like Steffens, Zhao describes a method where an object is identified based on a similarity between an image and a corresponding model (*i.e.*, “the resemblance between a contour [of an object in an image] and a model”). EX1006, 54. Accordingly, POSITA would have been motivated to combine Zhao’s probabilistic “Bayesian Similarity Measure” with Steffens and would have had a reasonable expectation of success, as

Zhao is elaborating and improving on a method already contemplated by Steffens. EX1004, ¶125.

2. Claims 1/9

a. Element 1/9[preamble]

This element was disclosed or at least rendered obvious over Steffens in view of Zhao. EX1004, ¶¶127-128. As an initial matter, this element was disclosed or at least rendered obvious by Zhao, as described above for Ground I and incorporated by reference.

Further, Steffens disclosed or at least rendered this element obvious. EX1004, ¶128. According to Steffens, its “present invention relates to vision-based object detection and tracking, and more particularly, to systems for detecting objects in video images, such as human faces, and tracking and identifying objects in real time.” EX1007, 1:14-17. As a system and method relating to object detection for video images, a POSITA would have understood Steffens to describe a computer implemented method for object detection. EX1004, ¶128. A POSITA would have understood this to further include “a computer readable medium embodying a computer program” to perform that method. EX1004, ¶128; EX1007, 4:55-61

(describing image recognition process of the invention being implemented in either “program code” stored in RAM or “application specific hardware”).

b. Element 1/9[a]

This element was disclosed or at least rendered obvious over Steffens in view of Zhao. EX1004, ¶¶129-137. As an initial matter, this element was disclosed or at least rendered obvious by Zhao, as described above for Ground I and incorporated by reference.

Steffens further disclosed or at least rendered this element obvious. EX1004, ¶130. As part of its method for object detection, Steffens provides a model for an object to be detected, which is a “spatio-temporal model.”

First, with respect to the “spatio” element, Steffens explains that, for its model for an object to be detected, “the nodes of the labeled graph refer to points on the object and are labeled by jets,” and “[e]dges of the graph are labeled with distance vectors between the nodes.” EX1007, 8:58-60. Further, “[n]odes and edges define the graph topology. Graphs with equal geometry may be compared.” EX1007, 60-62. Accordingly, a POSITA would have understood that Steffens describes a model where “spatial” elements are used for an object to be measured. EX1004, ¶131. For example, by incorporating the “distance vectors between the nodes,” the model is incorporating spatial information. As such, Steffens disclosed or at least rendered the “spatio” aspect of the spatio-temporal model obvious.

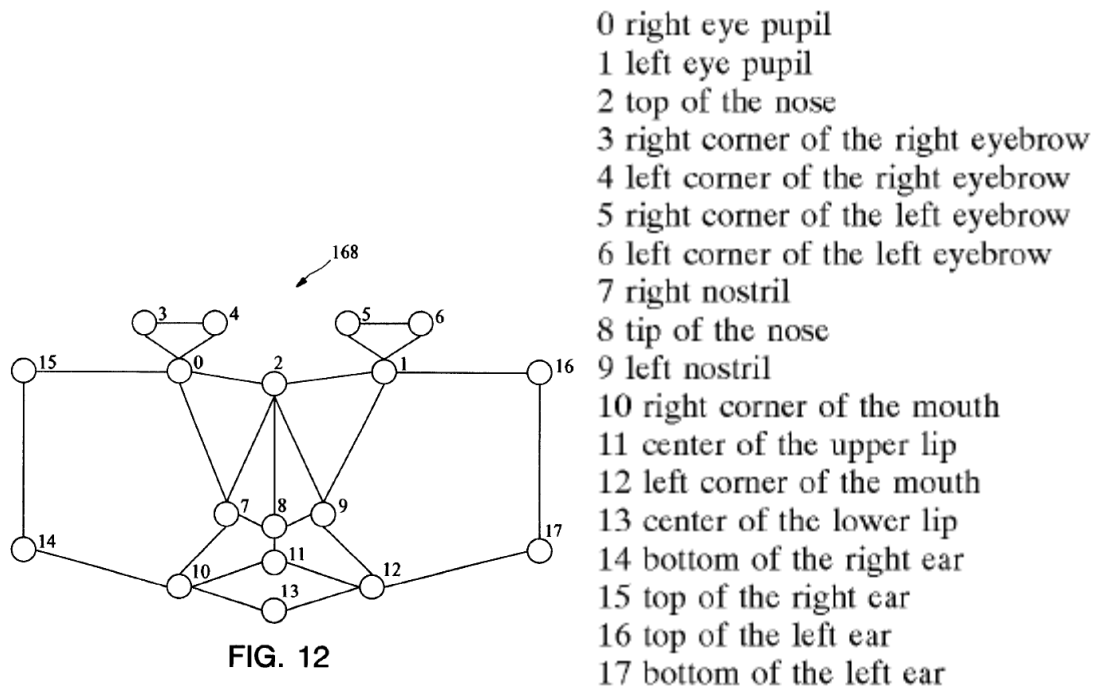
Second, with respect to the “temporal” element, Steffens explains that, for its model for an object to be detected, when “the object is in a sequence of images . . . the step of detecting an object further includes tracking the object between image frames based on a trajectory associated with the object. Also, the step of locating the nodes includes tracking the nodes between image frames and reinitializing a tracked node if the node's position deviates beyond a predetermined position constraint between image frames.” EX1007, 2:6-13. Steffens claims, as part of its process for recognizing objects, “tracking the object between image frames based on a trajectory associated with the object,” “tracking the nodes between image frames,” and “reinitializing a tracked node if the node’s position deviates beyond a predetermined position constraint between image frames.” *See, e.g.*, EX1007, Claims 14, 16, and 17. Accordingly, a POSITA would have understood that Steffens describes a model for an object to be detected that includes the use of “temporal” information, including for example the location of an object and its nodes between image frames. As such, Steffens disclosed or at least rendered obvious the “temporal” element of the spatio-temporal model.

Further, to the extent the term “spatio-temporal model” is construed as “a graphical model comprising nodes corresponding to each of the collection of components and to the object,” Steffens still disclosed or at least rendered this element obvious. As an initial matter, a POSITA would have understood that

graphical models for object detection were known in the prior art long before the '980 Patent. *See supra*, Ground 1, limitation 1[a]; EX1004, ¶133. In any event, a POSITA would have further found this element to be taught by Steffens, as Steffens' model is also described as a graphical model with nodes corresponding to an object and a collection of its components. EX1004, ¶133.

Steffens' model uses "bunch graph matching" to conduct similarity comparisons and identify an object. A bunch graph is a data structure like an image graph, which "is used to describe the aspects of an object." EX1007, 8:56-57, 9:22-25. The graph includes "nodes" that refer to points on the object, which are labeled with "jets." For a bunch graph, "instead of attaching only a single jet to each node, a whole bunch of jets 172 (a bunch jet) are attached to each node." EX1007, 9:22-25. Jets are derived from facial images and to form a bunch graph, "a collection of facial images (the bunch graph gallery) is marked with node locations at defined positions of the head," which are called "landmarks." EX1007, 9:25-29. Accordingly, Steffens describes providing a "graphical model comprising nodes corresponding to each of the collection of components and to the object."

Further, Steffens explains that a model graph that is designed for finding a human face is illustrated in Figure 12 with nodes corresponding to a collection of components:



EX1007, 13. Steffens teaches the use of a “coarse-to-fine” approach using fewer nodes to improve computational efficiency. EX1007, 15:32-41. After coarse landmark finding has been achieved, higher precision localization may be performed on higher resolution images for precise finding of a particular facial feature. An example of the coarse-to-fine procedure is shown in Figure 14:

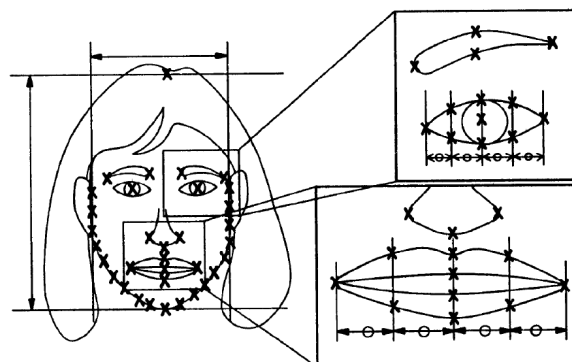


FIG. 14

EX1007, 15. As illustrated above, Steffens includes a “coarse” approach using fewer nodes that correspond to the object and its components. For example, a single node is used to represent an eye. EX1004, ¶137. Steffens then includes a “fine” approach using more nodes, including for example various nodes outlining the eye and a node at the center of the pupil. Accordingly, Steffens disclosed or at least rendered obvious providing a spatio-temporal model comprised of “nodes” corresponding to each of the collection of components and to the object (*i.e.*, a coarse-to-fine approach). *Id.*

c. Element 1/9[b]

This element was disclosed or at least rendered obvious over Steffens in view of Zhao. EX1004, ¶¶138-140. As an initial matter, this element was disclosed or least rendered obvious by Zhao, as described above for Ground I and incorporated by reference.

Steffens further disclosed this element or at least rendered it obvious. EX1004, ¶139. Steffens states, “[t]he present invention relates to vision-based object detection and tracking, and more particularly, to systems for detecting objects in *video images*, such as human faces...” EX1007, 1:14-16. Accordingly, “[t]he object recognition process 10 operates on digitized video image data provided by an image processing system 12. The image data includes an image of an object class, such as a human face. The image data may be a single video image frame or a series of sequential

monocular or stereo image frames.” EX1007, 3:48-53; EX1004 at ¶¶139-140. Accordingly, a POSITA would have understood that Steffens teaches providing a video comprising a plurality of images including the object, as its system “operates on digitized video image data provided by an image processing system” where “[t]he image data includes an image of an object class, such as a human face.” *Id.* As such, this element was disclosed or at least rendered obvious.

d. Element 1/9[c]

This element was disclosed or at least rendered obvious over Steffens in view of Zhao. EX1004, ¶¶141-143. As an initial matter, this element was disclosed or least rendered obvious by Zhao, as described above for Ground I and incorporated by reference.

Steffens further disclosed or at least rendered this element obvious. EX1004, ¶142. As described in element 1/9[a], Steffens uses a graph matching technique for measuring an object. In doing so, “[t]he nodes 164 of the labeled graph refer to points on the object and are labeled by jets 160. Edges 166 of the graph are labeled with distance vectors between the nodes. Nodes and edges define the graph topology.” EX1007, 8:58-61. This matches the process described in the specification of the ’980 Patent, which explains that nodes representing an object’s components “have an associated state vector $X^T=(x,y,s)$ defining the component’s real-valued position and scale within an image.” EX1001, 3:50-52; EX1004, ¶142. Like the ’980 Patent

specification, which describes nodes with an associated state vector to measure an object as a collection of components, Steffens also describes using nodes with an associated state vector (*e.g.*, distance vectors between the nodes) to measure the object as a collection of components. *Id.*

Accordingly, a POSITA would have found this element disclosed or at least obvious. EX1004, ¶¶141-143.

e. Element 1/9[d]

Steffens in combination with Zhao disclosed or at least rendered this element obvious. EX1004, ¶¶144-145. As an initial matter, Steffens describes its method for object detection as using a “similarity” measure. While this is not explicitly stated as calculating a “probability” that an object is an image, doing so would have been obvious to a POSITA. EX1004, ¶144. Indeed, probabilistic frameworks for object detection were widely known in the prior art and frequently used. *See* EX1004, ¶144.

Moreover, a POSITA would have been motivated to look for a probabilistic framework, such as Zhao, to combine with the teachings of Steffens. EX1004, ¶145; *see, supra*, Section X.B.1 (Motivation to Combine). Like Steffens, Zhao uses a similarity measure for the identification of objects in video data. As explained in Ground I, this includes the use of a “Bayesian Similarity Measure” that determines the probability that an object is in each image of a video. Accordingly, for the reasons

described above for Zhao, this element was disclosed or at least rendered obvious. EX1004, ¶¶144-145.

f. Element 1/9[e]

Steffens in combination with Zhao disclosed this element or at least rendered it obvious. EX1004, ¶¶146-149. As explained above for element 1/9[d], it would have been obvious to a POSITA to use a probabilistic framework, which was widely known in the prior art. *See* EX1004, ¶146. Further, 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. Indeed, the use of a “threshold” was widely known. *See* EX1004, ¶147.

Moreover, as explained above, a POSITA would have been motivated to look for a probabilistic framework, such as Zhao, to combine with the teachings of Steffens. *See, supra*, Section X.B.1. (Motivation to Combine). Like Steffens, Zhao uses a similarity measure for the identification of objects in video data. As explained in Ground I, this includes the use of a “Bayesian Similarity Measure” that determines the probability that an object is in each image of a video. This further includes comparing the probabilities for each image to a threshold for detecting an object. *See* Ground I limitation 1/9[e]. Accordingly, for the reasons described above for Zhao, this element was disclosed or at least rendered obvious.

3. Claims 2/10

Steffens in combination with Zhao disclosed or at least rendered obvious Claims 2/10. EX1004, ¶¶150-152. As explained above for Ground I, Zhao disclosed or at least rendered obvious Claims 2/10. *See, supra*, Ground I, Claims 2/10. That analysis is incorporated by reference.

Further, as described above, Steffens discloses a “spatio-temporal model” for an object to be detected. Steffens further teaches the use of detectors for each of the collection of components. For example, Steffens explains that its process uses a graph matching technique where the graphs include “nodes” that refer to points on the object, which are labeled with “jets.” EX1007, 8:58-59. “To compute the similarity between two graphs, the sum is taken over similarities of corresponding jets between the graphs.” EX1007, 8:65-67. Accordingly, to detect parts of the object (*i.e.*, the collection of components), Steffens matches the object parts to the “jets” of the graphics to detect the degree of similarity. Accordingly, it was known to a POSITA for a spatio-temporal model for an object to be detected to be comprised of “detectors” where each of the collection of components (*i.e.*, parts of the object) is matched with the “jets.” EX1004, ¶151. As such, Claims 2/10 were disclosed or at least rendered obvious. *Id.*, ¶150-152.

4. Claims 3/11

Steffens in combination with Zhao disclosed or at least rendered obvious Claims 3/11. As explained above for Ground I, Zhao disclosed or at least rendered Claims 3/11 obvious. *See, supra*, Ground I, Claims 3/11. That analysis is incorporated by reference.

The requirement of Claims 3/11 “wherein the spatio-temporal model is a graphical model comprising nodes corresponding to each of the collection of components and to the object” matches the Patent Owner’s purported construction of the term “spatio-temporal model.” Accordingly, Claims 3/11 were disclosed or at least rendered obvious for the same reasons discussed for element 1/9[a] and that purported construction. EX1004, ¶154. Indeed, Steffens describes a graphical model (*i.e.*, a bunch graph) comprised of nodes (*i.e.*, the nodes with jets) corresponding to each of the collection of components and to the object (*i.e.*, a coarse-to-fine approach where the nodes correspond to more or fewer object features for computational efficiency). EX1004, ¶154; *see supra*, element 1/9[a].

5. Claims 4/12

Steffens in combination with Zhao disclosed or at least rendered obvious Claims 4/12. As explained above for Ground I, Zhao rendered Claims 4/12 obvious. *See, supra*, Ground I, Claims 4/12. That analysis is incorporated by reference.

As described above in element 1/9[d], it was known to a POSITA to determine the probability that an object is in each image. Further, Steffens disclosed or at least rendered obvious detecting an object in an 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.” EX1004, ¶156. As Steffens explains, when “the object is in a sequence of images . . . the step of detecting an object further includes ***tracking the object between image frames*** based on a trajectory associated with the object. Also, the step of locating the nodes includes ***tracking the nodes between image frames*** and reinitializing a tracked node if the node's position deviates beyond a predetermined position constraint ***between image frames***.” EX1007, 2:6-13. Accordingly, Steffens claims as part of its process for recognizing objects, “tracking the object between image frames based on a trajectory associated with the object,” “tracking the nodes between image frames,” and “reinitializing a tracked node if the node’s position deviates beyond a predetermined position constraint between image frames.” *See, e.g.*, EX1007, Claims 14, 16, and 17.

By tracking the object and/or its associated nodes “between image frames” as part of its process for recognizing objects, a POSITA would have known to detect the object according to measurements of the object (which is measured as a collection of components, as discussed for element 1/9[a]) from a “prior image” and a “later image” relative to the current image. Indeed, a POSITA would have known

that tracking an object and/or its associated nodes “between frames” would include doing so according to a “prior image” and a “later image” relative to the current image. EX1004, ¶¶157-158. Notably, these steps are part of Steffens’ claimed “process for recognizing objects in a sequence of image frames,” and the existence of a “sequence of image frames” further indicates the obviousness of using measurements determined from image frames anywhere in the sequence, including both prior and later frames relatively to a current frame. *Id.*; EX1007, Claims 13-14, 16-17.

Accordingly, Claims 4/12 were disclosed or at least rendered obvious. EX1004, ¶¶155-159.

6. Claims 5/13

Steffens in combination with Zhao disclosed or at least rendered obvious Claims 5/13. As explained above for Ground I, Zhao disclosed or at least rendered Claims 5/13 obvious. *See, supra*, Ground I, Claims 5/13. That analysis is incorporated by reference.

As described above for element 1/9[a], Steffens describes a spatio-temporal model for an object to be detected. Further, when “the object is in a sequence of images . . . the step of detecting an object further includes *tracking the object between image frames* based on a trajectory associated with the object. Also, the step of locating the nodes includes *tracking the nodes between image frames* and

reinitializing a tracked node if the node's position deviates beyond a predetermined position constraint *between image frames*.” EX1007, 2:6-13. Accordingly, Steffens claims as part of its process for recognizing objects, “tracking the object between image frames based on a trajectory associated with the object,” “tracking the nodes between image frames,” and “reinitializing a tracked node if the node's position deviates beyond a predetermined position constraint between image frames.” *See, e.g.*, EX1007, Claims 14, 16, and 17.

It was known to a POSITA to use a “temporal window” defining a plurality of images in which measurements of components detected are passed to components detected in the current image, as a POSITA would need to define a window of “image frames” for tracking the object and/or its associated nodes “between image frames.” EX1004, ¶162. For example, Steffens describes a temporal window in an embodiment that includes two consecutive image frames, where “[t]he similarity (or confidence) value S may be calculated to indicate how well the two image regions in the two image frames correspond to each other...” EX1007, 13:22-25. Notably, Steffens claims a “process for recognizing objects in a sequence of image frames,” and the existence of a “sequence of image frames” illustrates the obviousness of providing the claimed “temporal window,” as the sequence is itself a temporal window and any subset of frames within that sequence could also be a temporal window. *Id.*, Claims 13-14, 16-17.

As such, Claims 5/13 were disclosed or at least rendered obvious. EX1004, ¶¶160-163.

7. Claims 6/14

Steffens in combination with Zhao disclosed or at least rendered obvious Claims 6/14. As explained above for Ground I, Zhao disclosed or at least rendered obvious Claims 6/14. *See, supra*, Ground I, Claims 6/14. That analysis is incorporated by reference.

As described above in element 1/9[d], Steffens in combination with Zhao describes determining the probability that an object is in each image. Further, Steffens states that its matching technique, includes for example, “provid[ing] more accurate information about the position and size of the face.” EX1007, 7:18-20. Zhao also states, “[t]he above probability distributions provide metrics to evaluate the *shape, size relationship, and configuration* similarities between the detected contour and the human body model.” EX1006, 45. By providing information on position and size (*e.g.*, of the face as in Steffens) or the shape, size, relationship, and configuration (*e.g.*, of the detected body as in Zhao), a POSITA would have known to determine the probability for a “position and size of the object” as claimed in the ’980 Patent. EX1004, ¶164. Accordingly, Claims 6/14 were disclosed or at least rendered obvious. *Id.*

8. Claims 7/15

Steffens in combination with Zhao disclosed or at least rendered obvious Claims 7/15.

As described above for element 1/9[e], Zhao uses a threshold as part of its probabilistic framework for detecting an object. Zhao additionally explains that this threshold can be selected empirically based on an “ROC curve” and the rate of false detections “generated by testing the RCR algorithm on 20 sequences (100 frames/sequences) of urban street scenes,” as explained above for Ground I. *See, supra*, Ground I, Claims 7/15. Claims 7/15 were disclosed or at least rendered obvious here for the same reasons.

9. Claims 8/16

Steffens in combination with Zhao rendered obvious Claims 8/16.

As described for Claims 1/9 element [d-e], using a probabilistic framework for a spatio-temporal model was disclosed or at least rendered obvious. Further, as described above in Ground I, the joint probability used in Zhao, and known to a POSITA is easily representable in the form of the joint probability distribution formulated by Claims 8/16. *See supra*, Ground I, Claims 8/16. Claims 8/16 would have been obvious here for the same reasons.

C. Ground III: Ozer In View Of Zhao Renders Claims 1-16 Obvious

1. Motivation to Combine

Ozer teaches a method for object detection in video sequences. EX1008, Abstract (“the methodology is able to identify object parts and to decide on the presence of the object of interest (human, bag, dog, etc.) in video sequences”). Zhao similarly concerns object detection in the field of computer vision. EX1006, 17 (“This dissertation presents an integrated human shape modeling, detection, and body part localization vision system.”); 97 (“This thesis has addressed several computer vision problems...”). Because Ozer and Zhao each concern object detection, a POSITA would have naturally looked to and combined their complementary teachings. EX1004, ¶170. Indeed, the similarities between Ozer and Zhao are substantial, such that a POSITA would have had reasonable expectation of success in combining Ozer and Zhao, as they both describe similar methods and algorithms for object detection, including using spatio-temporal models for detecting humans and classifying objects over time. *Id.*

During prosecution, the examiner found that there was no motivation or suggestion to combine Ozer with the Cham prior art reference “because the object detection algorithms in [the] two references are vastly different.” EX1002, 163. While Ozer described a graph matching algorithm that used a spatio-temporal model, Cham described a “spatial registration framework” and a kinematic model where no

temporal information was incorporated into the object detection algorithm. *Id.* (citing Cham).

But unlike Cham, Ozer and Zhao describe similar methods and algorithms for object detection, so a POSITA would have been motivated to combine their teachings. EX1004, ¶172. Like Zhao, Ozer compares extracted portions of an image with a set of stored models, which are expressed as graphical models. *Id.*; EX1008, 7:8-31. Both references also utilize a spatio-temporal model for an object to be detected and are not limited to the “kinematic model” described in Cham. EX1004, ¶172.

Further, while Ozer describes, for example, “matching the relational graphs of objects with [a] reference model” (EX1008, 4:24-25), Ozer does not precisely specify which technique to use to identify an object, just that the graph matching algorithm looks for “meaningful correspondence” between the object parts and the reference model. *Id.*, 7:8-31. As a POSITA would have known, finding “meaningful correspondence” through graph matching may include using probabilistic frameworks to perform similarity analyses. EX1004, ¶173. One such framework is expressly described in Zhao as a probabilistic “Bayesian Similarity Measure,” which “combines the local shape and global relationship constraints into a single equation to evaluate the degree of resemblance.” EX1006, 25. Like Ozer’s search for “meaningful correspondence,” Zhao describes a method where an object is identified

based on a similarity between an image and a corresponding model (*i.e.*, “the resemblance between a contour [of an object in an image] and a model”). EX1006, 54. Accordingly, a POSITA would have been motivated to combine Ozer with Zhao’s probabilistic Bayesian similarity measure and had a reasonable expectation of success in doing so, since Zhao is elaborating on and improving a process that Ozer already incorporates into its framework for object detection. EX1004, ¶173.

2. Claims 1/9

a. Element 1/9[preamble]

This element was disclosed or at least rendered obvious over Ozer in view of Zhao. EX1004, ¶¶175-176. As an initial matter, this element was disclosed or at least rendered obvious by Zhao, as described above for Ground I and incorporated by reference.

Further, Ozer describes a method for object detection in video sequences. EX1008, Abstract. As a system and method relating to object detection for video images, a POSITA would have understood Ozer to describe a computer implemented method for object detection. EX1004, ¶176. A POSITA would have understood this to further include “a computer readable medium embodying a computer program” to perform that method. *Id.* For example, Ozer describes its techniques being implemented on a “computer” that “may keep a database of graphs, HMM models

[hidden Markov models], and other information used during video analysis.”

EX1008, 6:30-32; *see also Id.* 3:28-33, 5:4-27.

b. Element 1/9[a]

This element was disclosed or at least rendered obvious over Ozer in view of Zhao. As an initial matter, this element was disclosed or at least rendered obvious by Zhao, as described above for Ground I and incorporated by reference.

Further, this element was disclosed or at least rendered obvious in view of Ozer. Ozer describes providing a “spatio-temporal model” for an object to be detected. Indeed, Ozer describes a process for “graph matching” where “we compare the object model with a set of stored models. Each extracted region modeled with ellipses corresponds to a node in the graphical representation of the object of interest.” EX1008, 7:8-11. Further, Ozer teaches “Classifying Over Multiple Frames” where “[t]he movements of the object parts are described as a spatio-temporal sequence of feature vectors that consist of the direction of the object part movement. 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, 7:33-45, 4:45-47 (describing the use of “Hidden Markov Models, which classifies activity patterns of the body parts in space-time and determines the movements of the object parts. It

also uses a distance classifier to detect specific gestures and activities.”). Accordingly, a POSITA would have found the claimed “spatio-temporal model” disclosed or at least obvious in view of Ozer. EX1004, ¶178.

Further, to the extent the term “spatio-temporal model” is construed as “a graphical model comprising nodes corresponding to each of the collection of components and to the object,” Ozer disclosed or at least rendered this element obvious. EX1004, ¶179. As part of Ozer’s method for object detection, each extracted part in an image “corresponds to a node in the graphical representation of the object of interest.” EX1008, 7:9-11. The “[o]utput of the graph-matching algorithm is the classified object parts.” *Id.*, 7:33-35. “For example, if the object of interest is a human body, the system will first detect different object parts, e.g. hands, head, arms, legs, torso and compare these part attributes with the human model attributes via graph matching.” *Id.*, 4:38-41. Further, “geometrical characteristics of the object regions is used to group object parts” and “[t]he object is decomposed for its presentation as a combination of component shapes,” such that “[t]he result will be unaffected by a partial occlusion of the object.” *Id.*, 4:1-6. Accordingly, a POSITA would have understood Ozer to describe providing a graphical model comprising nodes [*i.e.*, the graphical representation of the object of interest] corresponding to each of the collection of components [*i.e.*, the classified object

parts] and to the object [*i.e.*, the grouped object parts, forming the object]. EX1004, ¶179.

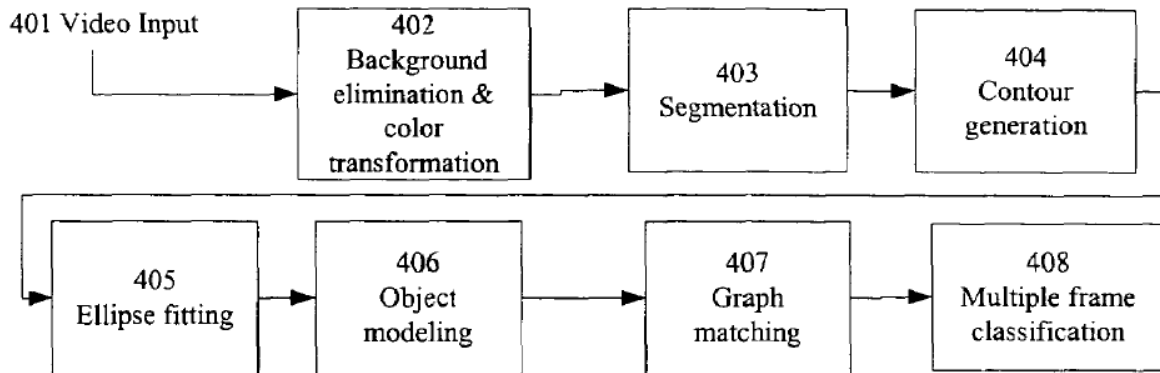
As such, this element was disclosed or at least rendered obvious. EX1004, ¶¶177-180.

c. Element 1/9[b]

This element was disclosed or at least rendered obvious over Ozer in view of Zhao. As an initial matter, this element was disclosed or least rendered obvious by Zhao, as described above for Ground I and incorporated by reference.

Ozer further disclosed or at least rendered obvious this element. Ozer states that its “methodology is able to identify object parts and to decide on the presence of the object of interest (human, bag, dog, etc.) in video sequences.” EX1008, Abstract. Accordingly, Ozer’s system provides “401 Video Input” to perform object detection, as illustrated in Figure 4:

Figure 4



EX1008, 6. Moreover, as Ozer defines, a “video sequence” includes “[a] sequence of images” and “can be any video recording or media (e.g. an MPEG-1 file, a video tape, a video disc, a DVD, etc.).” EX1008, 3:16-19. Accordingly, a POSITA would have found this element disclosed or at least rendered obvious. EX1004, ¶¶181-183.

d. Element 1/9[c]

This element was disclosed or at least rendered obvious over Ozer in view of Zhao. As an initial matter, this element was disclosed or least rendered obvious by Zhao, as described above for Ground I and incorporated by reference.

Ozer further disclosed or at least rendered this element obvious. Ozer explains that “[e]ach extracted region modeled with ellipses corresponds to a node in the graphical representation of the object of interest.” EX1008, 7:9-11. The graph matching algorithm looks for “meaningful correspondence” where “[e]ach object

part and meaningful combinations represent a class w where the combination of binary and unary features are represented by a feature vector X and computed off-line.” *Id.*, 7:11-19. As such, “[t]he computations needed for each node matching are then a function of the feature size and the previously matched nodes of the branch under consideration.” *Id.*, 7:25-28. Accordingly, this element was disclosed or at least rendered obvious to a POSITA. EX1004, ¶¶187-185.

e. Element 1/9[d]

This element was disclosed or at least rendered obvious over Ozer in view of Zhao. As explained in Section X.C.1 (Motivation to Combine), a POSITA would have looked to combine the teachings of Ozer with Zhao. Further, as described above for Ground I, this element was disclosed or at least rendered obvious in view of Zhao. Accordingly, the analysis under Ground I is incorporated by reference and this element was disclosed or at least rendered obvious for the same reasons. *See, supra*, Ground I element 1/9[d].

f. Element 1/9[e]

This element was disclosed or at least rendered obvious over Ozer in view of Zhao. As explained in Section X.C.1 (Motivation to Combine), a POSITA would have looked to combine the teachings of Ozer with Zhao. Further, as described above for Ground I, this element was disclosed or least rendered obvious by Zhao. Accordingly, the analysis under Ground I is incorporated by reference and this

element was disclosed or at least obvious for the same reasons. *See, supra*, Ground I element 1/9[e].

3. Claims 2/10

Ozer in combination with Zhao disclosed or at least rendered obvious Claims 2/10. EX1004, ¶¶188-189. As explained above for Ground I, Zhao disclosed or at least rendered obvious Claims 2/10. *See, supra*, Ground I, Claims 2/10. That analysis is incorporated by reference.

Further, a POSITA would have found it disclosed or at least obvious for a spatio-temporal model for an object to be detected to be comprised of “detectors” for each of the collection of components in view of Ozer. EX1004, ¶189. Ozer explains that each extracted part of an image “corresponds to a node in the graphical representation of the object of interest” (EX1008, 7:9-11) and “[o]utput of the graph-matching algorithm is the classified object parts.” EX1008, 7:34-35. Accordingly, it was known to a POSITA for a spatio-temporal model for an object to be detected to be comprised of “detectors” where each of the collection of components are matched to nodes in a graphical representation of the object of interest. EX1004, ¶189. As such, Claims 2/10 were disclosed or at least rendered obvious. *Id.*

4. Claims 3/11

Ozer in combination with Zhao disclosed or at least rendered obvious Claims 3/11. EX1004, ¶¶190-191. As explained above for Ground I, Zhao disclosed or at

least rendered obvious Claims 3/11. *See, supra*, Ground I, Claims 3/11. That analysis is incorporated by reference.

Further, the additional requirement of Claims 3/11 “wherein the spatio-temporal model is a graphical model comprising nodes corresponding to each of the collection of components and to the object” matches the Patent Owner’s purported construction of the term “spatio-temporal model.” Accordingly, Claims 3/11 were disclosed or at least obvious for the same reasons discussed for element 1/9[a] and that purported construction. EX1004, ¶191. As described above, a POSITA would have found the purported spatio-temporal model disclosed or at least obvious in view of Zhao, as it describes a graphical model where “ribbons” serve as “nodes” that correspond to each of the collection of components (*i.e.*, the component body parts) and the object (*i.e.*, the merged parts that form a human). EX1004, ¶191. Similarly, a POSITA would have understood Ozer to teach the use of a graphical model comprising nodes [*i.e.*, the graphical representation of the object of interest] corresponding to each of the collection of components [*i.e.*, the classified object parts] and to the object [*i.e.*, the grouped object parts, forming the object]. *Id.* As such, Claims 3/11 were disclosed or at least obvious. *Id.*

5. Claims 4/12

Ozer in combination with Zhao disclosed or at least rendered obvious Claims 4/12. EX1004, ¶¶192-195. As explained above for Ground I, Zhao rendered Claims

4/12 obvious. *See, supra*, Ground I, Claims 4/12. That analysis is incorporated by reference.

A POSITA would have found Claims 4/12 disclosed or at least obvious in further view of Ozer. Ozer teaches “Classifying Over Multiple Frames” where “[t]he movements of the object parts are described as a spatio-temporal sequence of feature vectors that consist of the direction of the object part movement. 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, 7:33-45. As such, it would have been known to a POSITA 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, ¶¶193-194.

6. Claims 5/13

Ozer in combination with Zhao disclosed or at least rendered obvious Claims 5/13. EX1004, ¶¶196-198. As explained above for Ground I, Zhao disclosed or at least rendered Claims 5/13 obvious. *See, supra*, Ground I, Claims 5/13. That analysis is incorporated by reference.

A POSITA would have found Claims 4/12 disclosed or at least obvious in further view of Ozer. Ozer teaches “Classifying Over Multiple Frames” where “[t]he

movements of the object parts are described as a spatio-temporal sequence of feature vectors that consist of the direction of the object part movement. 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, 7:33-45. By measuring the object parts for a “number of frames,” it would have been known to a POSITA to define a “temporal window” in which measurements of components are passed to components detected in the current image. EX1004, ¶197.

7. Claims 6/14

Claims 6/14 were disclosed or at least rendered obvious over Ozer in view of Zhao. As explained in Section X.C.1 (Motivation to Combine), a POSITA would have looked to combine the teachings of Ozer with Zhao. Further, as described above for Ground I, Claims 6/14 were disclosed or least rendered obvious by Zhao. Accordingly, the analysis under Ground I is incorporated by reference and Claims 6/14 were disclosed or at least rendered obvious for the same reasons here. *See, supra*, Ground I, Claims 6/14.

8. Claims 7/15

Claims 7/15 were disclosed or at least rendered obvious over Ozer in view of Zhao. As explained in Section X.C.1 (Motivation to Combine), a POSITA would

have looked to combine the teachings of Ozer with Zhao. Further, as described above for Ground I, Claims 7/15 were disclosed or least rendered obvious by Zhao. Accordingly, the analysis under Ground I is incorporated by reference and Claims 7/15 were disclosed or a least rendered obvious for the same reasons here. *See, supra*, Ground I, Claims 7/15.

9. Claims 8/16

Claims 8/16 would have been obvious over Ozer in view of Zhao. As explained in Section X.C.1 (Motivation to Combine), a POSITA would have looked to combine the teachings of Ozer with Zhao. Further, as described above for Ground I, Claims 8/16 would have been obvious in view of the teachings of Zhao. Accordingly, the analysis under Ground I is incorporated by reference and Claims 8/16 are rendered obvious for the same reasons here. *See, supra*, Ground I, Claims 8/16.

D. Ground IV: TLP Renders Claims 1-16 Obvious

1. Claims 1/9

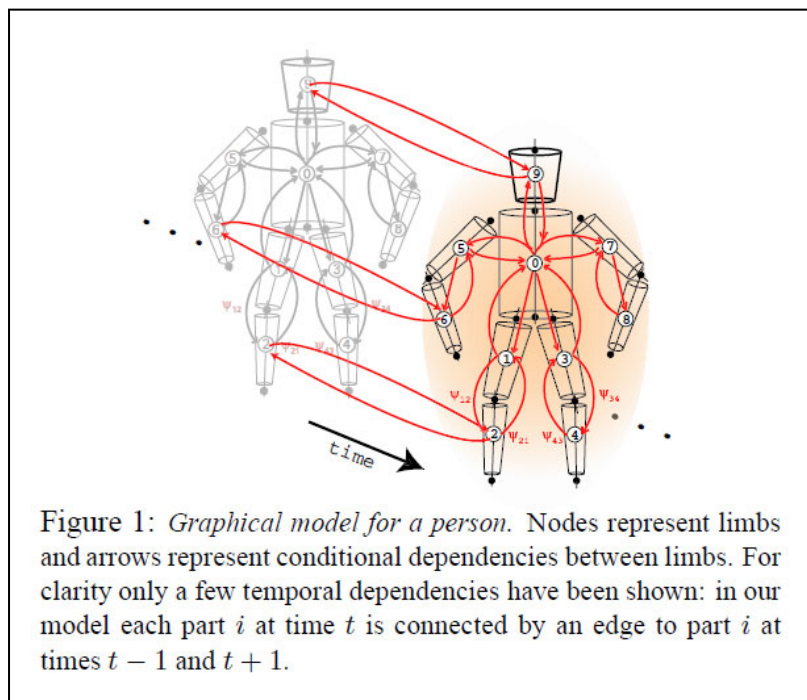
a. Element 1/9[preamble]

TLP disclosed this element or at least rendered it obvious. EX1004, ¶202. Specifically, TLP describes “a probabilistic method for *fully automatic* 3D human detection and tracking.” EX1009, 6. Accordingly, a computer implemented method for object detection would have been obvious to a POSITA. EX1004, ¶202. A

POSITA would have understood this to further include “a computer readable medium embodying a computer program” to perform that method. *Id.*

b. Element 1/9[a]

TLP disclosed this element or at least rendered it obvious. EX1004, ¶¶203-205. As part of its method for “human detection and tracking,” TLP provides a model for an object to be detected (*e.g.*, a human). TLP explains that “the body is represented by a graphical model in which each graph node corresponds to a body part (upper leg, torso, etc.). Each part has an associated configuration vector defining the part’s position and orientation in 3-space. ... Edges in the graphical model correspond to position and angle relationships between adjacent body parts in space and time, as illustrated in Figure 1.” EX1009, 2.



EX1009, 1. By incorporating spatial and temporal elements into a graphical model for an object to be detected, such as a person, TLP teaches the claimed “spatio-temporal.” EX1004, ¶203.

Further, even to the extent the term “spatio-temporal model” is construed as “a graphical model comprising nodes corresponding to each of the collection of components and to the object,” TLP still disclosed or at least rendered this element obvious. EX1004, ¶204. TLP’s model for an object to be detected is explicitly expressed as “a graphical model in which each graph node corresponds to a body part (upper leg, torso, etc.).” EX1009 at 2. Further, “[e]ach part has an associated configuration vector defining the part’s position and orientation in 3-space. Placing each part in a global coordinate frame enables the part detectors to operate independently while the full body is assembled by inference over the graphical model.” EX1009, 2. A POSITA would have found the claimed spatio-temporal model to be disclosed or at least obvious in view of TLP, as the nodes correspond to each of the collection of components and the object. EX1004, ¶204.

Accordingly, TLP disclosed or at least rendered this element obvious. EX1004, ¶¶203-205.

c. Element 1/9[b]

TLP disclosed or at least rendered this element obvious. EX1004, ¶206. TLP discloses a method for object detection and tracking for video. Accordingly,

“providing a video comprising a plurality of images including the object” would have been known to a POSITA because to detect an object in a video, providing the video itself would be required. *See* EX1004, ¶206; EX1009, 1 (“We illustrate the method by automatically tracking a walking person in video imagery using four calibrated cameras.”).

d. Element 1/9[c]

TLP disclosed or at least rendered this element obvious. EX1004, ¶¶207-211. TLP explains “our model of the body is a collection of loosely-connected limbs.” EX1009, 1. To perform object detection, this body model is combined “with a probabilistic image likelihood model” and “[i]nferring the body pose in our framework is defined as estimating belief in the graphical model.” EX1009, 4. “We define $\phi_i(X_i)$ to be the likelihood of observing the image measurements conditioned on the pose of limb i . Ideally this model would be robust to partial occlusions, the variability of image statistics across different input sequences, and variability among subjects. To that end, we combine a variety of cues including multi-scale edge and ridge filters following [20 [EX 1041]].” *Id.* at 4. Accordingly, measuring the object as a collection of components in each image would have been known to a POSITA, as the model itself is represented as “a collection of loosely-connected limbs” —*i.e.*, a collection of components and a POSITA would have understood the collection of

loosely-connected limbs to be equivalent to a collection of components. EX1009, 1; EX1004, ¶207.

Notably, to describe the parts in the graphical model, “[e]ach part has an associated configuration vector defining the part’s position and orientation in 3-space,” as shown in Figure 2:

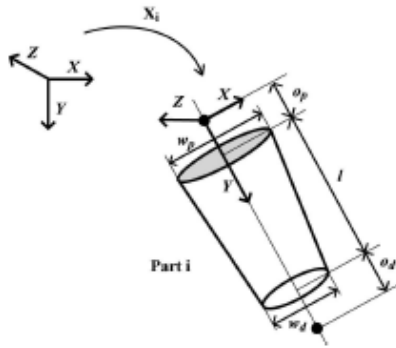
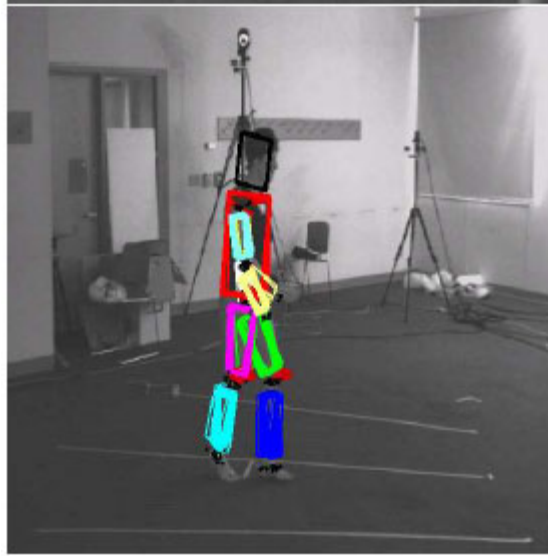


Figure 2: *Parameterization of part i.*

EX1009, 2, 3. “The estimated parameters $\mathbf{X}_i^T = (\mathbf{x}_i^T, \Theta_i^T)$ represent the configuration of the part i in a global coordinate frame where $\mathbf{x}_i \in \mathbb{R}^3$ and $\Theta_i \in \text{SO}(3)$ are the 3D position of the proximal joint and the angular orientation of the part respectively.” EX1009, 2-3. This matches the process described in the specification of the ’980 Patent, which explains that nodes representing an object’s components “have an associated state vector $\mathbf{X}^T=(x,y,s)$ defining the component’s real-valued position and scale within an image.” EX1001, 3:50-52.

Figure 6(c) illustrates, for example, the object being measured as a collection of components using the system defined in TLP:



c

EX1009, 7, Figure 6(c) (illustrating “Samples from the belief distribution after 30 iterations of belief propagation”). Accordingly, this element was disclosed or at least obvious over TLP. EX1004, ¶¶207-211.

e. Element 1/9[d]

TLP disclosed this element or at least rendered it obvious. EX1004, ¶¶212-213. TLP “present[s] a *probabilistic* method for fully automatic 3D human detection and tracking.” EX1009, 6. Accordingly, the inference algorithm “combines the body model described above with a probabilistic image likelihood model” and defines “ $\phi_i(\mathbf{X}_i)$ to be the likelihood of observing the image measurements conditioned on the pose of limb i .” EX1009, 4. To “be robust to partial occlusions, the variability of image statistics across different input sequences, and variability among subjects,” TLP incorporates the use of a “Gibbs [likelihood] model”:

$$p(\mathbf{f} \mid \mathbf{X}_i) \propto \exp \left(- \sum_i \langle \lambda^{(i)}, \xi^{(i)}(\mathbf{f}, \mathbf{X}_i) \rangle \right)$$

EX1009, 4. Accordingly, a POSITA would have understood that the probabilistic method described is equivalent to determining a probability an object is in each image, and this element was disclosed or at least obvious in view of TLP. EX1004, ¶¶212-213.

f. Element 1/9[e]

TLP disclosed this element or at least rendered it obvious. EX1004, ¶¶214-215. As explained above in Section X.D.1 (elements 1/9[d]-[e]), TLP “present[s] a ***probabilistic*** method for fully automatic 3D human detection and tracking” and determines the likelihood (*i.e.*, probability) that an object is in each image. EX1009, 6. Accordingly, it would have been known to a POSITA to further compare the probabilities to a “threshold” to detect the object. Indeed, TLP describes the use of an “image likelihood model” and detecting the object is “defined as estimating belief in the graphical model.” EX1009, 4. Indeed, to infer the belief in a model, it would have been known to a POSITA to compare a determined probability against a threshold. *See* EX1004, ¶214.

2. Claims 2/10

TLP disclosed or at least rendered obvious Claims 2/10. EX1004, ¶¶216-218.

Further, as described above for element 1/9[a], TLP discloses a “spatio-temporal model” for an object to be detected. TLP further teaches the use of detectors for each of the collection of components. For example, according to TLP, “[w]e adopt simple limb and head detectors to provide ‘bottom-up’ information that is incorporated into the inference process at every time-step.” EX1009, 1. Indeed, TLP acknowledges that “[m]any good head detectors exist and limb detectors have been used for some time.” EX1009, 1 (citing prior art). Accordingly, Claims 2/10 were disclosed or at least rendered obvious. EX1004, ¶¶216-218.

3. Claims 3/11

TLP disclosed or at least rendered obvious Claims 3/11. EX1004, ¶¶219-220.

Further, the additional requirement of Claims 3/11 “wherein the spatio-temporal model is a graphical model comprising nodes corresponding to each of the collection of components and to the object” matches the Patent Owner’s purported construction of the term “spatio-temporal model.” Accordingly, Claims 3/11 was disclosed or at least rendered obvious for the same reasons discussed for element 1/9[a] and that purported construction. EX1004, ¶220.

4. Claims 4/12

TLP disclosed or at least rendered obvious Claims 4/12. EX1004, ¶¶221-222.

As described for elements 1/9[d]-[e] above, TLP presents a probabilistic method that determines the probability that an object is in each image based on identified body parts, which includes temporal dependencies. As TLP explains, “in our model, each part i at time t is connected by an edge to part i at times $t - 1$ and $t + 1$.” EX1009, 1 (Figure 1). Accordingly, TLP uses “probabilistic models for the temporal evolution of each limb (forward and backward in time). Human pose and motion estimation is then solved with non-parametric belief propagation using a variation of particle filtering that can be applied over a general loopy graph.” EX1009, 1. Accordingly, TLP disclosed or at least rendered obvious the step of detecting an image according to measurements of the object as a collection of components determined from a prior image (i.e., $t - 1$) and a later image (i.e., $t + 1$), relative to the current image (i.e., t). EX1004, ¶222.

5. Claims 5/13

TLP disclosed or at least rendered obvious Claims 5/13. EX1004, ¶¶223-224.

As described above for element 1/9[a], TLP describes a spatio-temporal model for an object to be detected. As TLP explains, “in our model, each part i at time t is connected by an edge to part i at times $t - 1$ and $t + 1$.” EX1009, 1 (Figure 1). Accordingly, TLP uses “probabilistic models for the temporal evolution of each limb (forward and backward in time). Human pose and motion estimation is then solved with non-parametric belief propagation using a variation of particle filtering

that can be applied over a general loopy graph.” EX1009, 1. TLP further discloses using “a **3-frame windowed** smoothing algorithm for our tracking results where the estimates at time t are based on observations at times $(t - 1, t + 1)$.” EX1009, 5. Accordingly, a POSITA would have understood that TLP describes providing a “temporal window” defining a “plurality of images” in which measurements of components detected are passed to components detected in the current image. EX1004, ¶224.

6. Claims 6/14

TLP disclosed or at least rendered obvious Claims 6/14. EX1004, ¶¶225-227.

As described above, TLP describes determining the probability that an object is in each image. *See* element 1/9[d]. As part of that process, TLP explains “the body is represented by a graphical model in which each graph node corresponds to a body part (upper leg, torso, etc.). Each part has an associated configuration vector defining the part’s position and orientation in 3-space.” EX1009, 2. By determining the part’s position and orientation in 3-space, a POSITA would have known to determine the probability for a “position” of the object in each image.

Further, “[e]ach part/limb is modeled by a tapered cylinder having 5 fixed and 6 estimated parameters,” which include “the part length, width at the proximal and distal ends and the offset of the proximal and distal joints along the axis of the limb...” EX1009, 2. By determining the part’s parameters, such as length and width,

a POSITA would have known to determine the probability for a “size” of the object in each image.

7. Claims 7/15

TLP disclosed or at least rendered obvious Claims 7/15. EX1004, ¶¶228-229.

As described above for element 1/9[e], it was known to a POSITA to compare a determined probability against a threshold. *See* EX1004, ¶229. It would have been further known that this threshold is determined empirically, as opposed to arbitrarily chosen. *Id.* Indeed, TLP explains that “[c]onditional probabilities relating the 3D pose of connected limbs are learned from motion captured training data [i.e., empirically].” EX1009, 1. As a POSITA would have understood, determining a threshold “empirically” is done via observation or experimentation. EX1004, ¶229. By learning the conditional probabilities through training data, a POSITA would have known to determine a threshold “empirically” in view of TLP. *Id.* Accordingly, Claims 7/15 were disclosed or at least obvious. *Id.*

8. Claims 8/16

TLP rendered Claims 8/16 obvious. EX1004, ¶¶230-231.

As described for Claim 1/9, TLP presents a probabilistic method for object detection. *See* elements 1/9[d]-[e]. The joint probability distribution formulated in Claims 8/16 would have been known to a POSITA and is the very same pairwise Markov field representation of a spatio-temporal graphical model, which is cited in

TLP. *See* EX1009, 4 (“we use a form of non-parametric belief propagation [13 [citing EX1019], 25 [citing EX1012]].”); *see also* EX1004, ¶131 (“The equation represents a typical factorized form of the joint probability distribution $P(\dots)$ for a graphical model G of an object O with state X_i^O at time frame i of the video and $\{k = 1..N\}$ components state X_i^{C-k} ”). This form of the joint probability would have been known to a POSITA as a pairwise Markov field representation of a spatio-temporal graphical model. *Id.* (citing EX1012, §2.1).

XI. SECONDARY CONSIDERATIONS

Petitioners are unaware of any evidence of secondary considerations that would support a finding of non-obviousness. *See* EX1004 at ¶232. The asserted prior art demonstrates there is no evidence of failure by others and that the features recited in the Challenged Claims were readily available in the prior art. In any event, no secondary considerations could overcome the strong *prima facie* case of obviousness shown in this Petition. *See Wyers v. Master Lock Co.*, 616 F.3d 1231, 7276 (Fed. Cir. 2010). Should Patent Owner attempt to meet its burden to show secondary considerations, Petitioners reserve the right to respond.

XII. CONCLUSION

For the foregoing reasons, Petitioners request Trial be instituted and Claims 1-16 be canceled.

Petition for *Inter Partes* Review of U.S. Patent No. 7,436,980

Date: May 24, 2023

Respectfully submitted,

/s/ W. Todd Baker

W. Todd Baker (No. 45,265)

todd.baker@kirkland.com

Postal and Hand-Delivery Address:

KIRKLAND & ELLIS LLP

1301 Pennsylvania Ave., N.W.

Washington, D.C. 20004

Telephone: (202) 389-5000

Facsimile: (202) 389-5200

Ellisen Shelton Turner (No. 54,503)

ellisen.turner@kirkland.com

Postal and Hand-Delivery Address:

KIRKLAND & ELLIS LLP

2049 Century Park East, 37th Floor

Los Angeles, CA 90067

Telephone: (310) 552-4200

Facsimile: (310) 552-5900

Jonathan D. Brit (No. 75,646)

jonathan.brit@kirkland.com

Postal and Hand-Delivery Address:

KIRKLAND & ELLIS LLP

601 Lexington Avenue

New York, NY 10022

Telephone: (212) 390-4209

Facsimile: (212) 446-4900

Attorneys for Petitioners

CERTIFICATE OF COMPLIANCE

This Petition complies with the type-volume limitations as mandated in 37 C.F.R. § 42.24, totaling 13,998 words. Counsel has relied upon the word count feature provided by Microsoft Word.

/W. Todd Baker/

W. Todd Baker (No. 45,265)

CERTIFICATE OF SERVICE

The undersigned hereby certifies that a copy of the foregoing document with accompanying Exhibits 1001 through 1043 as listed in Petitioner's Exhibit List on page iv of this document, was served on May 24, 2023, via overnight delivery, upon agreement under 37 CFR § 42.105, directed to the attorney/agent of record for the patent as identified on USPTO Patent Center and the address on record for the current assignee as follows:

SIEMENS CORPORATION
IP Dept - Mail Code INT-244
3850 Quadrangle Blvd
Orlando, FL 32817

VL COLLECTIVE IP LLC
2303 Saint Francis Drive
Palo Alto, CA 94303

A courtesy copy was also served by electronic mail on PO's counsel of record in the following related matter:

VideoLabs, et al v. Meta Platforms, Inc., et al, Case No. 1:22-cv-00680 (D. Del.).

Petition for *Inter Partes* Review of U.S. Patent No. 7,436,980

Brian E. Farnan
Michael J. Farnan
919 N. Market St., 12th Floor
Wilmington, DE 19801
Telephone: (302) 777-0300
Fax: (302) 777-0301
bfarnan@farnanlaw.com
mfarnan@farnanlaw.com

Christine E. Lehman
Brian Baran
REICHMAN JORGENSEN
LEHMAN &
FELDBERG LLP
1909 K St, NW, Suite 800
Washington, DC 20006

Jaime F. Cardenas-Navia
Michael Matulewicz-Crowley
REICHMAN JORGENSEN LEHMAN &
FELDBERG LLP
400 Madison Ave, Suite 14D
New York, NY 10017

Taylor N. Mauze
Navid Bayar
REICHMAN JORGENSEN LEHMAN
& FELDBERG LLP
7500 Rialto Blvd., Ste. 1-250
Austin, Texas 78735

/W. Todd Baker/

W. Todd Baker (No. 45,265)

APPENDIX A

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

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

	components detected therein are passed to components detected in the current image.
14	The method of claim 9, wherein determining the probability that the object is in each image comprises determining the probability for a position and a size of the object in each image.
15	The method of claim 9, wherein the threshold is provided for the object to be detected, wherein the threshold is determined empirically.
16	The method of claim 9, wherein a joint probability distribution for the spatio-temporal model with N components is: $P(X_0^O, X_0^{C_0}, X_0^{C_1}, \dots, X_0^{C_N}, \dots, X_T^O, X_T^{C_0}, X_T^{C_1}, \dots, X_T^{C_N}) =$ $\frac{1}{Z} \prod_{ij} \psi_{ij}(X_i^O, X_j^O)$ $\prod_{ik} \psi_{ik}(X_i^O, X_i^{C_k}) \prod_{ikl} \psi_{kl}(X_i^{C_k}, X_i^{C_l}) \prod_i \phi_i(X_i^O) \prod_{ik} \phi_i(X_i^{C_k}).$