

Petitioners' Demonstrative Exhibits for Oral Argument

Meta Platforms, Inc., et al. v. VL Collective IP LLC

IPR2023-00924

U.S. Patent No. 7,436,980

Patent Trial and Appeal Board

September 10, 2024

Alexandria, Virginia

U.S. Patent No. 7,436,980

IPR2023-00924

Overview

(12) United States Patent Sigal et al.		(10) Patent No.: US 7,436,980 B2
		(45) Date of Patent: Oct. 14, 2008
(54) GRAPHICAL OBJECT MODELS FOR DETECTION AND TRACKING	7,200,266 B2* 4/2007 Ozer et al. 382/173	
(75) Inventors: Leonid Sigal , Providence, RI (US); Ying Zhu , Monmouth Junction, NJ (US); Dorin Comaniciu , Princeton Jct., NJ (US)	EP 1 217 572 A2 6/2002 FOREIGN PATENT DOCUMENTS	
(73) Assignee: Siemens Corporate Research, Inc. , Princeton, NJ (US)	OTHER PUBLICATIONS	
(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 691 days.	M. Isard, Pumpas: Real-valued graphical models for computer vision, CVPR, vol. 1, pp. 613-620, 2003. K. Okuma, A. Teleghani, N. de Freitas, J. Little and D. Lowe. A boosted particle filter: Multi-target detection and tracking, ECCV, 2004. L. Sigal, S. Bhatia, S. Roth, M. Black and M. Isard. Tracking loose-limbed people, CVPR, 2004. E. Sudderth, A. Ihler, W. Freeman and A. Willsky. Nonparametric belief propagation. MIT Laboratory for Information and Decision Systems Technical Report P-2551 (Oct. 15, 2002). P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features, CVPR, 2001. P. Viola, M. Jones and D. Snow. Detecting pedestrians using patterns of motion and appearance, ICCV, pp. 734-741, 2003.	
(21) Appl. No.: 11/135,210	(Continued)	
(22) Filed: May 23, 2005	Primary Examiner—Tom Y Lu (74) Attorney, Agent, or Firm—Donald B. Paschburg, F. Chau & Associates, LLC	
(65) Prior Publication Data US 2005/0286738 A1 Dec. 29, 2005	(57) ABSTRACT	
Related U.S. Application Data (60) Provisional application No. 60/574,799, filed on May 27, 2004.	A computer implemented method for object detection includes providing a spatio-temporal model for an object to be detected, providing a video including a plurality of images including the object, and measuring the object as a collection of components in each image. The method further includes determining a probability that the object is in each image, and detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.	
(51) Int. Cl. G06K 9/00 (2006.01)		
(52) U.S. Cl. 382/103		
(58) Field of Classification Search 382/103, 382/173		
See application file for complete search history.		
(56) References Cited U.S. PATENT DOCUMENTS 6,597,801 B1 7/2003 Cham et al.		
16 Claims, 5 Drawing Sheets		

501 PROVIDING A SPATIO-TEMPORAL MODEL FOR AN OBJECT

502 PROVIDING A VIDEO COMPRISING IMAGES INCLUDING THE OBJECT

503 MEASURING THE OBJECT AS A COLLECTION OF COMPONENTS IN EACH IMAGE

504 DETERMINING A PROBABILITY THAT THE OBJECT IS IN EACH IMAGE

505 DETECTING THE OBJECT IN AN IMAGE UPON COMPARING THE PROBABILITIES FOR THE IMAGE TO A THRESHOLD FOR DETECTING THE OBJECT

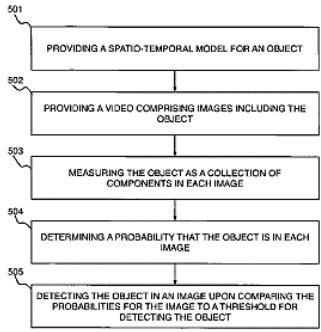
1. Background

2. Claim Construction

- 1/9[a]: spatio-temporal model
- 1/9[c]: measuring the object as a collection of components
- 1/9[e]: comparing the probabilities for each image

3. Grounds

- Claims 1-16: Obvious in view of Zhao
- Claims 1-16: Obvious over Steffens in view of Zhao
- Claims 1-16: Obvious over Ozer in view of Zhao
- Claims 1-16: Obvious in view of TLP

 US007436980B2	
(12) United States Patent Sigal et al.	(10) Patent No.: US 7,436,980 B2 (45) Date of Patent: Oct. 14, 2008
(54) GRAPHICAL OBJECT MODELS FOR DETECTION AND TRACKING	7,200,266 B2* 4/2007 Ozer et al. 382/173
(75) Inventors: Leonid Sigal , Providence, RI (US); Ying Zhu , Monmouth Junction, NJ (US); Dorin Comaniciu , Princeton Jct., NJ (US)	EP 1 217 572 A2 6/2002 FOREIGN PATENT DOCUMENTS
(73) Assignee: Siemens Corporate Research, Inc. , Princeton, NJ (US)	OTHER PUBLICATIONS M. Isard, Pumpas: Real-valued graphical models for computer vision, CVPR, vol. 1, pp. 613-620, 2003. K. Okuma, A. Teleghani, N. de Freitas, J. Little and D. Lowe. A boosted particle filter: Multi-target detection and tracking, ECCV, 2004. L. Sigal, S. Bhatia, S. Roth, M. Black and M. Isard. Tracking loose-limbed people, CVPR, 2004. E. Sudderth, A. Ihler, W. Freeman and A. Willsky. Nonparametric belief propagation, MIT Laboratory for Information and Decision Systems Technical Report P-2551 (Oct. 15, 2002). P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features, CVPR, 2001. P. Viola, M. Jones and D. Snow. Detecting pedestrians using patterns of motion and appearance, ICCV, pp. 734-741, 2003.
(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 691 days.	(Continued)
(21) Appl. No.: 11/135,210	Primary Examiner—Tom Y Lu (74) Attorney, Agent, or Firm—Donald B. Paschburg, F. Chau & Associates, LLC
(22) Filed: May 23, 2005	
(65) Prior Publication Data US 2005/0286738 A1 Dec. 29, 2005	
Related U.S. Application Data	
(60) Provisional application No. 60/574,799, filed on May 27, 2004.	(57) ABSTRACT A computer implemented method for object detection includes providing a spatio-temporal model for an object to be detected, providing a video including a plurality of images including the object, and measuring the object as a collection of components in each image. The method further includes determining a probability that the object is in each image, and detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.
(51) Int. Cl. G06K 9/00 (2006.01)	
(52) U.S. Cl. 382/103	
(58) Field of Classification Search 382/103, 382/173	
See application file for complete search history.	
(56) References Cited U.S. PATENT DOCUMENTS 6,597,801 B1 7/2003 Cham et al.	
16 Claims, 5 Drawing Sheets	
 <pre>graph TD 501[PROVIDING A SPATIO-TEMPORAL MODEL FOR AN OBJECT] --> 502[PROVIDING A VIDEO COMPRISING IMAGES INCLUDING THE OBJECT] 502 --> 503[MEASURING THE OBJECT AS A COLLECTION OF COMPONENTS IN EACH IMAGE] 503 --> 504[DETERMINING A PROBABILITY THAT THE OBJECT IS IN EACH IMAGE] 504 --> 505[DETECTING THE OBJECT IN AN IMAGE UPON COMPARING THE PROBABILITIES FOR THE IMAGE TO A THRESHOLD FOR DETECTING THE OBJECT]</pre>	

1. Background

2. Claim Construction

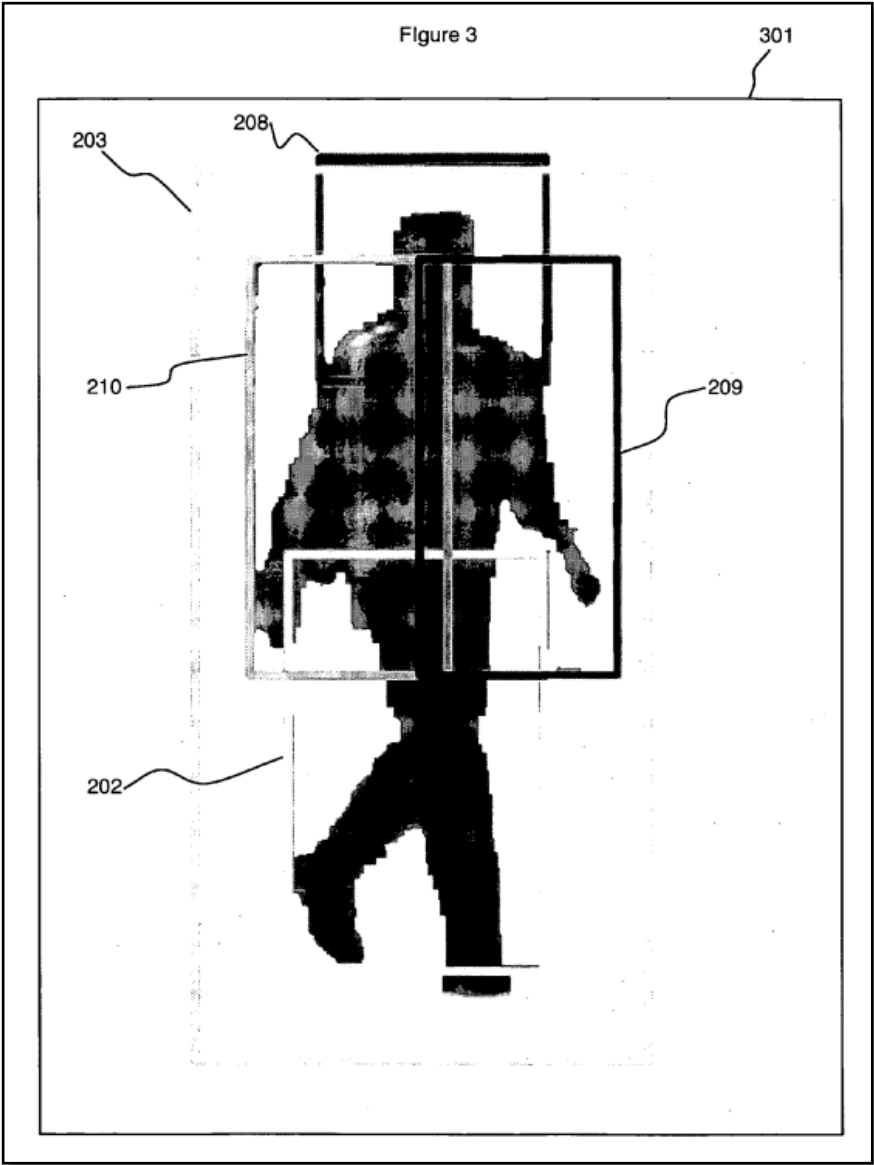
- a. 1/9[a]: spatio-temporal model
- b. 1/9[c]: measuring the object as a collection of components
- c. 1/9[e]: comparing the probabilities for each image

3. Grounds

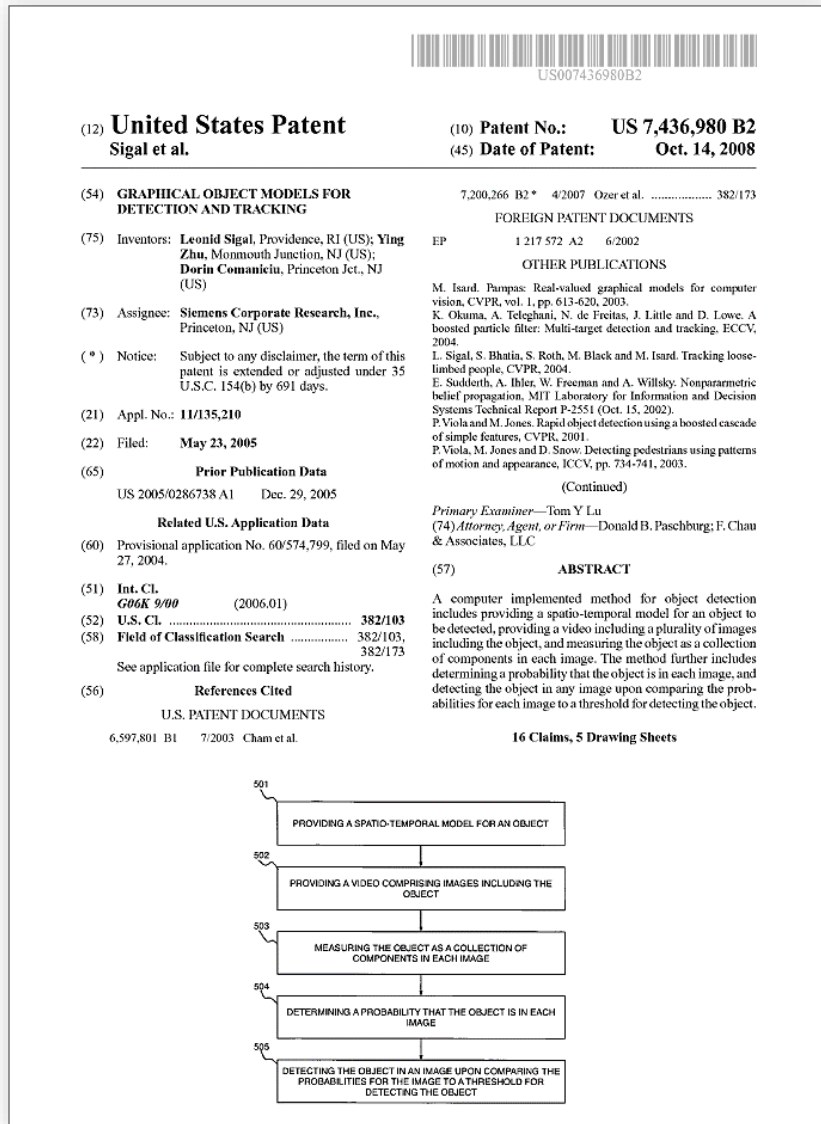
- a. Claims 1-16: Obvious in view of Zhao
- b. Claims 1-16: Obvious over Steffens in view of Zhao
- c. Claims 1-16: Obvious over Ozer in view of Zhao
- d. Claims 1-16: Obvious in view of TLP

Overview of Claim 1

Claim 1, '980 Patent	
Preamble [1]	A computer implemented method for object detection comprising:
Element 1[a]	providing a spatio-temporal model for an object to be detected;
Element 1[b]	providing a video comprising a plurality of images including the object;
Element 1[c]	measuring the object as a collection of components in each image;
Element 1[d]	determining a probability that the object is in each image; and
Element 1[e]	detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.



Overview



1. Background

2. Claim Construction

a. 1/9[a]: spatio-temporal model

- b. $1/9[c]$: measuring the object as a collection of components

c. $1/9[e]$: comparing the probabilities for each image

3. Grounds

a. Claims 1-16: Obvious in view of Zhao

b. Claims 1-16: Obvious over Steffens in view of Zhao

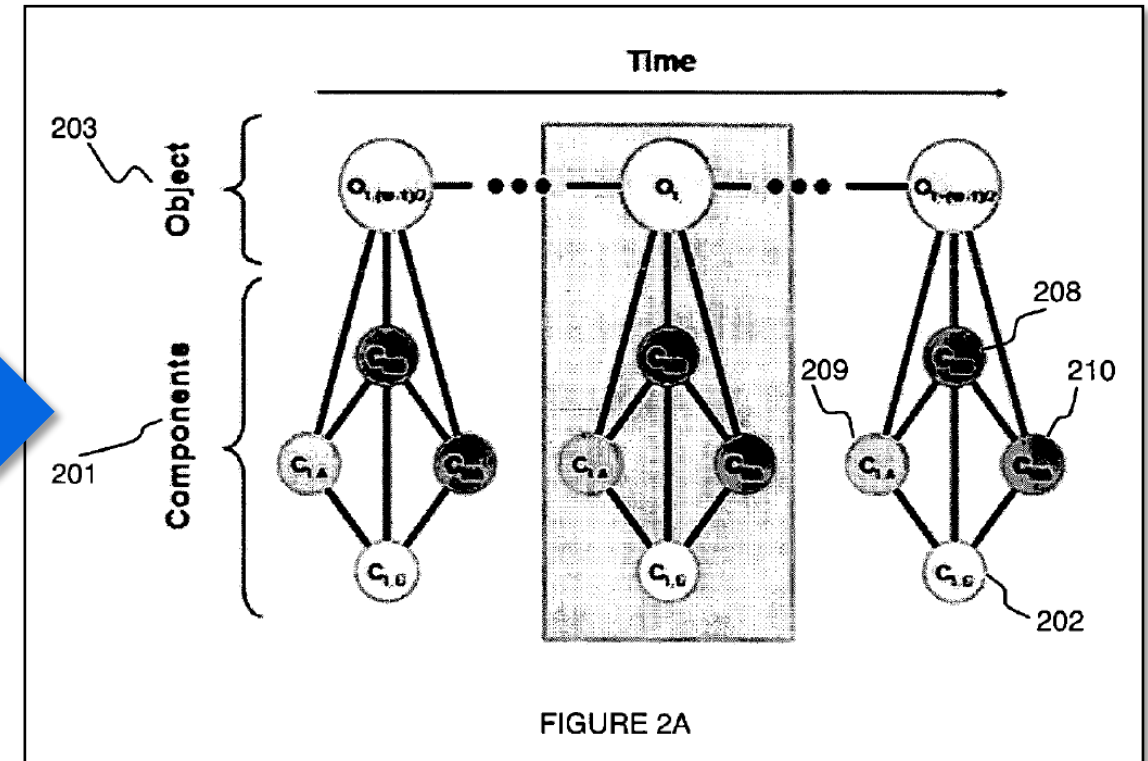
c. Claims 1-16: Obvious over Ozer in view of Zhao

d. Claims 1-16: Obvious in view of TLP

“Spatio-Temporal Model” Should Have Its Plain and Ordinary Meaning

Claim 3

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.



PO's Reliance on Lexicography Is Baseless

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

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

Petitioners

v.

VL COLLECTIVE IP LLC

Patent Owner

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

PATENT OWNER'S SUR-REPLY

B. “spatio-temporal model”

<u>Term</u>	<u>Patent Owner's Construction</u>	<u>Petitioners' Construction²</u>
“spatio-temporal model”	spatial model where an object is detected in an image based on the probabilities that the object is in preceding and/or subsequent images	spatial model where an image's analysis takes into account information from preceding and/or subsequent images

The inventors repeatedly and unambiguously defined “spatio-temporal model.” EX2007, ¶25. PO's construction is based on this lexicography and ensures that the claims capture the essence of the invention.

² Petitioners do not formally propose this construction but apply it in their invalidity analysis. Reply, 8-11 (citing Ex1052, ¶¶6-22, 23-35).

The '980 Does Not Clearly Define “Spatio-Temporal Model”



To act as its own lexicographer, a patentee must “clearly set forth a definition of the disputed claim term” other than its plain and ordinary meaning. *CCS Fitness, Inc. v. Brunswick Corp.*, 288 F.3d 1359, 1366 (Fed. Cir. 2002). It is not enough for a patentee to simply disclose a single embodiment or use a word in the same manner in all embodiments, the patentee must “clearly express an intent” to **redefine the term**. *Helmsderfer v. Bobrick Washroom Equip., Inc.*, 527 F.3d 1379, 1381 (Fed. Cir. 2008); *see also Kara Tech. Inc. v. Stamps.com*, 582 F.3d 1341, 1347–48 (Fed. Cir. 2009).

Thorner v. Sony Computer Entertainment America
669 F.3d 1362, 1365 (Fed. Cir. 2012)

PO's Lexicography Arguments Fail

PO's 1st argument (sur-reply at 17 and below) fails.

The patent repeatedly indicates that the function of the spatio-temporal model is to detect an object, and to do so based on the probabilities that object is in earlier or later images. EX2007, ¶26; EX1001, 1:28-33, 1:37-39, 1:48-52, 2:7-15, 2:59-60, 3:47-4:8, 4:34-52, 5:58-67, 7:45-55, 7:56-8:8, Figures 2A-2B.

Petitioners' response

- There is no clearly set forth definition (or disavowal) in the '980 or its file history that the spatio-temporal model is used to detect an object ***based on the probabilities that the object is in earlier or later images.***
- The '980 recites “spatio-temporal model” in various instances without expressing a clear intent to narrow the scope of that term.

PO's Lexicography Arguments Fail

PO's 2nd argument (sur-reply at 17 and below) fails.

is in preceding and/or subsequent images. *Id.* The patent repeatedly explains that a spatio-temporal model has probabilistic connectivity, *e.g.*, “edges,” “constraints,” or “compatibilities” connecting the object in multiple frames. *Id.*; EX1001, 2:59-60, 3:47-4:8; 4:49-52; Figs. 2A and 2B; EX1003, 4-7 Figs. 2(a)-2(d). The use of object

Petitioners' response

- Claims 1/9 require “a collection of components” without any requirement as to connectivity between components
- Each of the passages cited by PO is directed to limitations recited in dependent claims 3/11.
 - ▶ 2:59-60: **a graphical model** comprising nodes corresponding to each of the collection of components and to the object
 - ▶ 3:47-4:8: the joint probability distribution of a spatio-temporal **graphical model**
 - ▶ Figs. 2A and 2B: exemplary **graphical models** for a pedestrian and vehicle

PO's Lexicography Arguments Fail

PO's 3rd argument (sur-reply at 19 and below) also fails.

actually invented and intended to envelop with the claim.”). The patent demonstrates achievement of these things by comparing a 3-frame spatio-temporal model with a single-frame (non-spatio-temporal) model that had “poor” performance. EX2007, ¶28; EX1001, 7:56-8:3.

Petitioners' response

- Claims 1/9 is a method for object detection (with or without smoothing). PO misconstrues the “w-frame windowed smoothing algorithm” with $w \geq 3$ as a requirement of claims 1/9.

Spatio-Temporal Models Vary in Complexity and Were Well Known

Stochastic Tracking of 3D Human Figures Using 2D Image Motion

Hedvig Sidenbladh¹ Michael J. Black² David J. Fleet²

¹ Royal Institute of Technology (KTH), CVAP/NADA, S-100 44 Stockholm, Sweden
hedvig@nada.kth.se

² Xerox Palo Alto Research Center, 3333 Coyote Hill Rd., Palo Alto, CA 94304 USA
{black,fleet}@parc.xerox.com

Abstract. A probabilistic method for tracking 3D articulated human figures in monocular image sequences is presented. Within a Bayesian framework, we define a generative model of image appearance, a robust likelihood function based on image graylevel differences, and a prior probability distribution over pose and joint angles that models how humans move. The posterior probability distribution over model parameters is represented using a discrete set of samples and is propagated over time using particle filtering. The approach extends previous work on parameterized optical flow estimation to exploit a complex 3D articulated motion model. It also extends previous work on human motion tracking by including a perspective camera model, by modeling limb self occlusion, and by recovering 3D motion from a monocular sequence. The explicit posterior probability distribution represents ambiguities due to image matching, model singularities, and perspective projection. The method relies only on a frame-to-frame assumption of brightness constancy and hence is able to track people under changing viewpoints, in grayscale image sequences, and with complex unknown backgrounds.

1 Introduction

We present a Bayesian approach to tracking 3D articulated human figures in monocular video sequences. The human body is represented by articulated cylinders viewed under perspective projection. A *generative model* is defined in terms of the shape, appearance, and motion of the body, and a model of noise in the pixel intensities. This leads to a likelihood function that specifies the probability of observing an image given the model parameters. A prior probability distribution over model parameters depends on the temporal dynamics of the body and the history of body shapes and motions. With this likelihood function and temporal prior, we formulate the posterior distribution over model parameters at each time instant, given the observation history.

The estimation of 3D human motion from a monocular sequence of 2D images is challenging for a variety of reasons. These include the non-linear dynamics of the limbs, ambiguities in the mapping from the 2D image to the 3D model, the

Stochastic Tracking of 3D Human Figures Using 2D Image Motion

Hedvig Sidenbladh¹ Michael J. Black² David J. Fleet²

¹ Royal Institute of Technology (KTH), CVAP/NADA, S-100 44 Stockholm, Sweden
hedvig@nada.kth.se

² Xerox Palo Alto Research Center, 3333 Coyote Hill Rd., Palo Alto, CA 94304 USA
{black,fleet}@parc.xerox.com

Temporal Models. Temporal models of body limb or joint motion also vary in complexity; they include smooth motion [7], linear dynamical models [18], non-linear models learned from training data using dimensionality reduction [3, 16, 23], and probabilistic Hidden Markov Models (HMM's) (e.g., [4]). In many of these methods, image measurements are first computed and then the temporal models are applied to either smooth or interpret the results. For example, Leventon and Freeman [16] proposed a Bayesian framework for recovering 3D human motion from the motion of a 2D stick figure. They learned a prior distribution over human motions using vector quantization. Given the 2D motion of a set of joints, the most plausible 3D motion could be found. They required a pre-processing step to determine the 2D stick figure motion and did not tie the 3D motion directly to the image. Their Bayesian framework did not represent multi-modal distributions and therefore did not maintain multiple interpretations.

Spatio-Temporal Models Were Not Limited to Graphical Models

**Patent Owner's
Expert Witness**



Dr. Eli Saber

Q Would you describe Zhao as a temporal model within the scope of the definition provided on Page 3 of Exhibit 1016?

ATTORNEY EKLEM: Objection to form and scope.

A So again, as I mentioned, Zhao has a spatial model that has a probabilistic Bayesian, you know, framework to it. The temporal part is a deterministic motion estimation. So you can say that it includes smooth motion, like reference 7 that you saw on that sentence.

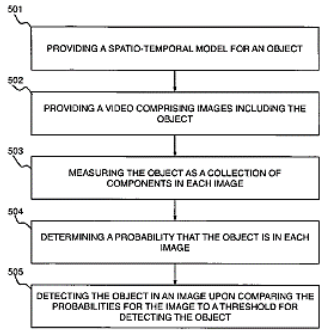
Spatio-Temporal Models Change Configuration Over Time

Petitioners' Expert



Dr. Chandra Bajaj

14. As a precursor to image by image processing steps, limitation 1[a]/9[a] provides a spatio-temporal model for an object, so each of the limitations 1[c]/9[c], 1[d]/9[d] and 1[e]/9[e] can be performed for each of the images comprising the video, provided by limitation 1[b]/9[b]. A POSITA would have understood that claims 1 and 9 are not limited to a graphical model, nor a two-layer graphical model, but a spatio-temporal model that consists of a set or collection of components, each with spatial dimensions that are measurable, and with the ability of this component collection model to change configuration in time.

 US007436980B2	
(12) United States Patent Sigal et al.	(10) Patent No.: US 7,436,980 B2 (45) Date of Patent: Oct. 14, 2008
(54) GRAPHICAL OBJECT MODELS FOR DETECTION AND TRACKING	7,200,266 B2* 4/2007 Ozer et al. 382/173 FOREIGN PATENT DOCUMENTS
(75) Inventors: Leonid Sigal , Providence, RI (US); Ying Zhu , Monmouth Junction, NJ (US); Dorin Comaniciu , Princeton Jct., NJ (US)	EP 1 217 572 A2 6/2002 OTHER PUBLICATIONS
(73) Assignee: Siemens Corporate Research, Inc. , Princeton, NJ (US)	M. Isard, Pumpas: Real-valued graphical models for computer vision, CVPR, vol. 1, pp. 613-620, 2003. K. Okuma, A. Teleghani, N. de Freitas, J. Little and D. Lowe. A boosted particle filter: Multi-target detection and tracking. ECCV, 2004. L. Sigal, S. Bhatia, S. Roth, M. Black and M. Isard. Tracking loose-limbed people, CVPR, 2004. E. Sudderth, A. Ihler, W. Freeman and A. Willsky. Nonparametric belief propagation. MIT Laboratory for Information and Decision Systems Technical Report P-2551 (Oct. 15, 2002). P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features, CVPR, 2001. P. Viola, M. Jones and D. Snow. Detecting pedestrians using patterns of motion and appearance, ICCV, pp. 734-741, 2003.
(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 691 days.	(Continued)
(21) Appl. No.: 11/135,210	Primary Examiner—Tom Y Lu (74) Attorney, Agent, or Firm—Donald B. Paschburg, F. Chau & Associates, LLC
(22) Filed: May 23, 2005	
(65) Prior Publication Data US 2005/0286738 A1 Dec. 29, 2005	
Related U.S. Application Data	
(60) Provisional application No. 60/574,799, filed on May 27, 2004.	(57) ABSTRACT
(51) Int. Cl. G06K 9/00 (2006.01)	A computer implemented method for object detection includes providing a spatio-temporal model for an object to be detected, providing a video including a plurality of images including the object, and measuring the object as a collection of components in each image. The method further includes determining a probability that the object is in each image, and detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.
(52) U.S. Cl. 382/103	
(58) Field of Classification Search 382/103, 382/173	
See application file for complete search history.	
(56) References Cited U.S. PATENT DOCUMENTS 6,597,801 B1 7/2003 Cham et al.	
16 Claims, 5 Drawing Sheets	
 <pre>graph TD 501[PROVIDING A SPATIO-TEMPORAL MODEL FOR AN OBJECT] --> 502[PROVIDING A VIDEO COMPRISING IMAGES INCLUDING THE OBJECT] 502 --> 503[MEASURING THE OBJECT AS A COLLECTION OF COMPONENTS IN EACH IMAGE] 503 --> 504[DETERMINING A PROBABILITY THAT THE OBJECT IS IN EACH IMAGE] 504 --> 505[DETECTING THE OBJECT IN AN IMAGE UPON COMPARING THE PROBABILITIES FOR THE IMAGE TO A THRESHOLD FOR DETECTING THE OBJECT]</pre>	

1. Background

2. Claim Construction

a. 1/9[a]: spatio-temporal model

b. 1/9[c]: measuring the object as a collection of components

c. 1/9[e]: comparing the probabilities for each image

3. Grounds

a. Claims 1-16: Obvious in view of Zhao

b. Claims 1-16: Obvious over Steffens in view of Zhao

c. Claims 1-16: Obvious over Ozer in view of Zhao

d. Claims 1-16: Obvious in view of TLP

1[c]: The Object Is Represented by a Collection of Components

Claim 1, '980 Patent	
Preamble [1]	A computer implemented method for object detection comprising:
Element 1[a]	providing a spatio-temporal model for an object to be detected;
Element 1[b]	providing a video comprising a plurality of images including the object;
Element 1[c]	measuring the <i>object as a collection of components</i> in each image;
Element 1[d]	determining a probability that the object is in each image; and
Element 1[e]	detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.

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

Providing the spatio-temporal model for the object to be detected comprises providing detectors for each of the collection of components.

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.

on character recognition task. Unlike all previous approaches however we provide a automatic probabilistic framework that explicitly models objects using components, integrates temporal information, and allows for simultaneous tracking. The closest in spirit comes the

The '980 Represents the Object as a Collection of Components

**Patent Owner's
Expert Witness**



Dr. Eli Saber

In an attempt to distinguish the '980 from Zhao, Dr. Saber acknowledged that the '980 represents the object as a collection of components:

Q ... Can contours represent components?

A Contours are contours. So the contour of this cup is not the same as the interior region of that cup. So the contour is the contour. That's what is being modeled in -- that's what Zhao relies on.

The '980 looks at the object as a series of components. The component represents the interior part of the object like a region, versus in Zhao you're looking at the contour. So they're not the same thing.

Claims 1/9 Do Not Distinguish the Object from Its Components

Petitioners' Expert



Dr. Chandra Bajaj

16. In light of the '980 specification, a POSITA would have understood that “a probability” that the object recited in 1[d]/9[d] is in each image is determined as a function of the respective probabilities of the “collection of components” recited in 1[c]/9[c]. The '980 specification indicates by way of example that a joint probability distribution for the spatio-temporal model with N components is trained to determine whether an object is present in an image. *See* '980 Patent at 3:53-65 and 5:25-67. In order to detect the object in “any” one of the plurality images as recited by 1[e]/9[e], the '980 Patent indicates that AdaBoost can be used to boost the confidence that the “collection of components” are present in the image.

Boosted detectors are trained for each component. For simplicity AdaBoost may be used without a cascade. Training with a cascade would likely improve the computational efficiency of the system. To reduce the number of false positives produced by the detectors, a bootstrapping method iteratively adds false positives that are collected by running the trained strong classifier over the set of background images (e.g., those not containing the desired object) and the detectors are retrained using the old positive and the new extended negative sets.

Petitioners' and Dr. Bajaj's Treatment of 1[c] Remains Unchanged

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEALS BOARD

META PLATFORMS, INC.; INSTAGRAM
WHATSAPP LLC; META PLATFORMS TECHNOLOGIES, INC.
and GIPHY, INC.
Petitioners

v.

VL COLLECTIVE IP LLC
Patent Owner

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

Case IPR2023-00924

DECLARATION OF CHANDRAJIT BAJAJ, PH.D.
PETITION FOR *INTER PARTES* REVIEW
U.S. PATENT NO. 7,436,980

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 was disclosed or at least would have been obvious 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 at 1.

PO's Argument that "Object" and "Components" Are Distinct Fails

PO's 1st argument (sur-reply at 6 and below) fails.

This distinctness is confirmed by the specification. Figures 2A and 2B depict the "object as a whole" 203 as a distinct entity from the collection of "components" 202. EX1001, 4:18-48; EX2007, ¶15. Modeling objects separately from their constituent components has a purpose: it "allows the inference process to reason explicitly about the object as a whole" and assume "conditional independence of components." *Id.*; EX1001, 4:37-42. Furthermore, the joint probability distribution

Petitioners' response

- Claims 1/9 are not limited to a spatio-temporal model comprising nodes corresponding to each of the collection of components and to the object. Compare with claims 3/11.

PO's Argument that "Object" and "Components" Are Distinct Fails

PO's 2nd argument (sur-reply at 6 and below) fails.

components.” *Id.*; EX1001, 4:37-42. Furthermore, the joint probability distribution described at 3:47-4:7 expressly includes distinct states of the object (“O”) ($X_0^O \dots X_T^O$) and each component (“C”) ($X_0^{C_0}, X_0^{C_1} \dots X_0^{C_N} \dots X_T^{C_0}, X_T^{C_1} \dots X_T^{C_N}$). *Id.* It also includes multiple probabilistic constraints (potential functions) that distinguish between the object and components: $\psi_{ij}(X_i^O, X_j^O)$ for object-to-object constraints;

Petitioners’ response

- Claims 1/9 only claim “probabilities” of the collection of components—and not a joint probability distribution as claimed in dependent claim 8/16

PO's Reliance on Claim Differentiation Is Misplaced



As we have previously held, “each claim in a patent is presumptively different in scope.” *RF Del., Inc. v. Pac. Keystone Techs., Inc.*, 326 F.3d 1255, 1263 (Fed.Cir.2003). Thus, in a situation where dependent claims have no meaningful difference other than an added limitation, the independent claim is not restricted by the added limitation in the dependent claim.

Trustees of Columbia Univ. in City of New York v. Symantec Corp.
811 F.3d 1359, 1364 (Fed. Cir. 2016)

PO's Argument that “Object” and “Components” Are Distinct Fails

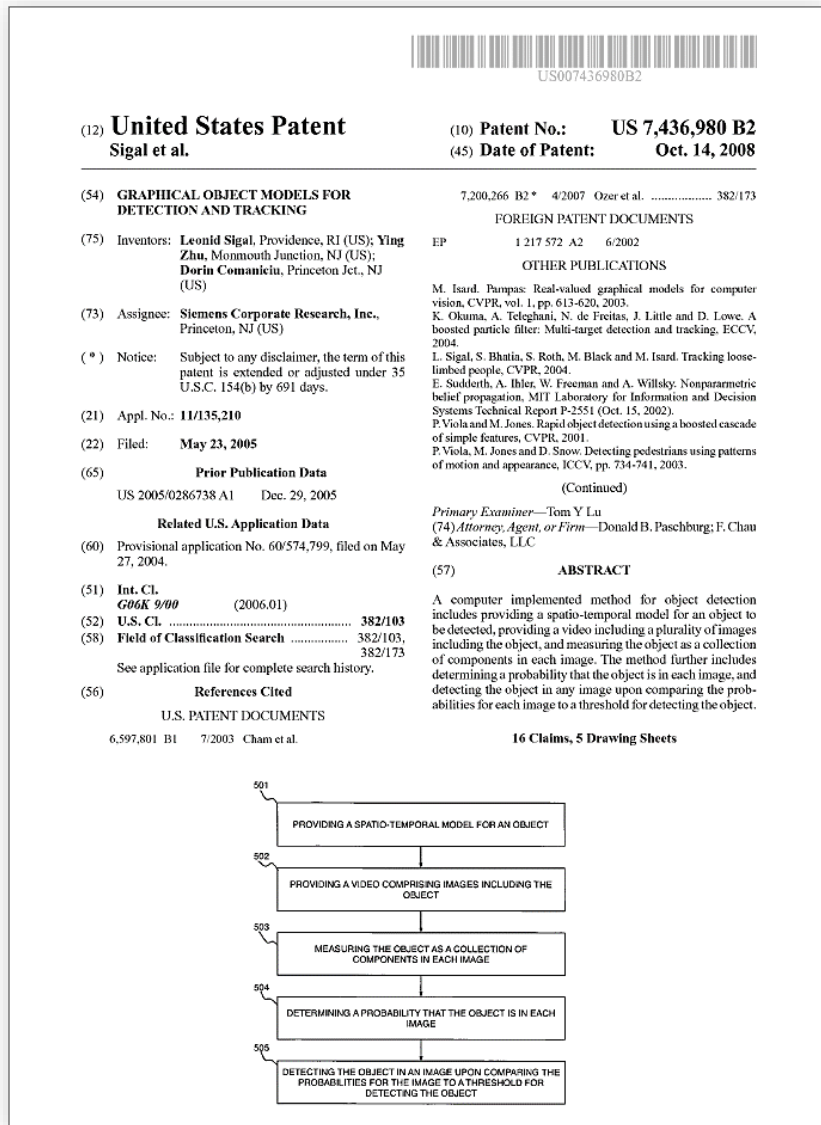
PO's 3rd argument (below)—that Petitioners' construction violates the doctrine of claim differentiation—also fails.

Also, because dependent claims 3/11 expressly distinguish between an “object” and a “collection of components,” independent claims 1/9 must be broad enough to encompass embodiments in which they are distinct. *Littelfuse, Inc. v. Mersen USA EP Corp.*, 29 F.4th 1376, 1380 (Fed. Cir. 2022). Yet under Bajaj's

Petitioners' response:

- Claims 1/9 are not limited to a graphical model (EX1052, ¶14), so dependent claims 3/11 are narrower (different) in scope than claims 1/9, satisfying the claim differentiation doctrine.

Overview



1. Background

2. Claim Construction

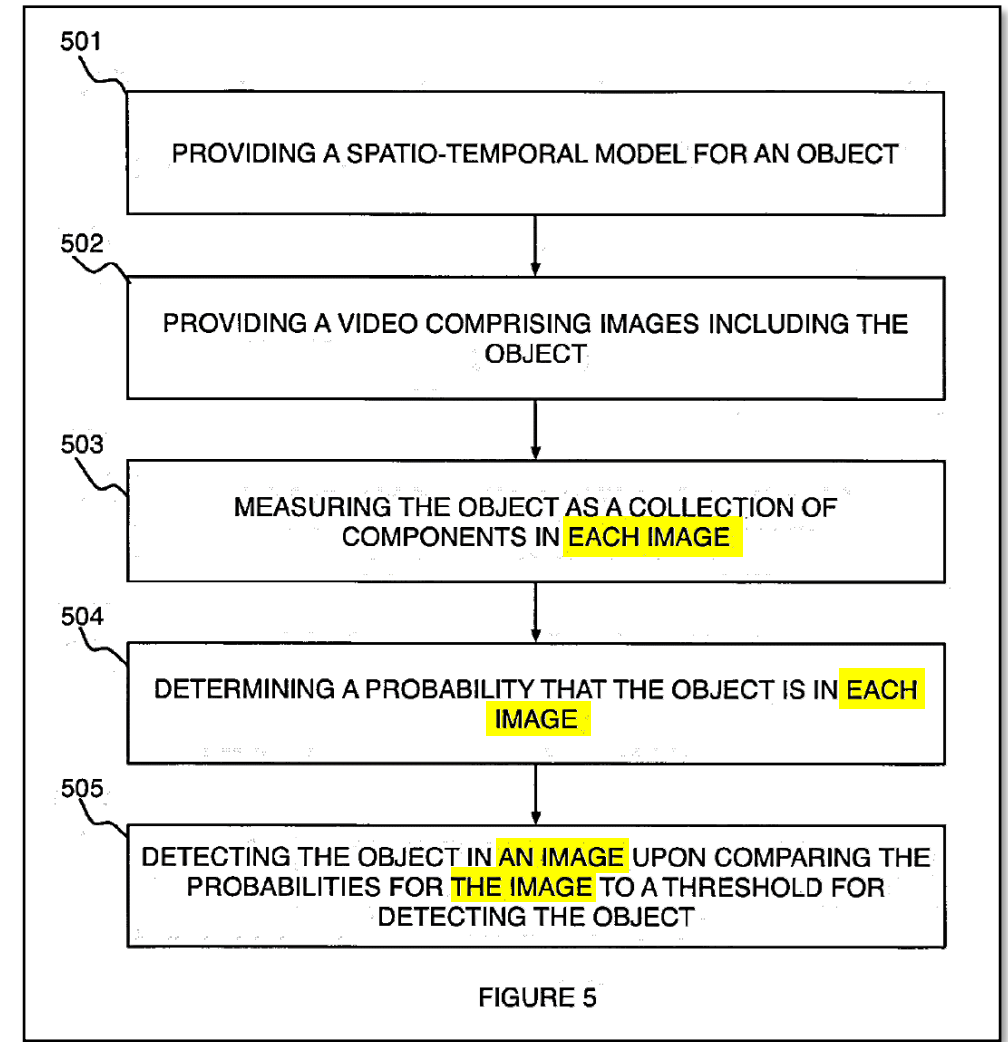
- a. $1/9[a]$: spatio-temporal model
- b. $1/9[c]$: measuring the object as a collection of components
- c. $1/9[e]$: comparing the probabilities for each image**

3. Grounds

- a. Claims 1-16: Obvious in view of Zhao
- b. Claims 1-16: Obvious over Steffens in view of Zhao
- c. Claims 1-16: Obvious over Ozer in view of Zhao
- d. Claims 1-16: Obvious in view of TLP

1[e] Does Not Require a Multi-Image Analysis for Detecting the Object

Claim 1, '980 Patent	
Preamble [1]	A computer implemented method for object detection comprising:
Element 1[a]	providing a spatio-temporal model for an object to be detected;
Element 1[b]	providing a video comprising a plurality of images including the object;
Element 1[c]	measuring the object as a collection of components in each image;
Element 1[d]	determining a probability that the object is in each image; and
Element 1[e]	detecting the object in any image upon <i>comparing the probabilities for each image to a threshold</i> for detecting the object.



1[e] Does Not Require a Multi-Image Analysis for Detecting the Object

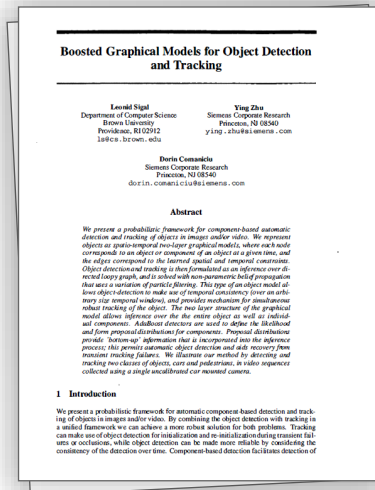
Petitioners' Expert



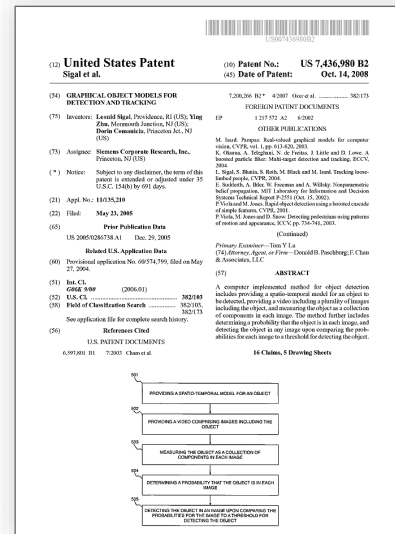
Dr. Chandra Bajaj

35. With regard to elements 1[d]/9[d] “determining a probability that the object is in each image” and 1[e]/9[e] “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object,” the image processing is similar and accomplished individually for a current image frame, and sequentially for all image frames of the video, an image frame at a time. That is, under the w=1 scenario, steps 1/9[c], [d] and [e] is performed image frame by image frame of the video image sequence, without any fusion from previous or future image frames, using the multi-component object probabilities for detecting the presence of the object in each image frame. A POSITA would have known

The '980 and its Provisional Expressly Indicate That w Is an Odd Integer ≥ 1



We use w -frame windowed smoothing algorithm for our tracking and object detection, where w is an odd integer ≥ 1 (see Fig. 2cd). There are two ways one can do windowed smoothing, what we call an object-detection centric way, and the tracking centric way. In an object-detection centric way we re-initialize all nodes every time we shift a window, hence the temporal integration is only applied in the window of size w . In the tracking centric way we only initialize the nodes associated with a new frame, which tends to force temporal consistency from before $t - (w - 1)/2$. While tracking-centric way tends to converge faster and produce more consistent results over time, it also less sensitive to objects entering and leaving the scene. Note that with $w = 1$, the algorithm resembles single frame component-based fusion [8].



While it is possible to perform inference over the spatio-temporal model defined for the entire image sequence, there are many applications for which this is not an option due to the lengthy off-line processing needed. A w -frame windowed smoothing algorithm is used where w is an odd integer ≥ 1 . There are two ways one can do windowed smoothing: in an object-detection centric way or a tracking centric way. In the former all nodes are re-initialized every time a window is shifted, hence the temporal integration is only applied in the window of size w . In the tracking centric way only the nodes associated with a new frame are initialized, which tends to enforce temporal consistency from before $t - (w - 1)/2$. While the tracking-centric way tends to converge faster and produce more consistent results over time, it is also less sensitive to objects entering and leaving the scene. Note that with $w=1$, the algorithm resembles single frame component-based fusion.

Any Criticism of the w=1 Scenario Does Not Rise to the Level of Disavowal



Mere criticism of a particular embodiment encompassed in the plain meaning of a claim term ***is not sufficient to rise to the level of clear disavowal***. *Epistar Corp. v. Int'l Trade Comm'n*, 566 F.3d 1321, 1335 (Fed.Cir.2009) (holding that even a direct criticism of a particular technique did not rise to the level of clear disavowal). In *Spine Solutions, Inc. v. Medtronic Sofamor Danek USA, Inc.*, we explained that even where a particular structure makes it “particularly difficult” to obtain certain benefits of the claimed invention, this does not rise to the level of disavowal of the structure. 620 F.3d 1305, 1315 (Fed.Cir.2010).

Thorner v. Sony Computer Entertainment America LLC,
669 F.3d 1362, 1366 (Fed. Cir. 2012)

Belief Propagation Is Still Applicable When $w=1$

Petitioners' Expert



Dr. Chandra Bajaj

PO's 1st argument:

The detection inference is defined as an estimation of “belief” in a model over a w -frame window. POR at 25.

Dr. Bajaj's response:

29. In this single image frame processing without windowing (*i.e.*, the $w=1$ scenario), BP is performed between the likelihoods of multiple components present within the image frame for their spatial location and spatial connectivity of the multi-component spatio-temporal model. That is, even when $w=1$, BP is still applicable to the detection and determination of the likelihood probabilities of the multiple components of the model in each image frame. Thus, even though the

'980 Object Detection Can Be Accomplished When $w=1$

PO's 2nd argument:

the '980 teaches that "object detection" is "formulated as an inference" in a "spatio-temporal model" that covers more than "a single frame" **because it is "extended over time"** and therefore includes instances of the object at a "**current time**" and "**earlier**" or "**later**" times. POR at 24-25.

Petitioners' response:

- ▶ The '980 indicates that in order to integrate temporal constraints the spatial **graphical models** are **extended over time** to an **arbitrary length temporal window**. 4:34-36.
 - ▶ The '980 expressly indicates that the temporal window can be set to $w=1$ frames. '980, 7:10.
 - ▶ Claims 1/9 are not limited to a graphical model
 - ▶ Claims 1/9 do not require w -frame windowed smoothing, $w \geq 3$

The '980 Model Is a Temporal Model When $w=1$

Patent Owner's
Expert Witness



Dr. Eli Saber

PO's Argument: Dr. Saber begrudgingly acknowledged that the configuration of the '980 model changes frame-to-frame when $w=1$:

Q Has it changed -- has the configuration of the model changed from frame to frame for W equals 1?

ATTORNEY EKLEM: Objection to form.

A For W equals 1, you have a frame. You have a model for that frame. That model has now been reduced to just a spatial model. You have no temporal elements. Your formulation here, you've just ripped off a section of that formulation.

Q How --

A Let me finish. You go to the next frame, you again have a spatial model. You ripped off the temporal. That's not what's being practiced in the '980.

Q Are those two spatial models you just mentioned, do they have different configurations?

A Well, they have to -- they have to model what's in the frame. So they potentially could be

- ▶ A POSITA would have understood that a spatial model which changes configuration over time is a spatio-temporal model.

Overview

US007436980B2

(12) **United States Patent**
Sigal et al.

(10) **Patent No.:** **US 7,436,980 B2**
(45) **Date of Patent:** **Oct. 14, 2008**

(54) **GRAPHICAL OBJECT MODELS FOR
DETECTION AND TRACKING**

7,200,266 B2 * 4/2007 Ozer et al. 382/173

FOREIGN PATENT DOCUMENTS

(75) Inventors: **Leonid Sigal**, Providence, RI (US); **Ying Zhu**, Monmouth Junction, NJ (US);
Dorin Comaniciu, Princeton Jct., NJ
(US)

EP 1 217 572 A2 6/2002

OTHER PUBLICATIONS

(73) Assignee: **Siemens Corporate Research, Inc.**,
Princeton, NJ (US)

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 691 days.

M. Isard, P. Papanicolaou, Real-valued graphical models for computer
vision, CVPR, vol. 1, pp. 613-620, 2003.

K. Okuma, A. Tachibana, N. de Freitas, J. Little and D. Lowe. A
boosted particle filter: Multi-target detection and tracking. ECCV,
2004.

L. Sigal, S. Bhatia, S. Roth, M. Black and M. Isard. Tracking loose-
ly-labeled people, CVPR, 2004.

E. Sudderth, A. Ihler, W. Freeman and A. Willsky. Nonparametric
belief propagation. MIT Laboratory for Information and Decision
Systems Technical Report P-2551 (Oct. 15, 2002).

P. Viola and M. Jones. Rapid object detection using a boosted cascade
of simple features, CVPR, 2001.

P. Viola, M. Jones and D. Snow. Detecting pedestrians using patterns
of motion and appearance, ICCV, pp. 734-741, 2003.

(21) Appl. No.: **11/135,210**

(22) Filed: **May 23, 2005**

(65) **Prior Publication Data**

US 2005/0286738 A1 Dec. 29, 2005

(Continued)

Related U.S. Application Data

(60) Provisional application No. 60/574,799, filed on May
27, 2004.

Primary Examiner—Tom Y Lu

(74) Attorney, Agent, or Firm—Donald B. Paschburg; F. Chau
& Associates, LLC

(51) Int. Cl.
G06K 9/00 (2006.01)

(52) U.S. Cl. **382/103**

(58) Field of Classification Search **382/103,**
382/173

See application file for complete search history.

(57) **ABSTRACT**

A computer implemented method for object detection
includes providing a spatio-temporal model for an object to
be detected, providing a video including a plurality of images
including the object, and measuring the object as a collection
of components in each image. The method further includes
determining a probability that the object is in each image, and
detecting the object in any image upon comparing the prob-
abilities for each image to a threshold for detecting the object.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,597,801 B1 7/2003 Cham et al.

16 Claims, 5 Drawing Sheets

```
graph TD
    501[PROVIDING A SPATIO-TEMPORAL MODEL FOR AN OBJECT] --> 502[PROVIDING A VIDEO COMPRISING IMAGES INCLUDING THE OBJECT]
    502 --> 503[MEASURING THE OBJECT AS A COLLECTION OF COMPONENTS IN EACH IMAGE]
    503 --> 504[DETERMINING A PROBABILITY THAT THE OBJECT IS IN EACH IMAGE]
    504 --> 505[DETECTING THE OBJECT IN AN IMAGE UPON COMPARING THE PROBABILITIES FOR THE IMAGE TO A THRESHOLD FOR DETECTING THE OBJECT]
```

1. Background

2. Claim Construction

- a. $1/9[a]$: spatio-temporal model
- b. $1/9[c]$: measuring the object as a collection of components
- c. $1/9[e]$: comparing the probabilities for each image

3. Grounds

- a. **Claims 1-16: Obvious in view of Zhao**
- b. Claims 1-16: Obvious over Steffens in view of Zhao
- c. Claims 1-16: Obvious over Ozer in view of Zhao
- d. Claims 1-16: Obvious in view of TLP

Prior Art - Zhao (EX1006)

**Dressed Human Modeling, Detection,
and Parts Localization**

Liang Zhao

CMU-RI-TR-01-19

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

July 26, 2001

*Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy*

© 2001 Liang Zhao

This research was partially sponsored by PennDOT Agreement No. 62N111 Project
TA-34 executed under the Federal Technical Assistance Grant Program 001-008-701-99-7
and SAIC.

Meta 1006
p. 1

- ▶ Zhao is a dissertation submitted to the Robotics Institute of Carnegie Mellon University
- ▶ Primary reference in Ground I
- ▶ Secondary reference in Grounds II and III
- ▶ Publicly available no later than September 23, 2002

Ground I: Obvious in View of Zhao

Claim 1, '980 Patent

Preamble [1]	A computer implemented method for object detection comprising:
Element 1[a]	providing a <i>spatio-temporal model</i> for an object to be detected;
Element 1[b]	providing a video comprising a plurality of images including the object;
Element 1[c]	measuring the object as a collection of components in each image;
Element 1[d]	determining a probability that the object is in each image; and
Element 1[e]	detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.

Zhao’s dynamic human model is a “spatio-temporal model” when that term is accorded:

- 1. Its plain and ordinary meaning; and
- 2. PO’s construction in D. Del.

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.

Spatio-Temporal Model -- Plain and Ordinary Meaning

Dressed Human Modeling, Detection, and Parts Localization

Liang Zhao
CMU-RI-TR-01-19

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

July 26, 2001

*Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy*

© 2001 Liang Zhao

This research was partially sponsored by PennDOT Agreement No. 62N111 Project
TA-34 executed under the Federal Technical Assistance Grant Program 001-008-701-99-7
and SAIC.

Meta 1006
p. 1

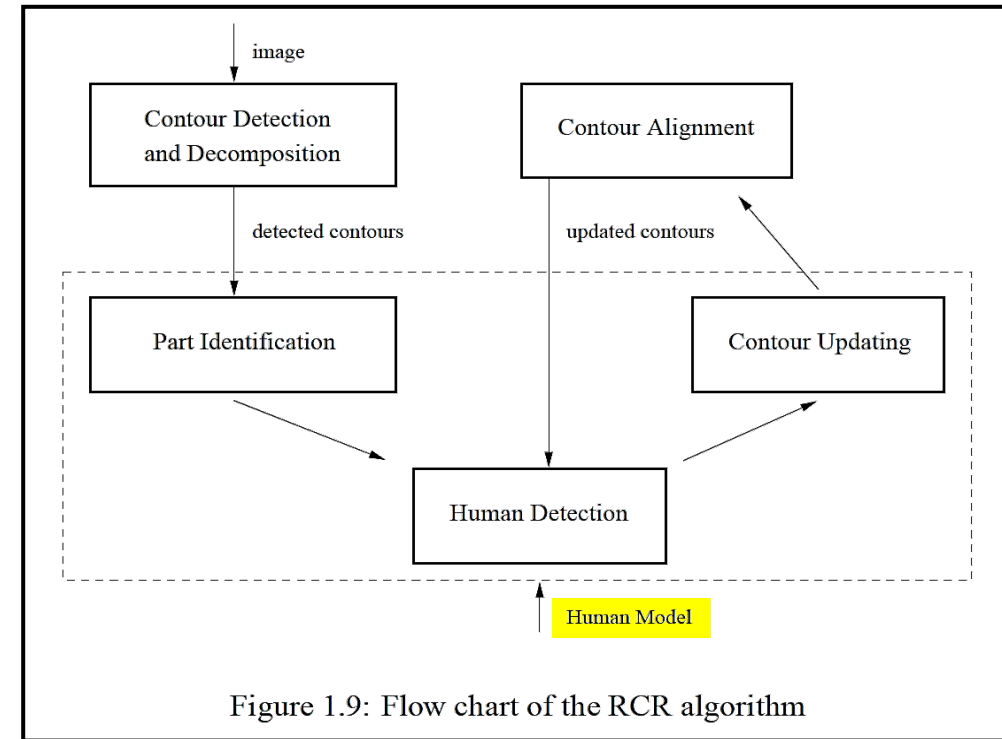
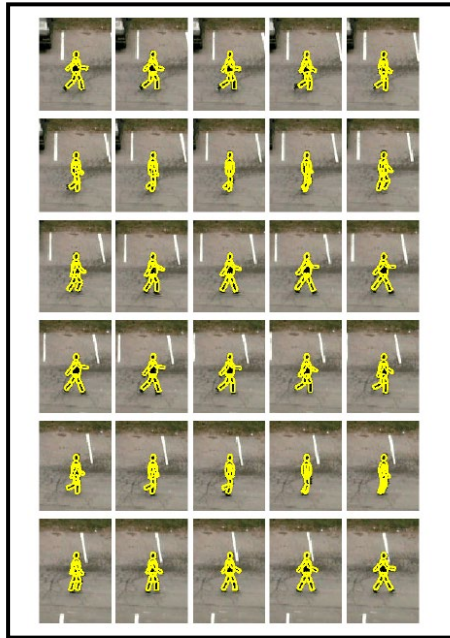
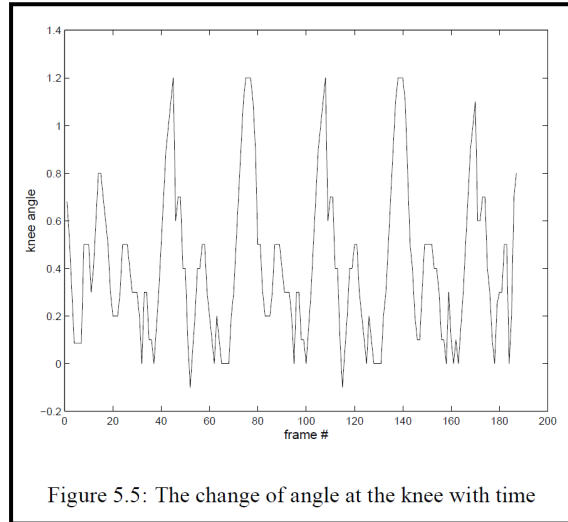


Figure 1.9: Flow chart of the RCR algorithm

are readily applicable to other tasks and to detecting other objects. First, the TRS-invariant probabilistic representation can be used to model other object categories as well. Second, the Bayesian similarity measure can be used to classify other objects and to perform shape retrieval. Third, the RCR algorithm is useful for vision tasks such as object tracking as well by incorporating motion information into the model.

Spatio-Temporal Model -- Plain and Ordinary Meaning



- ▶ Zhao, Figures 5.5 and 5.6 illustrate the effect of incorporating temporal motion information in Zhao's system and demonstrates the change of angle at a subject's knee with time.

Another problem is that the left and right limbs tend to be confused in a side-view. Motion information obtained from previous frames can be used to predict the orientations of the limbs and to solve the ambiguity. Fig. 5.6 presents a full cycle of a walking person. In the first half cycle, no motion information is used to resolve the ambiguity with limbs' orientations, while in the second half of the cycle, the prediction from previous frames (constant angular velocity is assumed) is used to get better results of body part localization. Fig. 5.5 illustrates how the left knee angle changes with time. The motion information is not used from frames 1 to 30, but is from frames 31 to 190. Therefore, the left and right legs are switched sometimes during the first part. From frames 31 to 190, an obvious pattern of walking cycle can be observed.

Saber: Zhao's Modeling Constitutes Spatio-Temporal Modeling

Patent Owner's Expert Witness



Dr. Eli Saber

Q -- does Zhao teach a temporal body model?

ATTORNEY EKLEM: Objection to form.

A See, I don't know what you mean by model that changes configuration over time. That's why Dr. Bajaj is incorrect in his -- in his definition.

This model in Zhao, yes, the human is moving from frame to frame so that the human is changing from frame to frame. The probabilistic modeling that Zhao does on a given frame, he does that -- Zhao does that spatially.

There's deterministic motion estimation from frame to frame. There's no probabilistic modeling. There's no temporal model as in the context of the '980. His model is strictly motion deterministic motion model.

So his temporal modeling is deterministic motion modeling. There's no probabilistic

Saber Acknowledges That Zhao's Model Is a Spatio-Temporal Model

**Patent Owner's
Expert Witness**



Dr. Eli Saber

Q Would you describe Zhao as a temporal model within the scope of the definition provided on Page 3 of Exhibit 1016?

ATTORNEY EKLEM: Objection to form and scope.

A So again, as I mentioned, Zhao has a spatial model that has a probabilistic Bayesian, you know, framework to it. The temporal part is a deterministic motion estimation. So you can say that it includes smooth motion, like reference 7 that you saw on that sentence.

Consistent with PO's Position in Dist. Ct. Briefing –Zhao Uses a Graphical Model

Petitioners' Expert



Dr. Chandra Bajaj

84. Further, in my opinion a POSITA would have found this element disclosed or at least obvious, as the Zhao object model is a graphical model with nodes corresponding to an object and a collection of its components. Zhao explains, “[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 at 42. The model is defined as a graphical model that’s comprised of “ribbons” for these body parts, as shown in Figure 2.6(a):

Consistent with PO's D. Del. Briefing—Zhao Uses a Graphical Model

Dressed Human Modeling, Detection, and Parts Localization

Liang Zhao
CMU-RI-TR-01-19

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

July 26, 2001

*Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy*

© 2001 Liang Zhao

This research was partially sponsored by PennDOT Agreement No. 62N111 Project
TA-34 executed under the Federal Technical Assistance Grant Program 001-008-701-99-7
and SAIC.

Meta 1006
p. 1

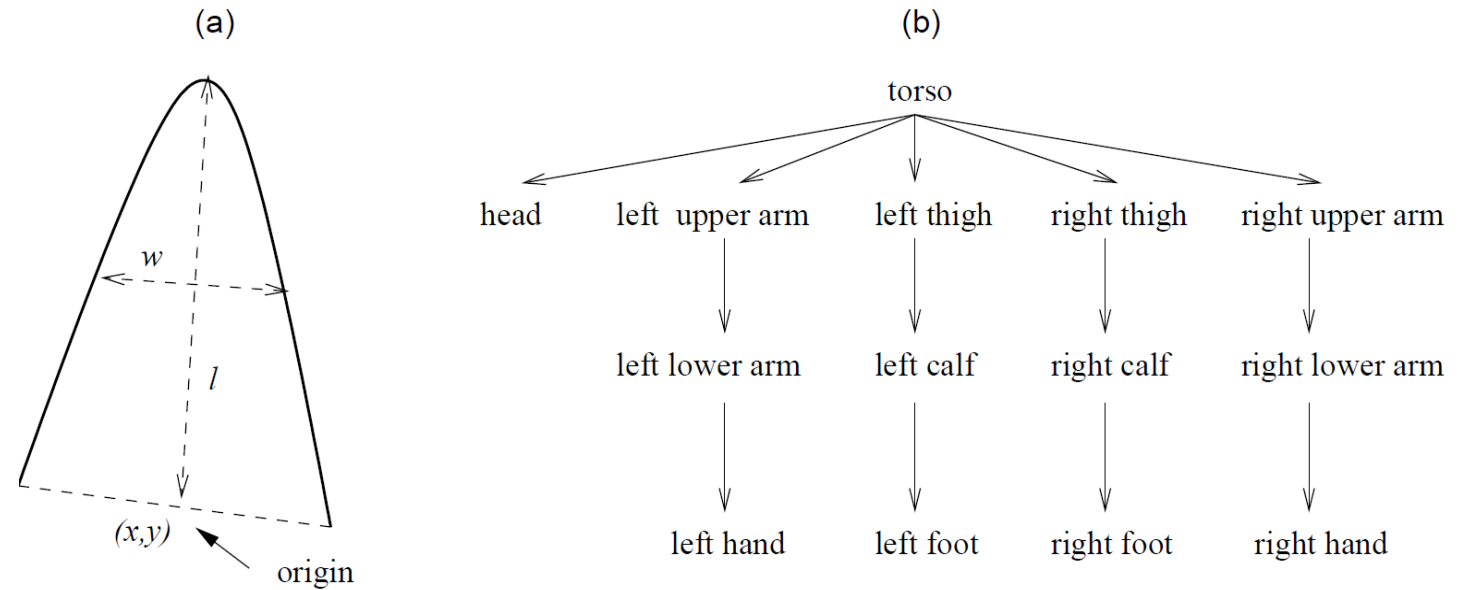


Figure 2.6: (a) Body part model (b) “connect-to” hierarchy

The body parts are modeled with ribbons¹ [114, 115]
as shown in Fig. 2.6(a).

Consistent with PO's D. Del. Briefing—Zhao Models the body and components

Petitioners' Expert



Dr. Chandra Bajaj

person.” EX1006 at 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:

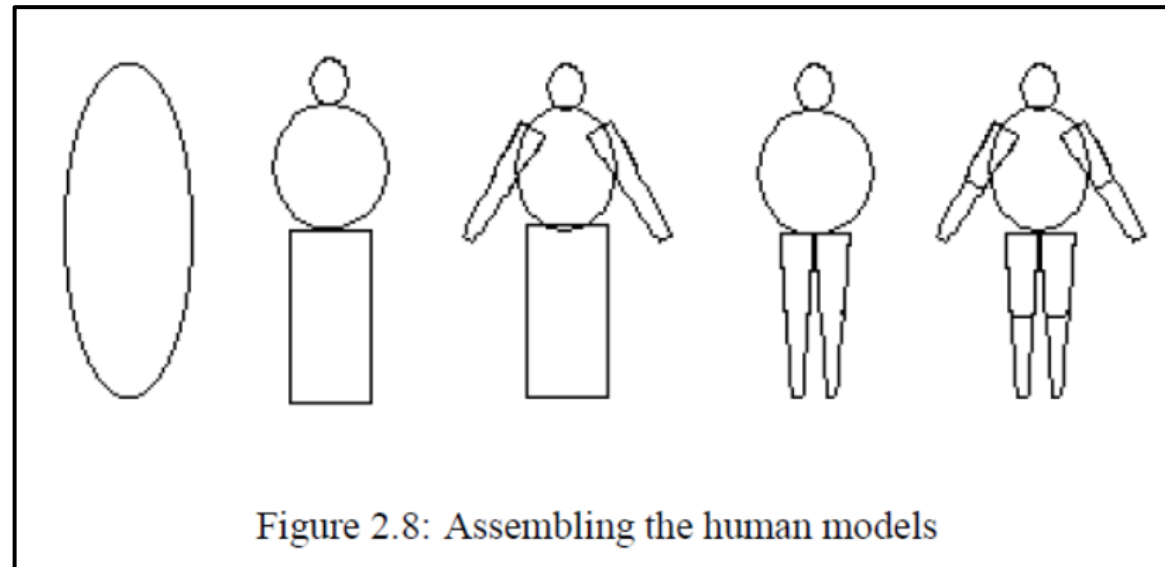


Figure 2.8: Assembling the human models

Ground I: Obvious in View of Zhao

Claim 1, '980 Patent

Preamble [1]	A computer implemented method for object detection comprising:
Element 1[a]	providing a spatio-temporal model for an object to be detected;
Element 1[b]	providing a video comprising a plurality of images including the object;
Element 1[c]	measuring the object as a collection of components in each image;
Element 1[d]	<i>determining a probability that the object is in each image;</i> and
Element 1[e]	detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.

- ▶ Zhao uses a Bayesian similarity measure to determine a probability that an object is in each image
- ▶ A contour C corresponds to a person if

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

1[d]: Zhao Determines a Probability That the Object Is in Each Image

Dressed Human Modeling, Detection, and Parts Localization

Liang Zhao
CMU-RI-TR-01-19

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

July 26, 2001

*Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy*

© 2001 Liang Zhao

This research was partially sponsored by PennDOT Agreement No. 62N111 Project
TA-34 executed under the Federal Technical Assistance Grant Program 001-008-701-99-7
and SAIC.

Meta 1

3.3 Bayesian Similarity Measure and Body Part Identification

A part-based similarity measure evaluates the resemblance between a contour and a model based on the best match between their body parts. This section presents the definition of a Bayesian Similarity Measure derived from the probabilistic human model, and illustrates how to use this measure to guide the matching between a set of decomposed ribbons and a set of model body parts through coarse-to-fine model assembling.

The conditional probability $P(\text{person}|H)$ 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. According to the above observation, the following formula is designed to estimate $P(\text{person}|H)$:

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

Ground I: Obvious in View of Zhao

Claim 1, '980 Patent

Preamble [1]	A computer implemented method for object detection comprising:
Element 1[a]	providing a spatio-temporal model for an object to be detected;
Element 1[b]	providing a video comprising including the object;
Element 1[c]	measuring the object as a co components in each image;
Element 1[d]	determining a probability th image; and
Element 1[e]	<i>detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.</i>

According to the assumption made in section 3.3.1, $P(H^*)/P(person)$ is a constant. Thus, the decision rule can be further simplified as: the contour C corresponds to a person if

$$P(C|H^*, person)P(person|H^*) > threshold \quad (3.17)$$

where the left hand side of Eq. (3.17) is just the similarity measure defined in Eq. (3.5).

$$BSM(C) \triangleq G(H^*), \quad (3.5)$$

Zhao's Testing Confirm that Zhao's Model Is Temporal

Dressed Human Modeling, Detection, and Parts Localization

Liang Zhao
CMU-RI-TR-01-19

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

July 26, 2001

*Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy*

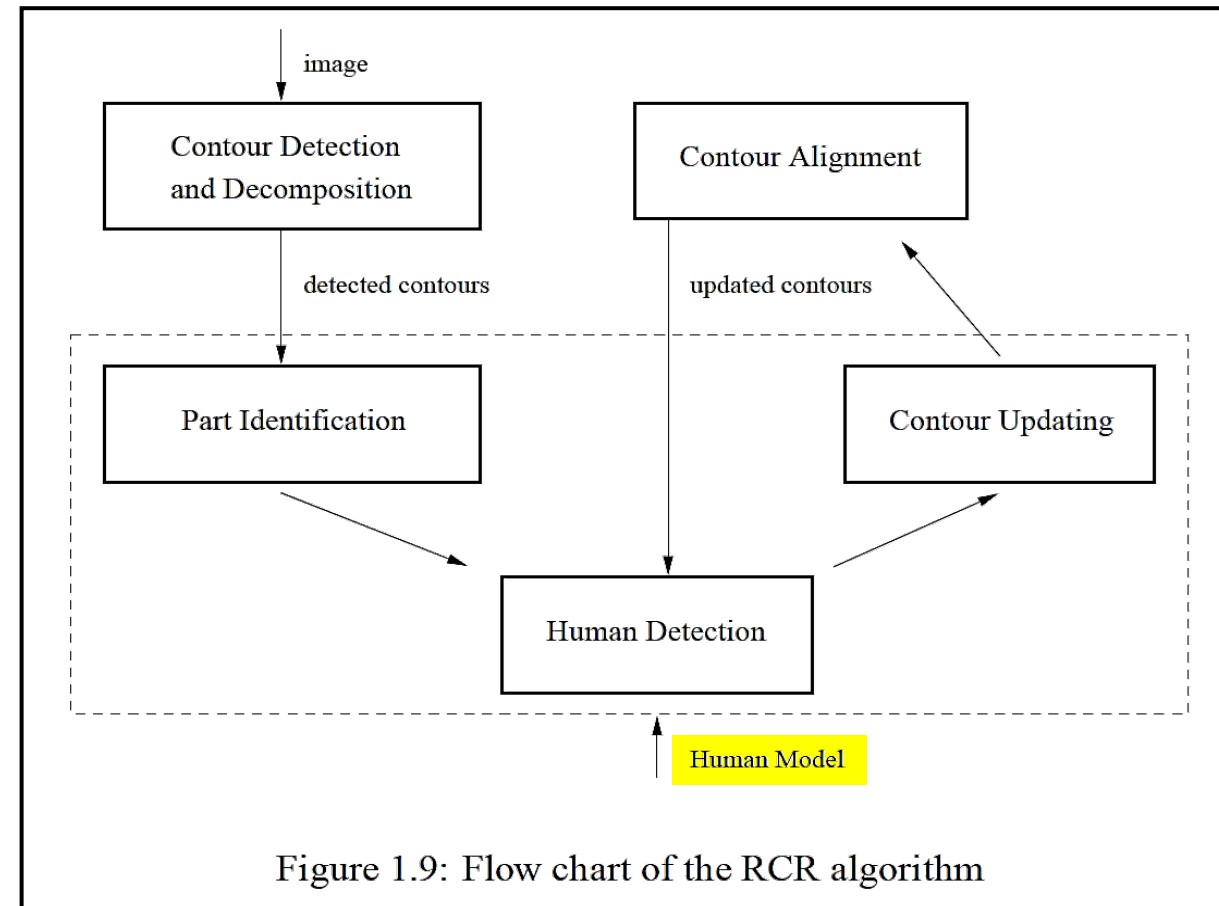
© 2001 Liang Zhao

This research was partially sponsored by PennDOT Agreement No. 62N111 Project
TA-34 executed under the Federal Technical Assistance Grant Program 001-008-701-99-7
and SAIC.

Meta 1006
p. 1

5.2.2 Experimental Results

The body parts of a person are located in a coarse-to-fine manner using the RCR algorithm.



Zhao's Object Detection Tested on Video

Dressed Human Modeling, Detection, and Parts Localization

Liang Zhao
CMU-RI-TR-01-19

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

July 26, 2001

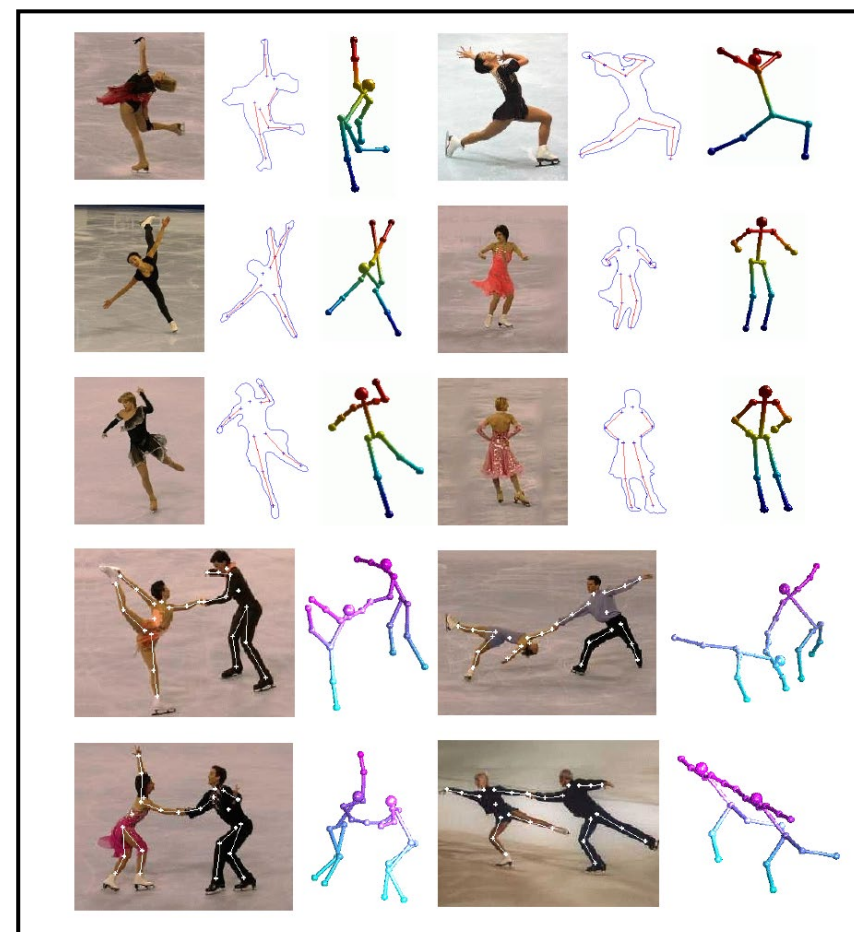
*Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy*

© 2001 Liang Zhao

This research was partially sponsored by PennDOT Agreement No. 62N111 Project
TA-34 executed under the Federal Technical Assistance Grant Program 001-008-701-99-7
and SAIC.

Meta 1006
p. 1

- ▶ Figure 5.7 demonstrates that covered body parts can still be located accurately using the dynamically assembled human model, the Bayesian similarity measure, and the coarse-to-fine localization approach.



Ground I: Obvious in View of Zhao

Claim 4, '980 Patent	
Preamble 4	The computer implemented method of claim 1,
Element 4[a]	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.

- ▶ Zhao teaches that motion information obtained from previous image-frames can be used to detect the object in a current image-frame.
- ▶ 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.

Zhao's Object Detection Is Based on Measurements from a Previous Frame

Dressed Human Modeling, Detection, and Parts Localization

Liang Zhao

CMU-RI-TR-01-19

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

July 26, 2001

*Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy*

© 2001 Liang Zhao

This research was partially sponsored by PennDOT Agreement No. 62N111 Project
TA-34 executed under the Federal Technical Assistance Grant Program 001-008-701-99-7
and SAIC.

Meta 1006
p. 1

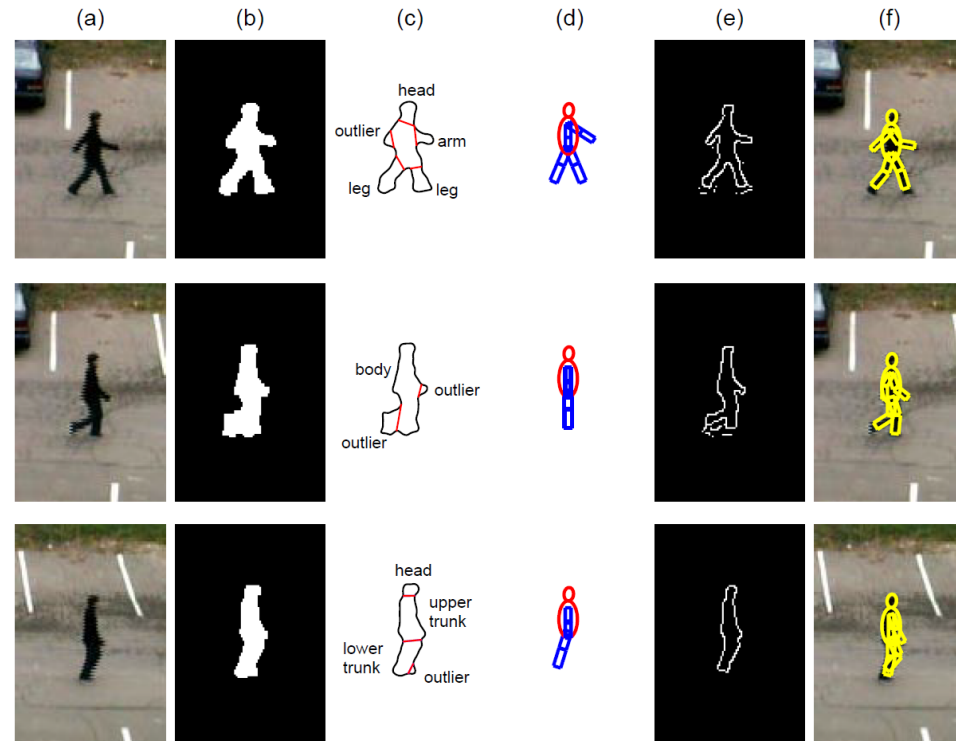


Figure 5.4: Body part localization: (a) images (b) foreground object detected from background subtraction (c) identified body parts (d) updated/predicted locations and outlines of the body parts (e) edge images (f) aligned body parts.

the first half cycle, no motion information is used to resolve the ambiguity with limbs' orientations, while in the second half of the cycle, the prediction from previous frames (constant angular velocity is assumed) is used to get better results of body part localization.

Using Measurements from Later Images Would Have Been Obvious

Generating Spatiotemporal Models from Examples

Adam Baumberg and David Hogg
School of Computer Studies
University of Leeds, Leeds LS2 9JT, U.K.
amb@scs.leeds.ac.uk

Abstract

Physically based vibration modes have been shown to provide a useful mechanism for describing non-rigid motions of articulated and deformable objects. The approach relies on assumptions being made about the elastic properties of an object to generate a compact set of orthogonal shape parameters which can then be used for tracking and data approximation. We present a method for automatically generating an equivalent physically based model using a training set of examples of the object deforming, tuning the elastic properties of the object to reflect how the object actually deforms. The resulting **model** provides a low dimensional shape description that **allows accurate temporal extrapolation based on the training motions**. Results are shown in which the method is applied to an automatically acquired training set of the outline of a walking pedestrian.

Hence we assume an observed set of matched, aligned shape vectors consisting of nodal (or point) positions observed over short intervals of time. e.g. a set of shape vectors $\mathbf{x}^{(k)}$ each consisting of n control points.

$$\mathbf{x}^{(0)} = (P_x^1, P_y^1, \dots, P_x^n, P_y^n)$$

with $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}$ observations of the nodes at time **$t = 0, \Delta t, 2\Delta t$** . From this data set a set of nodal displacements $\mathbf{u}^{(k)}$ is extracted by subtracting off the mean shape vector. The corresponding nodal velocities $\dot{\mathbf{u}}^{(k)}$ and nodal accelerations $\ddot{\mathbf{u}}^{(k)}$ are then calculated by finite difference approximations.

- It was well known to extrapolate from adjacent frames, and accordingly, determine probability that the object is in each image based on a prior image and a later image relative to the current image.

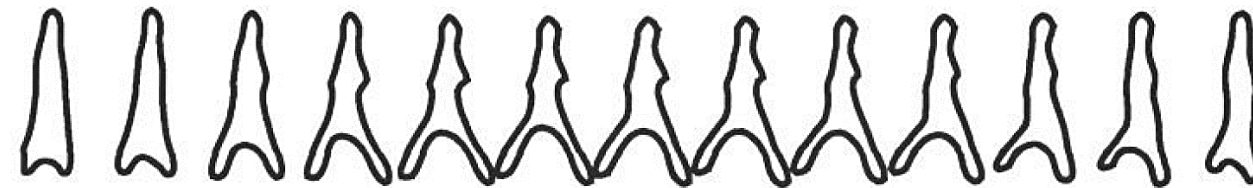
Using Measurements from Later Images Would Have Been Obvious

Petitioners' Expert

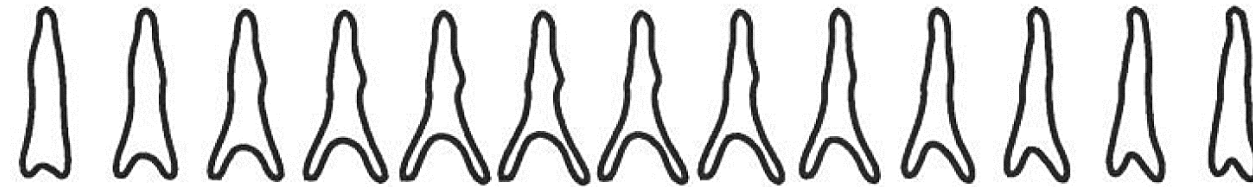


Dr. Chandra Bajaj

39. Figure 5 of Ex1011 shows the approximation of a spatio-temporal shape model walking sequence (bottom row (b)) $\mathbf{x}^{(k)}$, given the input video image sequence (top row (a)), through measurement estimations of shape displacements **in each image frame (k)**, and finite difference estimations of velocity $\dot{\mathbf{u}}^{(k)}$, acceleration $\ddot{\mathbf{u}}^{(k)}$ between 3 adjacent image frames $t = 0, \Delta t, 2\Delta t$, and with temporal consistency along the entire walking sequence.



(a) input walk sequence



(b) approximated sequence (24 parameters)

Using Measurements from Later Images Would Have Been Obvious

Attractive People: Assembling Loose-Limbed Models using Non-parametric Belief Propagation

Leonid Sigal
Department of Computer Science
Brown University
Providence, RI 02912
ls@cs.brown.edu

Benjamin H. Sigelman
Department of Computer Science
Brown University
Providence, RI 02912
bhsigelm@cs.brown.edu

Abstract

The detection and pose estimation of people in images is challenging by the variability of human appearance in natural scenes, and the high dimensionality of the problem. To cope with these problems we represent the human body as a graphical model in which the relationships between the body parts are represented by conditional probability distributions. We formulate the pose estimation problem as one of probabilistic inference over a graphical model where the random variables correspond to the individual limb parameters (position and orientation). Because the limbs are described by 6-dimensional vectors encoding pose in 3-space, discretization is impractical and the random variables in our model must be continuous-valued. To approximate belief propagation in such a graph we exploit a recently introduced generalization of the particle filter. This framework facilitates the automatic initialization of the body-model from low level cues and is robust to occlusion of body parts and scene clutter.

1 Introduction

Recent approaches to person detection and tracking exploit articulated body models in which the body is viewed as a kinematic tree in 2D [14], 2.5D [16, 23], or 3D [2, 5, 6, 19, 21] leading to a parametric state-space representation of roughly 25–35 dimensions. The high dimensionality of the resulting state-space has motivated the development of specialized stochastic search algorithms that either exploit the highly redundant dynamics of typical human motions [19], or use hierarchical sampling schemes to exploit the tree-structured nature of the model [5, 15]. These schemes have been effective for tracking people wearing increasingly complex clothing in increasingly complex cluttered backgrounds [21]. There are however a number of important shortcomings of these ap-

tracker. There are several advantages of this approach compared with traditional particle filtering: the complexity of the search task is linear rather than exponential in the number of body parts; bottom-up initialization information can be incorporated in every frame; and forward-backward smoothing, either over a time-window or an entire sequence, is straightforward.

- ▶ The '980 inventors admit that using measurements from adjacent frames, including previous and later frames, was straightforward.

Meta 1015
p. 1

Ground I: Obvious in View of Zhao

Claim 8, '980 Patent	
Preamble 8	The computer implemented method of claim 1,
Element 8[a]	wherein a joint probability distribution for the spatio-temporal model with N components is: $P(X_0^O, X_0^{C_0}, X_0^{C_1}, \dots, X_0^{C_N}, \dots, X_T^O, X_T^{C_0}, X_T^{C_1}, \dots, X_T^{C_N}) =$ $\frac{1}{Z} \prod_{ij} \psi_{ij}(X_i^O, X_j^O)$ $\prod_{ik} \psi_{ik}(X_i^O, X_i^{C_k}) \prod_{ikl} \psi_{kl}(X_i^{C_k}, X_i^{C_l}) \prod_i \phi_i(X_i^O) \prod_{ik} \phi_i(X_i^{C_k})$

- ▶ The joint probability used by Zhao is representable in the form of the joint probability distribution formulated by claim 8. EX1004, ¶121.
- ▶ The '980 joint probability distribution represents a typical factorized probability distribution P(...) for a graphical model G of an object O.

Zhao's Joint Probability Is Representable in the Form of Claim 8

- ▶ Zhao's joint probability for a graphical model of a human body:

The conditional probability $P(\text{person}|H)$ 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. According to the above observation, the following formula is designed to estimate $P(\text{person}|H)$:

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

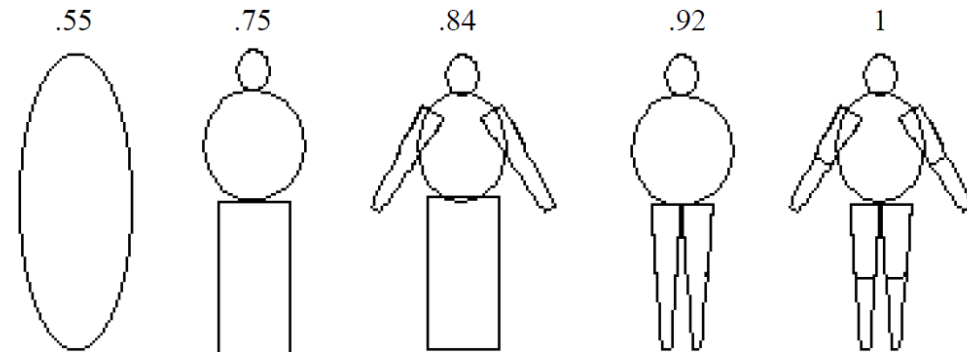


Figure 3.1: Examples of the calculated $P(\text{person}|H)$ given the degeneration factor $\alpha = .9$ and the contour parts matching the model parts exactly

Dressed Human Modeling, Detection, and Parts Localization

Liang Zhao
CMU-RI-TR-01-19

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

July 26, 2001

*Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy*

© 2001 Liang Zhao

This research was partially sponsored by PennDOT Agreement No. 62N111 Project
TA-34 executed under the Federal Technical Assistance Grant Program 001-008-701-99-7
and SAIC.

Meta 1006
p. 1

Zhao's Joint Probability Is Representable in the Form of Claim 8

TO APPEAR AT 2003 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION

Nonparametric Belief Propagation

Erik B. Sudderth, Alexander T. Ihler, William T. Freeman, and Alan S. Willsky
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
esuddert@mit.edu, ihler@mit.edu, wtf@ai.mit.edu, willsky@mit.edu

Abstract

In many applications of graphical models arising in computer vision, the hidden variables of interest are most naturally specified by continuous, non-Gaussian distributions. There exist inference algorithms for discrete approximations to these continuous distributions, but for the high-dimensional variables typically of interest, discrete inference becomes infeasible. Stochastic methods such as particle filters provide an appealing alternative. However, existing techniques fail to exploit the rich structure of the graphical models describing many vision problems.

Drawing on ideas from regularized particle filters and belief propagation (BP), this paper develops a nonparametric belief propagation (NBP) algorithm applicable to general graphs. Each NBP iteration uses an efficient sampling procedure to update kernel-based approximations to the true, continuous likelihoods. The algorithm can accommodate an extremely broad class of potential functions, including nonparametric representations. Thus, NBP extends particle filtering methods to the more general vision problems that graphical models can describe. We apply the NBP algorithm to infer component interrelationships in a part-based face model, allowing location and reconstruction of occluded features.

1. Introduction

Graphical models provide a powerful, general framework for developing statistical models of computer vision problems [7, 8, 10]. However, graphical formulations are only useful when combined with efficient algorithms for inference and learning. Computer vision problems are particularly challenging because they often involve high-dimensional, continuous variables and complex, multimodal distributions. For example, the articulated models used in many tracking applications have dozens of degrees of freedom to be estimated at each time step [17]. Realistic graphical models for these problems must represent outliers, bimodalities, and other non-Gaussian statistical features. The corresponding optimal inference procedures for these models typically involve integral equations for which no closed form solution exists. Thus, it is necessary to de-

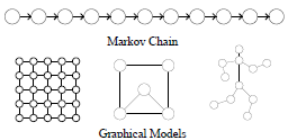


Figure 1. Particle filters assume variables are related by a Markov chain. The NBP algorithm extends particle filtering techniques to arbitrarily structured graphical models.

velop families of approximate representations, and corresponding methods for updating those approximations.

The simplest method for approximating intractable continuous-valued graphical models is discretization. Although exact inference in general discrete graphs is NP hard, approximate inference algorithms such as loopy belief propagation (BP) [16, 21, 23, 24] produce excellent empirical results in many cases. Certain vision problems, including stereo vision [20], are well suited to discrete formulations. For problems involving high-dimensional variables, however, exhaustive discretization of the state space is intractable. In some cases, domain-specific heuristics may be used to dynamically exclude those configurations which appear unlikely based upon the local evidence [2, 7]. In more challenging applications, however, the local evidence at some nodes may be inaccurate or misleading, and these approximations will heavily distort the resulting estimates.

For temporal inference problems, particle filters [5, 10] have proven to be an effective, and influential, alternative to discretization. They provide the basis for several of the most effective visual tracking algorithms [15, 17]. Particle filters approximate conditional densities nonparametrically as a collection of representative elements. Although it is possible to update these approximations deterministically using local linearizations [1], most implementations use Monte Carlo methods to stochastically update a set of weighted point samples. The stability and robustness of particle filters can often be improved by regularization methods [5, Chapter 12] in which smoothing kernels [18] explicitly represent the uncertainty associated with each point sample.

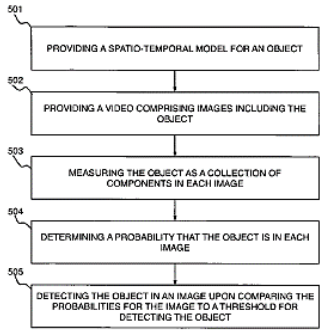
Meta 1012
p. 1

2. Undirected Graphical Models

An undirected graph $\mathcal{G}$ is defined by a set of nodes $\mathcal{V}$, and a corresponding set of edges $\mathcal{E}$. The *neighborhood* of a node $s \in \mathcal{V}$ is defined as $\Gamma(s) \triangleq \{t \mid (s, t) \in \mathcal{E}\}$. Graphical models associate each node $s \in \mathcal{V}$ with an unobserved, or hidden, random variable x_s , as well as a noisy local observation y_s . Let $x = \{x_s \mid s \in \mathcal{V}\}$ and $y = \{y_s \mid s \in \mathcal{V}\}$ denote the sets of all hidden and observed variables, respectively. For simplicity, we consider models with **pairwise potential functions**, for which **$p(x, y)$ factorizes as**

$$p(x, y) = \frac{1}{Z} \prod_{(s,t) \in \mathcal{E}} \psi_{s,t}(x_s, x_t) \prod_{s \in \mathcal{V}} \psi_s(x_s, y_s) \quad (1)$$

- ▶ Sudderth teaches the same joint probabilistic distribution formulas of claims 8/16, using the same belief propagation approach as the '980 Patent.
- ▶ A POSITA would have understood that the joint probability distribution of Zhao would be representable in the form recited by claim 8. EX1052, ¶¶121 and 122.

 US007436980B2	
(12) United States Patent Sigal et al.	(10) Patent No.: US 7,436,980 B2 (45) Date of Patent: Oct. 14, 2008
(54) GRAPHICAL OBJECT MODELS FOR DETECTION AND TRACKING	7,200,266 B2* 4/2007 Ozer et al. 382/173 FOREIGN PATENT DOCUMENTS 1 217 572 A2 6/2002 OTHER PUBLICATIONS M. Isard, Pumpas: Real-valued graphical models for computer vision, CVPR, vol. 1, pp. 613-620, 2003. K. Okuma, A. Teleghani, N. de Freitas, J. Little and D. Lowe: A boosted particle filter: Multi-target detection and tracking, ECCV, 2004. L. Sigal, S. Bhatia, S. Roth, M. Black and M. Isard: Tracking loose-limbed people, CVPR, 2004. E. Sudderth, A. Ihler, W. Freeman and A. Willsky: Nonparametric belief propagation, MIT Laboratory for Information and Decision Systems Technical Report P-2551 (Oct. 15, 2002). P. Viola and M. Jones: Rapid object detection using a boosted cascade of simple features, CVPR, 2001. P. Viola, M. Jones and D. Snow: Detecting pedestrians using patterns of motion and appearance, ICCV, pp. 734-741, 2003. (Continued) Primary Examiner—Tom Y Lu (74) Attorney, Agent, or Firm—Donald B. Paschburg, F. Chau & Associates, LLC
(75) Inventors: Leonid Sigal , Providence, RI (US); Ying Zhu , Monmouth Junction, NJ (US); Dorin Comaniciu , Princeton Jct., NJ (US)	
(73) Assignee: Siemens Corporate Research, Inc. , Princeton, NJ (US)	
(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 691 days.	
(21) Appl. No.: 11/135,210	
(22) Filed: May 23, 2005	
(65) Prior Publication Data US 2005/0286738 A1 Dec. 29, 2005	
Related U.S. Application Data	
(60) Provisional application No. 60/574,799, filed on May 27, 2004.	
(51) Int. Cl. G06K 9/00 (2006.01)	
(52) U.S. Cl. 382/103	
(58) Field of Classification Search 382/103, 382/173	
See application file for complete search history.	
(56) References Cited U.S. PATENT DOCUMENTS 6,597,801 B1 7/2003 Cham et al.	
16 Claims, 5 Drawing Sheets	
 <pre>graph TD 501[PROVIDING A SPATIO-TEMPORAL MODEL FOR AN OBJECT] --> 502[PROVIDING A VIDEO COMPRISING IMAGES INCLUDING THE OBJECT] 502 --> 503[MEASURING THE OBJECT AS A COLLECTION OF COMPONENTS IN EACH IMAGE] 503 --> 504[DETERMINING A PROBABILITY THAT THE OBJECT IS IN EACH IMAGE] 504 --> 505[DETECTING THE OBJECT IN AN IMAGE UPON COMPARING THE PROBABILITIES FOR THE IMAGE TO A THRESHOLD FOR DETECTING THE OBJECT]</pre>	

1. Background

2. Claim Construction

- a. 1/9[a]: spatio-temporal model
- b. 1/9[c]: measuring the object as a collection of components
- c. 1/9[e]: comparing the probabilities for each image

3. Grounds

- a. Claims 1-16: Obvious in view of Zhao
- b. Claims 1-16: Obvious over Steffens in view of Zhao**
- c. Claims 1-16: Obvious over Ozer in view of Zhao
- d. Claims 1-16: Obvious in view of TLP

Prior Art - Steffens (EX1007)

(12) **United States Patent**
Steffens et al.

(10) Patent No.: **US 6,301,370 B1**
(45) Date of Patent: **Oct. 9, 2001**

(54) **FACE RECOGNITION FROM VIDEO IMAGES**

(75) Inventors: **Johannes Bernhard Steffens**, Culver City; **Egor Valerievich Elagin**, Los Angeles; **Laciano Pasquale Agostino Nocera**, Los Angeles; **Thomas Maurer**, Los Angeles; **Hartmut Neven**, Santa Monica, all of CA (US)

(73) Assignee: **Eyematic Interfaces, Inc.**, Inglewood, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/206,195**

(22) Filed: **Dec. 4, 1998**

Related U.S. Application Data

(60) Provisional application No. 60/081,615, filed on Apr. 13, 1998.

(51) Int. Cl.⁷ **G06K 9/00; G06K 9/36; G06K 9/62; H04N 5/225; G01S 13/00**

(52) U.S. Cl. **382/103; 382/117; 382/118; 382/190; 382/209; 382/276; 342/290; 348/169**

(58) Field of Search **382/117, 118, 382/168, 181, 190, 195, 199, 103, 209, 216, 219, 279, 282, 115, 100, 276; 348/169, 342/90**

References Cited

U.S. PATENT DOCUMENTS

4,725,824 2/1988 Yoshioka 340575
4,805,234 2/1989 Kozuka et al. 382/141

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

44 06 020 C1 6/1995 (DE) G06K/962

0 807 902 A2 11/1997 (EP) .

OTHER PUBLICATIONS

Sara, R. et al "3-D Data Acquisition and Interpretation for Virtual Reality and Telepresence", Proceedings IEEE Workshop Computer Vision for Virtual Reality Based Human Communication, Bombay, Jan. 1998, 7 pp.
Sara, R. et al "Fish-Scales: Representing Fuzzy Manifolds", Proceedings International Conference Computer Vision, ICCV '98, pp. 811-817, Bombay, Jan. 1998.
Wiskott, L. "Phantom Faces for Face Analysis", *Pattern Recognition*, vol. 30, No. 6, pp. 637-646, 1997.
Wurtz, R. "Object Recognition Robust Under Translations, Deformations, and Changes in Background", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, No. 7, Jul. 1997, pp. 769-775.
Akimoto, T. et al. "Automatic Creation of 3-D Facial Models", *IEEE Computer Graphics & Applications*, pp. 16-22, Sep. 1993.

(List continued on next page.)

Primary Examiner—Leo Boudreau

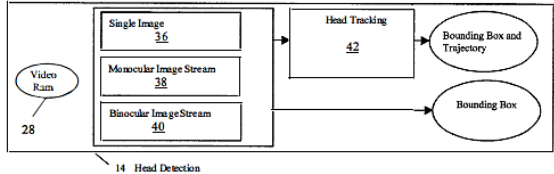
Assistant Examiner—Daniel G. Mariani

(74) **Attorney, Agent, or Firm**—Gray Cary Ware Freidenrich

ABSTRACT

(57) The present invention is embodied in an apparatus, and related method, for detecting and recognizing an object in an image frame. The object may be, for example, a head having particular facial characteristics. The object detection process uses robust and computationally efficient techniques. The object identification and recognition process uses an image processing technique based on model graphs and bunch graphs that efficiently represent image features as jets. The jets are composed of wavelet transforms and are processed at nodes or landmark locations on an image corresponding to readily identifiable features. The system of the invention is particularly advantageous for recognizing a person over a wide variety of pose angles.

42 Claims, 21 Drawing Sheets



Meta 1007
p. 1

- ▶ Steffens was filed December 4, 1998, and issued on October 9, 2001
- ▶ Primary reference in Ground II

Ground II: Steffens in View of Zhao

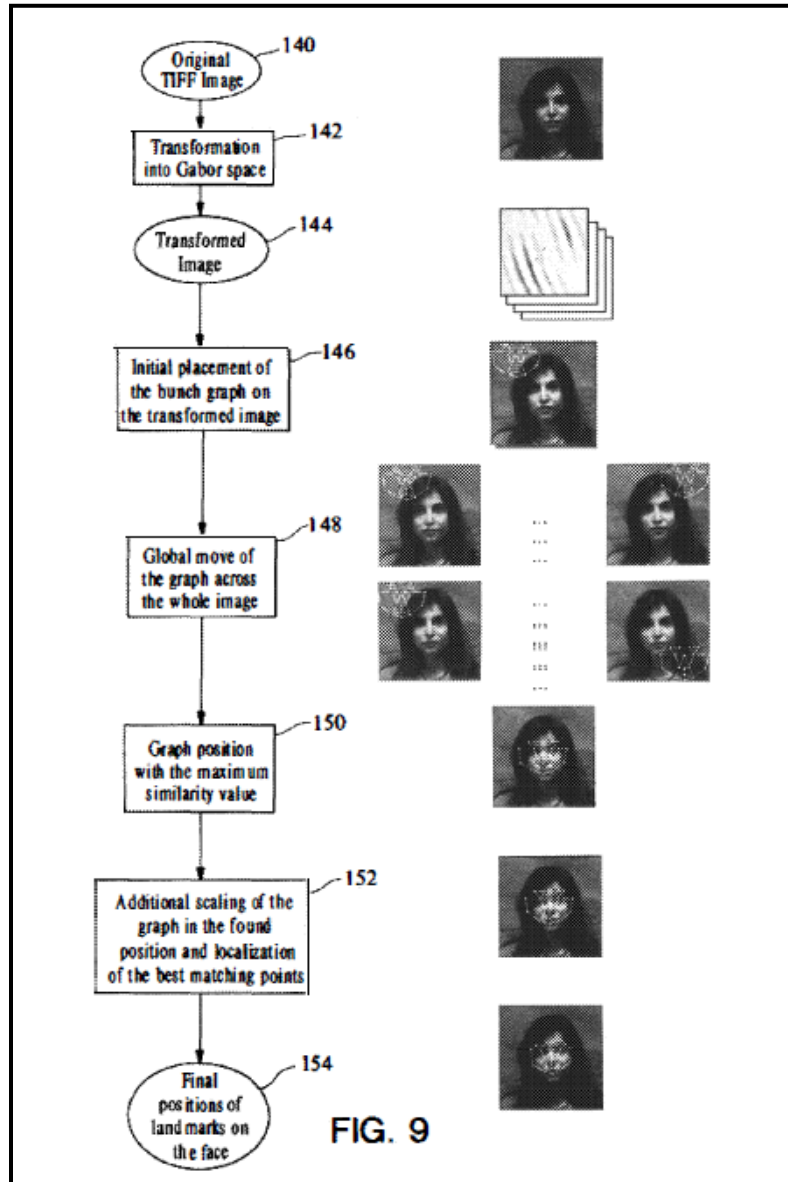
Claim 1, '980 Patent

Preamble [1]	A computer implemented method for object detection comprising:
Element 1[a]	providing a spatio-temporal model for an object to be detected;
Element 1[b]	providing a video comprising a plurality of images including the object;
Element 1[c]	measuring the object as a collection of components in each image;
Element 1[d]	determining a probability that the object is in each image; and
Element 1[e]	detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.

- ▶ Both Steffens and Zhao use a spatio-temporal model and **similarity measures** to detect an object in a video. Petition at 8.
- ▶ Steffens does not expressly teach that its similarity measures include **probabilistic frameworks** for identifying objects. Petition at 36.
- ▶ A POSITA would have been motivated to use **Zhao's probabilistic "Bayesian Similarity Measure"** in Steffens. *Id.*

Steffens Detects an Object Using Similarity Measures

- ▶ Steffens describes “identifying the object based on a similarity” between an image and a corresponding model.



The facial landmarks and features of the head may be located using an elastic graph matching technique shown in FIG. 9. In the elastic graph matching technique, a captured image (block 140) is transformed into Gabor space using a wavelet transformation (block 142) which is described below in more detail with respect to FIG. 10. The transformed image (block 144) is represented by 40 complex values, representing wavelet components, per each pixel of the original image. Next, a rigid copy of a model graph, which is described in more detail below with respect to FIG. 12, is positioned over the transformed image at varying model node positions to locate a position of optimum similarity (block 146). The search for the optimum similarity may be performed by positioning the model graph in the upper left hand corner of the image, extracting the jets at the nodes, and determining the similarity between the image graph and the model graph. The search continues by sliding

Zhao Uses a Probabilistic Similarity Measure

Dressed Human Modeling, Detection, and Parts Localization

Liang Zhao
CMU-RI-TR-01-19

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

July 26, 2001

*Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy*

© 2001 Liang Zhao

This research was partially sponsored by PennDOT Agreement No. 62N111 Project
TA-34 executed under the Federal Technical Assistance Grant Program 001-008-701-99-7
and SAIC.

Meta 1006
p. 1

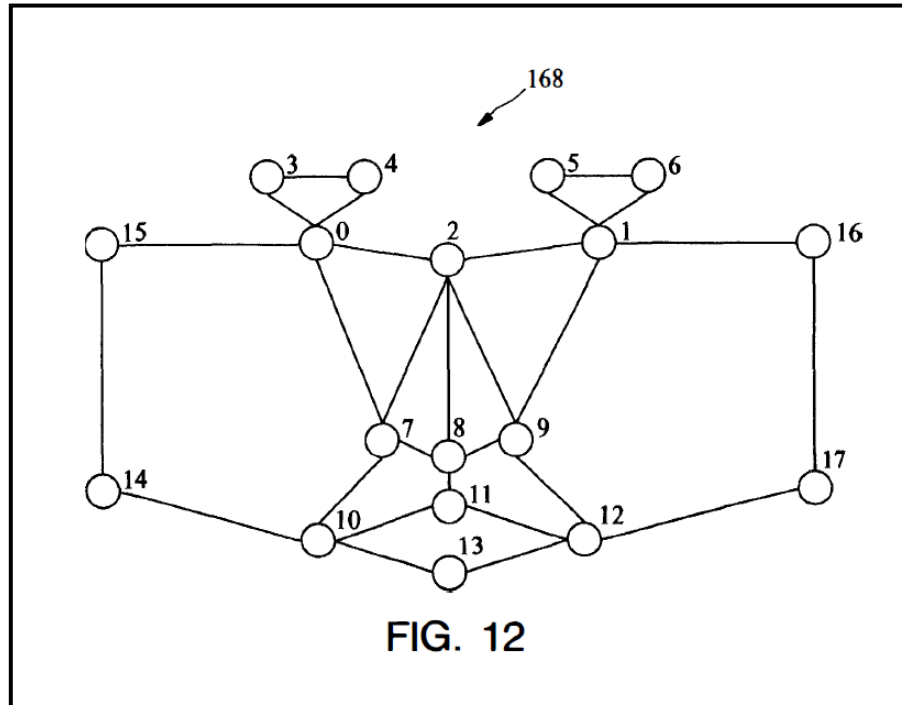
- ▶ Zhao expressly describes a probabilistic similarity measure.
- ▶ Like Steffens, Zhao describes a method where an object is identified based on a similarity between an image and a corresponding model.

3.3 Bayesian Similarity Measure and Body Part Identification

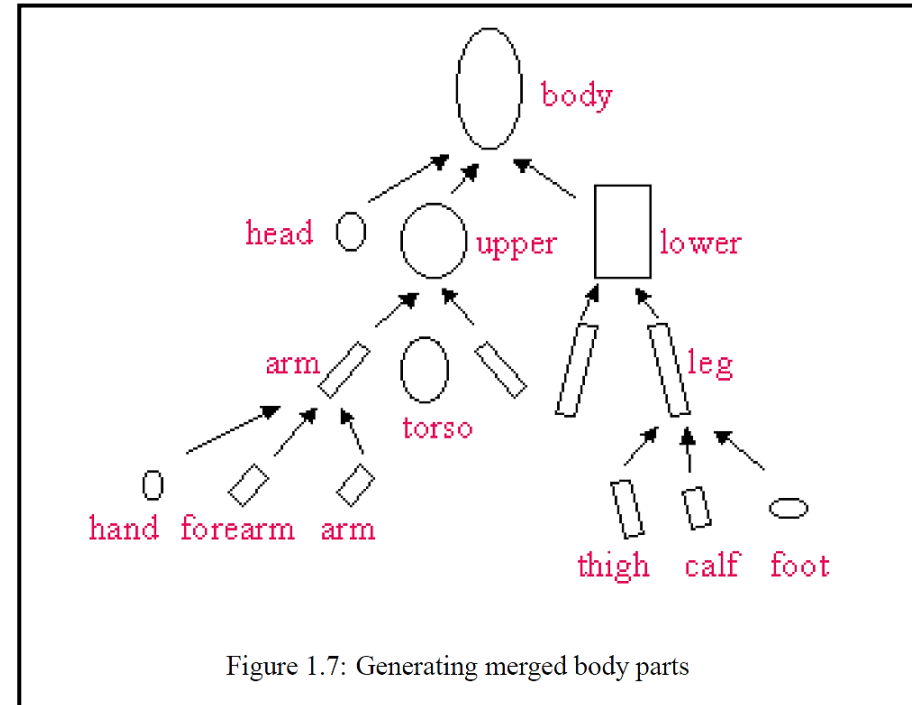
A part-based similarity measure evaluates the resemblance between a contour and a model based on the best match between their body parts. This section presents the definition of a Bayesian Similarity Measure derived from the probabilistic human model, and illustrates how to use this measure to guide the matching between a set of decomposed ribbons and a set of model body parts through coarse-to-fine model assembling.

Steffens Does Not Employ a Fundamentally Different Framework than Zhao

Steffens' multi-component, graphical model



Zhao's multi-component, graphical model



- ▶ Although Steffens focuses more on the human face for facial recognition, both Steffens and Zhao concern object detection in video images based on spatio-temporal models and similarity matching.

Dr. Saber's Distinction Between Landmarks and Components is Artificial

Patent Owner's Expert Witness



Dr. Eli Saber

Q If the face were considered the object, are you saying the tip of the nose would not be a component of the face?

A So component has a certain meaning in the '980. In the context of the '980, component has a certain meaning. The word component, these things in the Steffens are all landmarks and features. They would not be described as components of the face.

And because the word component in the '980 has a certain context features. Steffens describes these as features and landmarks. The word component doesn't show up in Steffens in his description.

Steffens' Similarity Measure Corresponds to a Measure of Probability

Petitioners' Expert



Dr. Chandra Bajaj

52. As I explained in my First Declaration (¶144), Steffens uses a similarity measure to match and determine whether an object is in an image. Steffens describes two different similarity functions that are both normalized so take on real similarity values ranging between $[0, 1]$ with 0 being not-similar and 1 being exactly similar, and real values between them giving a continuous measure function of similarity. As a POSITA would know this corresponds to a measure of probability of similarity between two entities. Ex1007 at 9:47-65.

Steffens' Similarity Measure is Closely Related to Probability

Patent Owner's
Expert Witness



Dr. Eli Saber

A Whether it's normalized or not, it doesn't matter. The dot product between two vectors is essentially how much of Vector 1 is represented or rides along Vector 2. So it provides a sense of similarity between the vectors.

Ground II: Steffens in View of Zhao

Claim 4, '980 Patent

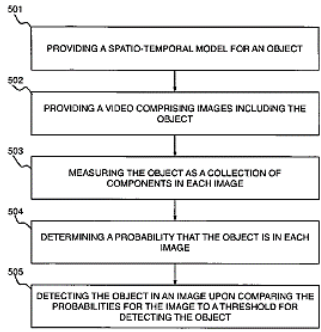
Preamble 4

The computer implemented method of claim 1,

Element 4[a]

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.

- ▶ Steffens teaches tracking an object between image frames based on a trajectory associated with the object. Petition at 48; EX1007, 2:6-13 and Claims 14, 16, and 17.
- ▶ PO acknowledges that Steffens' object tracking process operates between frames in the forward direction. PO sur-reply at 32.
- ▶ It would have been obvious to extrapolate from later images as well. See slides 48-52 above.

 US007436980B2	
(12) United States Patent Sigal et al.	(10) Patent No.: US 7,436,980 B2 (45) Date of Patent: Oct. 14, 2008
(54) GRAPHICAL OBJECT MODELS FOR DETECTION AND TRACKING	7,200,266 B2* 4/2007 Ozer et al. 382/173 FOREIGN PATENT DOCUMENTS
(75) Inventors: Leonid Sigal , Providence, RI (US); Ying Zhu , Monmouth Junction, NJ (US); Dorin Comaniciu , Princeton Jct., NJ (US)	EP 1 217 572 A2 6/2002 OTHER PUBLICATIONS
(73) Assignee: Siemens Corporate Research, Inc. , Princeton, NJ (US)	M. Isard, Pumpas: Real-valued graphical models for computer vision, CVPR, vol. 1, pp. 613-620, 2003. K. Okuma, A. Teleghani, N. de Freitas, J. Little and D. Lowe. A boosted particle filter: Multi-target detection and tracking, ECCV, 2004. L. Sigal, S. Bhatia, S. Roth, M. Black and M. Isard. Tracking loose-limbed people, CVPR, 2004. E. Sudderth, A. Ihler, W. Freeman and A. Willsky. Nonparametric belief propagation, MIT Laboratory for Information and Decision Systems Technical Report P-2551 (Oct. 15, 2002). P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features, CVPR, 2001. P. Viola, M. Jones and D. Snow. Detecting pedestrians using patterns of motion and appearance, ICCV, pp. 734-741, 2003.
(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 691 days.	(Continued)
(21) Appl. No.: 11/135,210	Primary Examiner—Tom Y Lu (74) Attorney, Agent, or Firm—Donald B. Paschburg, F. Chau & Associates, LLC
(22) Filed: May 23, 2005	
(65) Prior Publication Data US 2005/0286738 A1 Dec. 29, 2005	
Related U.S. Application Data	
(60) Provisional application No. 60/574,799, filed on May 27, 2004.	(57) ABSTRACT
(51) Int. Cl. G06K 9/00 (2006.01)	A computer implemented method for object detection includes providing a spatio-temporal model for an object to be detected, providing a video including a plurality of images including the object, and measuring the object as a collection of components in each image. The method further includes determining a probability that the object is in each image, and detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.
(52) U.S. Cl. 382/103	
(58) Field of Classification Search 382/103, 382/173	
See application file for complete search history.	
(56) References Cited U.S. PATENT DOCUMENTS 6,597,801 B1 7/2003 Cham et al.	
16 Claims, 5 Drawing Sheets	
 <pre>graph TD 501[PROVIDING A SPATIO-TEMPORAL MODEL FOR AN OBJECT] --> 502[PROVIDING A VIDEO COMPRISING IMAGES INCLUDING THE OBJECT] 502 --> 503[MEASURING THE OBJECT AS A COLLECTION OF COMPONENTS IN EACH IMAGE] 503 --> 504[DETERMINING A PROBABILITY THAT THE OBJECT IS IN EACH IMAGE] 504 --> 505[DETECTING THE OBJECT IN AN IMAGE UPON COMPARING THE PROBABILITIES FOR THE IMAGE TO A THRESHOLD FOR DETECTING THE OBJECT]</pre>	

1. Background

2. Claim Construction

- a. 1/9[a]: spatio-temporal object
- b. 1/9[c]: measuring the object as a collection of components
- c. 1/9[e]: comparing the probabilities for each image

3. Grounds

- a. Claims 1-16: Obvious in view of Zhao
- b. Claims 1-16: Obvious over Steffens in view of Zhao
- c. **Claims 1-16: Obvious over Ozer in view of Zhao**
- d. Claims 1-16: Obvious in view of TLP

Ground III: Ozer in View of Zhao

Claim 1, '980 Patent

Preamble [1]	A computer implemented method for object detection comprising:
Element 1[a]	providing a spatio-temporal model for an object to be detected;
Element 1[b]	providing a video comprising a plurality of images including the object;
Element 1[c]	measuring the object as a collection of components in each image;
Element 1[d]	determining a probability that the object is in each image; and
Element 1[e]	detecting the object in any image upon <i>comparing the probabilities for each image to a threshold</i> for detecting the object.

- ▶ For element 1/9[e], PO relies on its Zhao arguments (Ground I), but makes no additional argument. POR, 50.
- ▶ PO's contention that element 1/9[e] is not rendered obvious by Ozer in view of Zhao is based on a flawed and impermissibly narrow claim construction.

Ground III: Ozer in View of Zhao

Claim 4, '980 Patent

Preamble 4

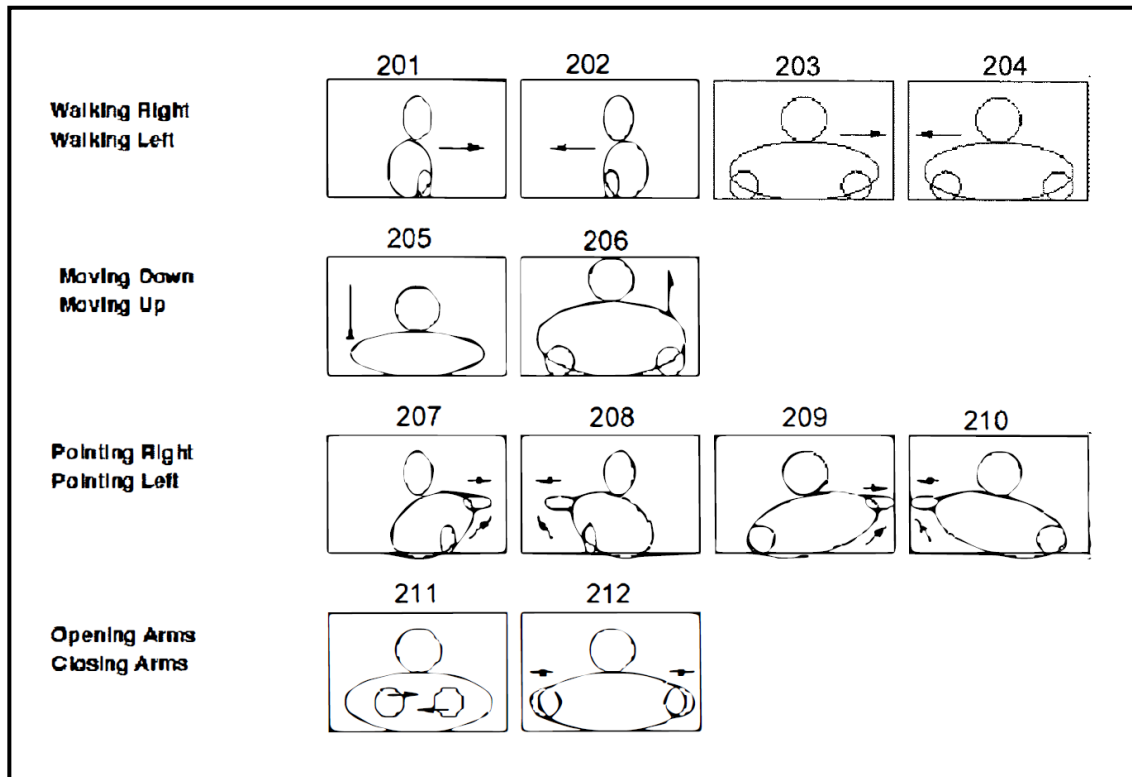
The computer implemented method of claim 1,

Element 4[a]

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.

- ▶ Ozer's "activity" classification encompasses object tracking over multiple frames. Petitioner Reply at 19.

Ozer's "activity" Classification Encompasses Object Tracking over Multiple Frames



Classifying Over Multiple Frames

Output of the graph-matching algorithm is the classified object parts. The movements of the object parts are described as a spatio-temporal sequence of feature vectors that consist of the direction of the object part movement. 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). The activity of each part is then combined by a quadratic distance classifier to find the overall activity of the object of interest. As shown in FIG. 2, frame 201 shows walking right and frame 202 shows walking left.

Ozer Uses Probabilistic Hidden Markov Models for Its Inference of the Motion

Petitioners' Expert



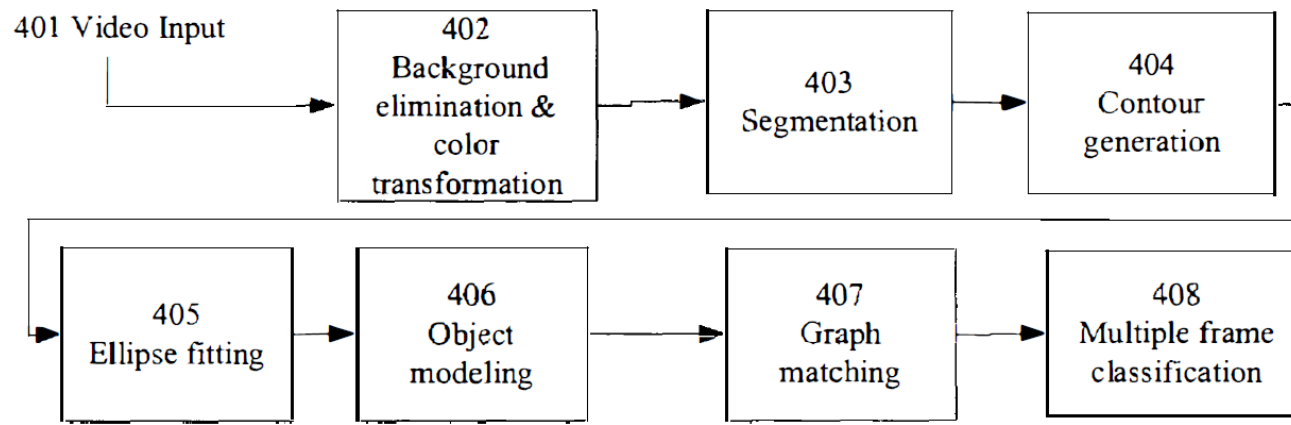
Dr. Chandra Bajaj

60. In light of this passage from Ozer, among others, a POSITA would have understood that Ozer's "activity" classification encompasses object motion tracking over multiple frames to classify or discriminate between various motion trajectories. Ozer teaches both the detection of the object parts and the movements (motions) of the object parts and moreover uses probabilistic Hidden Markov models for its inference of the motion (*e.g.*, walking activity) from video of the spatio-temporal model.

Ozer Uses Probabilistic Hidden Markov Models for Its Inference of the Motion

The **system** of the present invention can **detect non-rigid** (e.g. human body) and rigid object **parts** and recognize their activities in compressed and uncompressed domains. To

Figure 4



- ▶ Non-rigid object detection infers that the object is moving.
- ▶ Figure 4 illustrates a video input involving multiple frames. See 408 Multiple frame classification.

Ozer Teaches Object Tracking Over Multiple Frames

Patent Owner's Expert Witness



Dr. Eli Saber

Q So what does that mean, that those are tracked objects?

A Tracks only objects that fit the model of the user defined subject. So it's not going to track -- for instance, if I'm interested in humans and humans walking their dog, I'm not going to track a car driving down the road.

So I'm tracking objects that I'm specifically interested in. You know, that's why the next sentence says: This makes the system more accurate because it easily rejects many elements in the scene that may be moving but are not objects of interest.

I have a deer walking in the backyard. I'm not interested in deer. I just want to know if somebody's walking around my backyard for security purposes.

Ozer Tracks An Object by Using Ellipse Parameters For Consecutive Frames

Petitioners' Expert



Dr. Chandra Bajaj

63. Further, Ozer teaches that after the detection of the object parts, the system is ready to recognize the activities of each object part and the overall activity of the object. *Id.* at 4:35-37. Ozer “tracks only objects that fit the model of the user defined subject, such as human, any rigid object, dog.” *Id.* at 6:23-24. Ozer explains that “matched nodes” identified using the graph-matching algorithm are “tracked by using ellipse parameters for the consecutive frames.” *Id.* at 7:25-31.

- ▶ To the extent that it was not explicitly taught by Ozer and/or Zhao, it would have been obvious to detect an object using measurements from later image frames. See slides 48-52.

Overview

US007436980B2

(12) **United States Patent**
Sigal et al.

(10) **Patent No.:** **US 7,436,980 B2**
(45) **Date of Patent:** **Oct. 14, 2008**

(54) **GRAPHICAL OBJECT MODELS FOR
DETECTION AND TRACKING**

7,200,266 B2 * 4/2007 Ozer et al. 382/173

FOREIGN PATENT DOCUMENTS

(75) Inventors: **Leonid Sigal**, Providence, RI (US); **Ying Zhu**, Monmouth Junction, NJ (US);
Dorin Comaniciu, Princeton Jct., NJ
(US)

EP 1 217 572 A2 6/2002

OTHER PUBLICATIONS

(73) Assignee: **Siemens Corporate Research, Inc.**,
Princeton, NJ (US)

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 691 days.

M. Isard, P. Pong, Real-valued graphical models for computer
vision, CVPR, vol. 1, pp. 613-620, 2003.

K. Okuma, A. Tachibana, N. de Freitas, J. Little and D. Lowe. A
boosted particle filter: Multi-target detection and tracking. ECCV,
2004.

L. Sigal, S. Bhatia, S. Roth, M. Black and M. Isard. Tracking loose-
ly-labeled people, CVPR, 2004.

E. Sudderth, A. Ihler, W. Freeman and A. Willsky. Nonparametric
belief propagation. MIT Laboratory for Information and Decision
Systems Technical Report P-2551 (Oct. 15, 2002).

P. Viola and M. Jones. Rapid object detection using a boosted cascade
of simple features, CVPR, 2001.

P. Viola, M. Jones and D. Snow. Detecting pedestrians using patterns
of motion and appearance, ICCV, pp. 734-741, 2003.

(21) Appl. No.: **11/135,210**

(22) Filed: **May 23, 2005**

(65) **Prior Publication Data**

US 2005/0286738 A1 Dec. 29, 2005

(Continued)

Related U.S. Application Data

(60) Provisional application No. 60/574,799, filed on May
27, 2004.

Primary Examiner—Tom Y Lu

(74) Attorney, Agent, or Firm—Donald B. Paschburg; F. Chau
& Associates, LLC

(51) Int. Cl.
G06K 9/00 (2006.01)

(52) U.S. Cl. **382/103**

(58) **Field of Classification Search** **382/103,**
382/173

See application file for complete search history.

(57) **ABSTRACT**

A computer implemented method for object detection
includes providing a spatio-temporal model for an object to
be detected, providing a video including a plurality of images
including the object, and measuring the object as a collection
of components in each image. The method further includes
determining a probability that the object is in each image, and
detecting the object in any image upon comparing the prob-
abilities for each image to a threshold for detecting the object.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,597,801 B1 7/2003 Cham et al.

16 Claims, 5 Drawing Sheets

```
graph TD
    501[PROVIDING A SPATIO-TEMPORAL MODEL FOR AN OBJECT] --> 502[PROVIDING A VIDEO COMPRISING IMAGES INCLUDING THE OBJECT]
    502 --> 503[MEASURING THE OBJECT AS A COLLECTION OF COMPONENTS IN EACH IMAGE]
    503 --> 504[DETERMINING A PROBABILITY THAT THE OBJECT IS IN EACH IMAGE]
    504 --> 505[DETECTING THE OBJECT IN AN IMAGE UPON COMPARING THE PROBABILITIES FOR THE IMAGE TO A THRESHOLD FOR DETECTING THE OBJECT]
```

1. Background

2. Claim Construction

- a. 1/9[a]: spatio-temporal model
- b. 1/9[c]: measuring the object as a collection of components
- c. 1/9[e]: comparing the probabilities for each image

3. Grounds

- a. Claims 1-16: Obvious in view of Zhao
- b. Claims 1-16: Obvious over Steffens in view of Zhao
- c. Claims 1-16: Obvious over Ozer in view of Zhao
- d. Claims 1-16: Obvious in view of TLP**

Prior Art - TLP (EX1009)

Tracking Loose-limbed People

Leonid Sigal* Sidharth Bhatia* Stefan Roth* Michael J. Black* Michael Isard†

*Department of Computer Science, Brown University Providence, RI 02912

†Microsoft Research Silicon Valley, Mountain View, CA 94043

{ls,sibhatia,roth,black}@cs.brown.edu, misard@microsoft.com

Abstract

We pose the problem of 3D human tracking as one of inference in a graphical model. Unlike traditional kinematic tree representations, our model of the body is a collection of loosely-connected limbs. Conditional probabilities relating the 3D pose of connected limbs are learned from motion-captured training data. Similarly, we learn 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. The loose-limbed model and decentralized graph structure facilitate the use of low-level visual cues. We adopt simple limb and head detectors to provide “bottom-up” information that is incorporated into the inference process at every time-step; these detectors permit automatic initialization and aid recovery from transient tracking failures. We illustrate the method by automatically tracking a walking person in video imagery using four calibrated cameras. Our experimental apparatus includes a marker-based motion capture system aligned with the coordinate frame of the calibrated cameras with which we quantitatively evaluate the accuracy of our 3D person tracker.

1 Introduction

We present a fully automatic method for tracking human bodies in 3D. Initialization and failure recovery are facilitated by the use of a loose-limbed body model [22] in which limbs are connected via learned probabilistic constraints. The tracking problem is formulated as one of inference in a graphical model and belief propagation is used to estimate the pose of the body at each time-step. Each node in the graphical model represents the 3D position and orientation of a limb (Figure 1). Directed edges between nodes represent statistical dependencies and these constraints between limbs are used to form messages that are sent to neighboring nodes in space and time. Additionally, each node has an associated likelihood defined over a rich set of image cues using a learned Gibbs model [19, 28]. The combination of

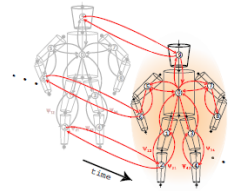


Figure 1: Graphical model for a person. Nodes represent limbs and arrows represent conditional dependencies between limbs. For clarity only a few temporal dependencies have been shown: in our model each part i at time t is connected by an edge to part i at times $t - 1$ and $t + 1$.

highly non-Gaussian likelihoods and a six-dimensional continuous parameter space (3D position and orientation) for each limb makes standard belief propagation algorithms infeasible. Consequently we exploit a form of non-parametric belief propagation that uses a variation of particle filtering and can be applied over a loopy graph [13, 25].

There are a number of significant advantages to this approach as compared to traditional methods for tracking human motion. Most current techniques model the body as a kinematic tree in 2D [14], 2.5D [17], or 3D [2, 5, 21, 23] leading to a high-dimensional parameter space (25–50 dimensions is not uncommon). Searching such a space directly is impractical and so current methods typically rely on manual initialization of the body model. Additionally, they often exploit strong priors characterizing the types of motions present. When such algorithms lose track (as they always do), the dimensionality of the state space makes it difficult to recover.

While the full body pose is hard to recover directly, the location and pose of individual limbs is much easier to compute. Many good head detectors exist and limb detectors have been used for some time (e.g. [18]). The approach we take here can use bottom up information from feature

Meta 1009
p. 1

- ▶ TLP was publicly available for the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (“CVPR”) no later than April 9, 2004.
- ▶ The ’980 provisional was filed May 27, 2004 and identifies TLP as reference [2] with a date of 2004.
- ▶ If TLP does not receive the benefit of the filing date of its provisional, then there is no dispute that TLP is prior art to the ’980.
- ▶ TLP is the Primary reference in Ground IV.

Petitioners' Reply Does Not Introduce New Theories

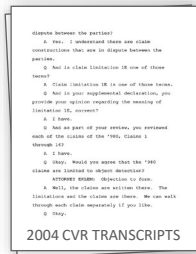


the effective filing date of the '980 patent. To be clear, we understand Petitioner's theory to be that TLP was made publicly available for the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition on April 9, 2004. Dec. 48; Pet. 6. We also understand Petitioner to argue that Mr. Black's purported publication of TLP on the ResearchGate website on January 31, 2004, makes it more likely that TLP was made publicly available for the 2004 IEEE Conference. Reply 3.

Paper 28 at 4

TLP's Public Availability Timeline

11/19/2003
Due date for authors
to submit full paper
for the 2004 CVPR



November 2003

2 Months

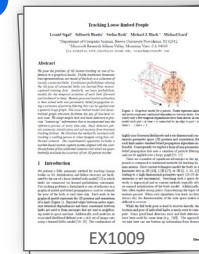
01/31/2004
According to Research
Gate, TLP author Michael
Black makes TLP publicly
available



January 2004

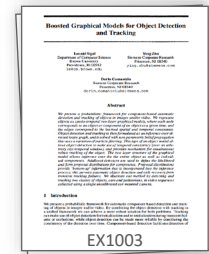
3 Months

04/09/2004
Deadline for
final manuscript
submission



April 2004

05/27/2004
'980 Provisional
filed and identifies
TLP as a reference



May 2004

Dr. Hall-Ellis Confirmed that TLP Was Publicly Available by January 31, 2004

Petitioners'

Librarian Expert Witness



Dr. Sylvia Hall-Ellis

46. First, I examined the conference Call for Papers. The Call for Papers for the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition required authors to submit manuscripts by November 19, 2003.¹⁸

47. Second, I examined the 29 *ResearchGate* entries that Michael J. Black posted during the period 2003-2005 (see Attachment 1E). For each of the entries the date indicated closely aligned with an official publication date for the document. I identified the publication date for each conference paper and journal article. The Sigal paper reflected the longest gap but aligned appropriately with the requirements stated in the Conference Call for Papers. Moreover, there was a strong indicia of reliability in that the *ResearchGate* entries did not simply list “January” for every paper/article, and different months were indicated for different papers/articles.

TLP Author Michael Black's Research Gate Entries Closely Align with Official Publication Dates



“[e]vidence of a person's habit ... may be admitted to prove that on a particular occasion the person ... acted in accordance with the habit or routine practice.”

Nobel Biocare Services v. Intradent, 903 F.3d 1365, 1377 (Fed. Cir. 2018)
(quoting *In re Hall*, 781 F.2d 897, 898–99 (Fed. Cir. 1986))

Dr. Saber Has No Personal Knowledge of Exhibits 2009 and 2010

**Patent Owner's
Expert Witness**



Dr. Eli Saber

Q Do you have any personal knowledge regarding the circumstances of the Exhibits 2009 and 2010, which I'm going to hand to you now.

A What do you mean, personal knowledge?

Q Did you receive them in the 2000 time frame?

A No. I --

Q Were you CC'd?

A No. I wasn't. No. I don't mean -- I've seen these in the confines of this case.

Q So the first time you saw them is after you'd been retained as an expert in this case?

A That's correct.

Overview

US007436980B2

(12) **United States Patent**
Sigal et al.

(10) **Patent No.:** **US 7,436,980 B2**
(45) **Date of Patent:** **Oct. 14, 2008**

(54) **GRAPHICAL OBJECT MODELS FOR
DETECTION AND TRACKING**

7,200,266 B2 * 4/2007 Ozer et al. 382/173

FOREIGN PATENT DOCUMENTS

(75) Inventors: **Leonid Sigal**, Providence, RI (US); **Ying Zhu**, Monmouth Junction, NJ (US);
Dorin Comaniciu, Princeton Jct., NJ
(US)

EP 1 217 572 A2 6/2002

OTHER PUBLICATIONS

(73) Assignee: **Siemens Corporate Research, Inc.**,
Princeton, NJ (US)

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 691 days.

M. Isard, P. Pong, Real-valued graphical models for computer
vision, CVPR, vol. 1, pp. 613-620, 2003.

K. Okuma, A. Tachibana, N. de Freitas, J. Little and D. Lowe. A
boosted particle filter: Multi-target detection and tracking. ECCV,
2004.

L. Sigal, S. Bhatia, S. Roth, M. Black and M. Isard. Tracking loose-
ly-labeled people, CVPR, 2004.

E. Sudderth, A. Ihler, W. Freeman and A. Willsky. Nonparametric
belief propagation. MIT Laboratory for Information and Decision
Systems Technical Report P-2551 (Oct. 15, 2002).

P. Viola and M. Jones. Rapid object detection using a boosted cascade
of simple features, CVPR, 2001.

P. Viola, M. Jones and D. Snow. Detecting pedestrians using patterns
of motion and appearance, ICCV, pp. 734-741, 2003.

(21) Appl. No.: **11/135,210**

(22) Filed: **May 23, 2005**

(65) **Prior Publication Data**

US 2005/0286738 A1 Dec. 29, 2005

(Continued)

Related U.S. Application Data

(60) Provisional application No. 60/574,799, filed on May
27, 2004.

Primary Examiner—Tom Y Lu

(74) Attorney, Agent, or Firm—Donald B. Paschburg; F. Chau
& Associates, LLC

(51) Int. Cl.
G06K 9/00 (2006.01)

(52) U.S. Cl. **382/103**

(58) Field of Classification Search **382/103,**
382/173

See application file for complete search history.

(57) **ABSTRACT**

A computer implemented method for object detection
includes providing a spatio-temporal model for an object to
be detected, providing a video including a plurality of images
including the object, and measuring the object as a collection
of components in each image. The method further includes
determining a probability that the object is in each image, and
detecting the object in any image upon comparing the prob-
abilities for each image to a threshold for detecting the object.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,597,801 B1 7/2003 Cham et al.

16 Claims, 5 Drawing Sheets

```
graph TD
    501[PROVIDING A SPATIO-TEMPORAL MODEL FOR AN OBJECT] --> 502[PROVIDING A VIDEO COMPRISING IMAGES INCLUDING THE OBJECT]
    502 --> 503[MEASURING THE OBJECT AS A COLLECTION OF COMPONENTS IN EACH IMAGE]
    503 --> 504[DETERMINING A PROBABILITY THAT THE OBJECT IS IN EACH IMAGE]
    504 --> 505[DETECTING THE OBJECT IN AN IMAGE UPON COMPARING THE PROBABILITIES FOR THE IMAGE TO A THRESHOLD FOR DETECTING THE OBJECT]
```

1. Background

2. Claim Construction

- a. 1/9[a]: spatio-temporal model
- b. 1/9[c]: measuring the object as a collection of components
- c. 1/9[e]: comparing the probabilities for each image

3. Grounds

- a. Claims 1-16: Obvious in view of Zhao
- b. Claims 1-16: Obvious over Steffens in view of Zhao
- c. Claims 1-16: Obvious over Ozer in view of Zhao
- d. Claims 1-16: Obvious in view of TLP**

Prior Art - TLP (EX1009)

Tracking Loose-limbed People

Leonid Sigal^{*} Sidharth Bhatia^{*} Stefan Roth^{*} Michael J. Black^{*} Michael Isard[†]

^{*}Department of Computer Science, Brown University Providence, RI 02912

[†]Microsoft Research Silicon Valley, Mountain View, CA 94043

{ls,sibhatia,roth,black}@cs.brown.edu, misard@microsoft.com

Abstract

We pose the problem of 3D human tracking as one of inference in a graphical model. Unlike traditional kinematic tree representations, our model of the body is a collection of loosely-connected limbs. Conditional probabilities relating the 3D pose of connected limbs are learned from motion-captured training data. Similarly, we learn 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. The loose-limbed model and decentralized graph structure facilitate the use of low-level visual cues. We adopt simple limb and head detectors to provide “bottom-up” information that is incorporated into the inference process at every time-step; these detectors permit automatic initialization and aid recovery from transient tracking failures. We illustrate the method by automatically tracking a walking person in video imagery using four calibrated cameras. Our experimental apparatus includes a marker-based motion capture system aligned with the coordinate frame of the calibrated cameras with which we quantitatively evaluate the accuracy of our 3D person tracker.

1 Introduction

We present a fully automatic method for tracking human bodies in 3D. Initialization and failure recovery are facilitated by the use of a loose-limbed body model [22] in which limbs are connected via learned probabilistic constraints. The tracking problem is formulated as one of inference in a graphical model and belief propagation is used to estimate the pose of the body at each time-step. Each node in the graphical model represents the 3D position and orientation of a limb (Figure 1). Directed edges between nodes represent statistical dependencies and these constraints between limbs are used to form messages that are sent to neighboring nodes in space and time. Additionally, each node has an associated likelihood defined over a rich set of image cues using a learned Gibbs model [19, 28]. The combination of

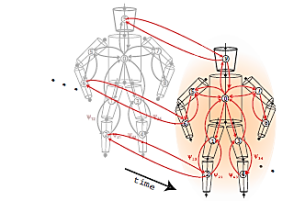


Figure 1: Graphical model for a person. Nodes represent limbs and arrows represent conditional dependencies between limbs. For clarity only a few temporal dependencies have been shown: in our model each part i at time t is connected by an edge to part i at times $t-1$ and $t+1$.

highly non-Gaussian likelihoods and a six-dimensional continuous parameter space (3D position and orientation) for each limb makes standard belief propagation algorithms infeasible. Consequently we exploit a form of non-parametric belief propagation that uses a variation of particle filtering and can be applied over a loopy graph [13, 25].

There are a number of significant advantages to this approach as compared to traditional methods for tracking human motion. Most current techniques model the body as a kinematic tree in 2D [14], 2.5D [17], or 3D [2, 5, 21, 23] leading to a high-dimensional parameter space (25–50 dimensions is not uncommon). Searching such a space directly is impractical and so current methods typically rely on manual initialization of the body model. Additionally, they often exploit strong priors characterizing the types of motions present. When such algorithms lose track (as they always do), the dimensionality of the state space makes it difficult to recover.

While the full body pose is hard to recover directly, the location and pose of individual limbs is much easier to compute. Many good head detectors exist and limb detectors have been used for some time (e.g. [18]). The approach we take here can use bottom up information from feature

Meta 1009
o. 1

► Primary reference in Ground IV

2 Graphical Object Models

Following the framework of [1, 2] we model an object as a spatio-temporal directed graphical model. Each node in this graph represents an object or a component of an object at time t . The direct edges correspond to the learned spatial and temporal constraints. Each node has an associated state vector $X^T = (x, y, s)$ defining the component's position and scale within an image (where $x, y, s \in \mathbb{R}$). Finding and tracking of an object is formulated as an inference over this graphical model.

Fig. 2c and 2d. Unlike the model presented in [2] having a second-layer also allows us to assume conditional independence of components over time given the whole appearance, hence reducing the complexity of the overall graphical model.

References

- [1] L. Sigal, M. Isard, B.H. Sigelman and M.J. Black. Attractive people: Assembling loose-limbed models using non-parametric belief propagation, *Advances in Neural Info. Proc. Sys.* 16, 2003.
- [2] L. Sigal, S. Bhatia, S. Roth, M.J. Black and M. Isard. Tracking Loose-limbed People, *CVPR*, 2004.

Ground IV: Obvious in view of TLP

Claim 1, '980 Patent

Preamble [1]	A computer implemented method for object detection comprising:
Element 1[a]	providing <i>a spatio-temporal model</i> for an object to be detected;
Element 1[b]	providing a video comprising a plurality of images including the object;
Element 1[c]	measuring the object as a collection of components in each image;
Element 1[d]	determining a probability that the object is in each image; and
Element 1[e]	detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.

Under plain and ordinary meaning

Claim 1 does do not require the spatial temporal model (i) to be graphical, (ii) to be two-layers, or (iii) to include a separate node corresponding to the object.

Under PO's D. Del. position

TLP taught or rendered obvious a graphical model having a node corresponding to the object.

TLP Taught or Rendered Obvious a Graphical Model Having a Node Corresponding to the Object

Tracking Loose-Limbed

Leonid Sigal* Sidharth Bhatia* Stefan Roth†

*Department of Computer Science, Brown University

†Microsoft Research Silicon Valley

{ls,sibhatia,roth,black}@cs.brown.edu

Abstract

We pose the problem of 3D human tracking as one of inference in a graphical model. Unlike traditional kinematic tree representations, our model of the body is a collection of loosely-connected limbs. Conditional probabilities relating the 3D pose of connected limbs are learned from motion-captured training data. Similarly, we learn 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. The loose-limbed model and decentralized graph structure facilitate the use of low-level visual cues. We adopt simple limb and head detectors to provide “bottom-up” information that is incorporated into the inference process at every time-step; these detectors permit automatic initialization and aid recovery from transient tracking failures. We illustrate the method by automatically tracking a walking person in video imagery using four calibrated cameras. Our experimental apparatus includes a marker-based motion capture system aligned with the coordinate frame of the calibrated cameras with which we quantitatively evaluate the accuracy of our 3D person tracker.

1 Introduction

We present a fully automatic method for tracking human bodies in 3D. Initialization and failure recovery are facilitated by the use of a loose-limbed body model [22] in which limbs are connected via learned probabilistic constraints. The tracking problem is formulated as one of inference in a graphical model and belief propagation is used to estimate the pose of the body at each time-step. Each node in the graphical model represents the 3D position and orientation of a limb (Figure 1). Directed edges between nodes represent statistical dependencies and these constraints between limbs are used to form messages that are sent to neighboring nodes in space and time. Additionally, each node has an associated likelihood defined over a rich set of image cues using a learned Gibbs model [19, 28]. The combination of

2 Loose-limbed body model

Following the framework in [22] 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. 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. Edges in the graphical model correspond to position and angle relationships between adjacent body parts in space and time, as illustrated in Figure 1.

and can be applied over a loopy graph [13, 25].

There are a number of significant advantages to this approach as compared to traditional methods for tracking human motion. Most current techniques model the body as a kinematic tree in 2D [14], 2.5D [17], or 3D [2, 5, 21, 23] leading to a high-dimensional parameter space (25–50 dimensions is not uncommon). Searching such a space directly is impractical and so current methods typically rely on manual initialization of the body model. Additionally, they often exploit strong priors characterizing the types of motions present. When such algorithms lose track (as they always do), the dimensionality of the state space makes it difficult to recover.

While the full body pose is hard to recover directly, the location and pose of individual limbs is much easier to compute. Many good head detectors exist and limb detectors have been used for some time (e.g. [18]). The approach we take here can use bottom up information from feature

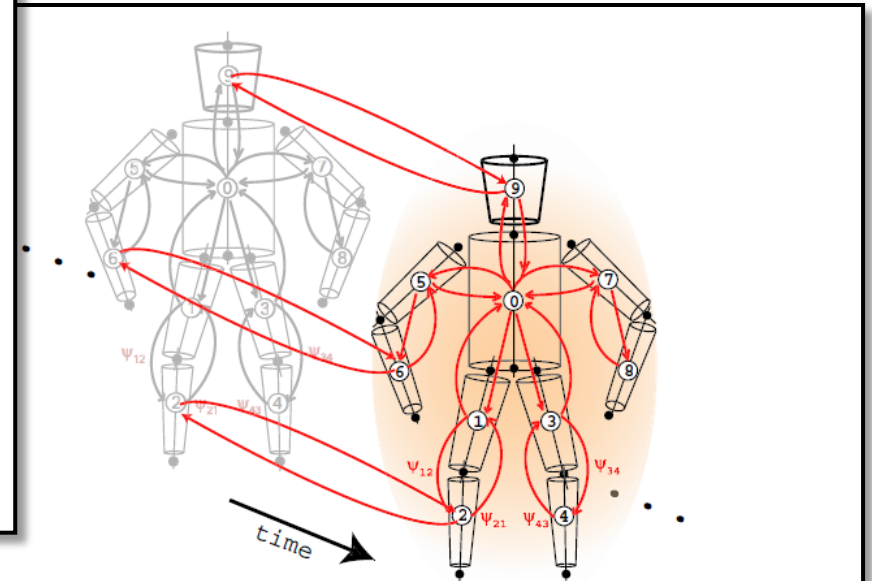


Figure 1: *Graphical model for a person.* Nodes represent limbs and arrows represent conditional dependencies between limbs. For clarity only a few temporal dependencies have been shown: in our model each part i at time t is connected by an edge to part i at times $t - 1$ and $t + 1$.

Meta 1009
n. 1

TLP Model Has a Probabilistic Relationship Between Joints Over Time

Patent Owner's Expert Witness



Dr. Eli Saber

Q Doesn't TLP teach that its loose-limbed body model requires a specification of the probabilistic relationships between joints at a given time instant and over time?

ATTORNEY EKLEM: Objection. Form.

A Yes. You read that correctly.

Q What does that mean to you?

A The loose-limbed body model, just like you see in Figure 1, you have a loose-limbed body model requires a specification of the probabilistic relationship between joints. So you have the probabilistic relationship between joints at a given time instant. So at a given time instant, you have a probabilistic relationship between these joints and over time.

Petitioners' Expert



Dr. Chandra Bajaj

204. ...

this element obvious. 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 at 2. A POSITA would have found the claimed spatio-temporal model disclosed or at least rendered obvious in view of TLP, as the nodes correspond to each of the collection of components and the object.

Dr. Bajaj Elaborated in His Supplemental Declaration

Petitioners' Expert



Dr. Chandra Bajaj

75. A POSITA would have understood from PAMPAS to select a component of the object (e.g., the torso) to be representative of the object node x_1 (i.e., a 1st layer), and to conditionally link all other components of the body (head, limbs, etc.) forming a 2nd layer of x_i nodes. See PAMPAS, Ex1019, § 3.3 Finding a jointed object in clutter. Due to this two-layered model, even when the object

PAMPAS (EX1019)

PAMPAS: Real-Valued Graphical Models for Computer Vision

M. Isard
Microsoft Research
Mountain View, CA 94043

Abstract

Probabilistic models have been adopted for many computer vision applications, however inference in high-dimensional spaces remains problematic. As the state-space of a model grows, the dependencies between the dimensions lead to an exponential growth in computation when performing inference. Many common computer vision problems naturally map onto the graphical model framework; the representation is a graph where each node contains a portion of the state-space and there is an edge between two nodes only if they are not independent conditional on the other nodes in the graph. When this graph is sparsely connected, belief propagation algorithms can turn an exponential inference computation into one which is linear in the size of the graph. However belief propagation is only applicable when the variables in the nodes are discrete-valued or jointly represented by a single multivariate Gaussian distribution, and this rules out many computer vision applications.

This paper combines belief propagation with ideas from particle filtering; the resulting algorithm performs inference on graphs containing both cycles and continuous-valued latent variables with general conditional probability distributions. Such graphical models have wide applicability in the computer vision domain and we test the algorithm on example problems of low-level edge linking and locating jointed structures in clutter.

1. Introduction

Probabilistic techniques for modelling uncertainty have found widespread success in computer vision. Their application ranges from pixel-level image generation models [13, 7] to tracking in high-level parametric spaces [11, 22] however as the size of the state-space grows the “curse of dimensionality” can cause an exponential growth in the computations necessary to perform inference. Many probabilistic vision models naturally decompose into a graphical model structure [14]; basic components become nodes in a graph where each node is conditionally independent of all

but its near neighbours. When the nodes are image patches the neighbours may be spatially close in an image, adjacent levels in a multi-scale representation or nearby time instants in an image sequence. Alternatively many complex objects can be decomposed into graphs where the nodes are subparts of the object and an edge indicates that two subparts are physically connected. The advantage of this representation is that some formerly exponential inference computations become linear in the size of the graph.

Exact inference on graphical models is possible only in restricted circumstances [14]: the (directed) graph must be acyclic and the joint distribution over those latent variables which are not discrete must be a single multivariate Gaussian. When a model violates these restrictions, Gibbs sampling can be used to draw approximate samples from the joint distribution [9] but for most computer vision applications this technique remains intractable. Approximate inference methods can be used for Conditional Linear Gaussian models [16] including time series [3]. Recently two methods have been widely studied in the computer vision literature to perform approximate inference on more general classes of graph: loopy belief propagation [26] (LBP) can be applied to graphs with cycles, though it still only applies to discrete or jointly Gaussian random variables; and particle filtering [5] allows the use of very general distributions over continuous-valued random variables but applies only to graphs with a simple linear chain structure.

The restriction to Gaussian latent variables is particularly onerous in computer vision applications because clutter in image generation models invariably leads to highly non-Gaussian, multi-modal posterior distributions. This effect largely accounts for the popularity of particle filtering in the computer vision tracking community. In this paper we describe the PAMPAS algorithm (Particle Message Passing) which combines LBP with ideas from particle filtering. It has wide applicability in the computer vision domain, and increases the range of complex continuous-valued models which admit tractable probabilistic inference. We investigate the behaviour of the algorithm on test problems including illusory-contour finding and locating a jointed structure. Additionally we argue that an algorithm which performs

Meta 1019
p. 1

3.3. Finding a jointed object in clutter

Finally we constructed a graph representing a jointed object and applied the PAMPAS algorithm (combined with the Gibbs sampler from [24]) to locating the object in a cluttered scene. The model consists of 9 nodes; a central circle with four jointed arms each made up of two rectangular links. The circle node $\mathbf{x}_1 = (x_0, y_0, r_0)$ encodes a position and radius. Each arm node $\mathbf{x}_i = (x_i, y_i, \alpha_i, w_i, h_i)$, $2 \leq i \leq 9$ encodes a position, angle, width and height and prefers one of the four compass directions (to break symmetry). The arms pivot around their inner joints, so the po-

The second example of figure 5 demonstrates the power of the probabilistic modelling approach, since now the circle at the centre of the object has not been displayed, simulating occlusion. Of course no $\mathbf{x}_1$ initialisation samples are generated near the “occluded” circle, but even after one iteration the belief at that node has high weight at the correct location due to the agreement of the nearby arm nodes. After a few iterations the belief has converged on the correct object position despite the lack of the central circle.

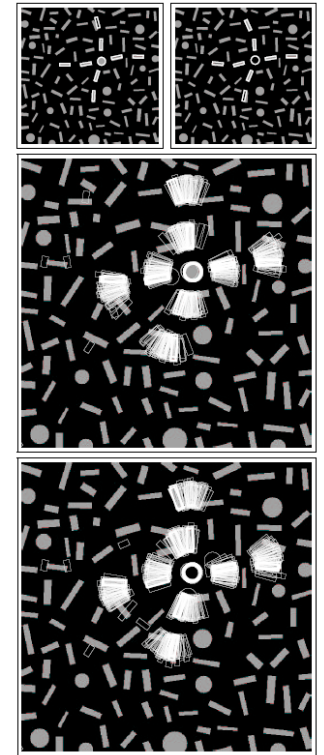


Figure 5. A jointed object is located in clutter. Two experiments are shown, with and without a simulated occlusion. The top two images show a sample from the joint distribution for each experiment, while the lower images show the full distributions. The second experiment models an occluded object with no circle present in the image where the belief has placed it. This experiment uses the Gibbs sampler from [24] to form the message product for the central node.