



REICHMAN JORGENSEN
LEHMAN & FELDBERG LLP

SILICON VALLEY • NEW YORK • ATLANTA • WASHINGTON, D.C.

Patent Owner's Presentation
IPR2023-00924
U.S. Patent No. 7,436,980

Meta Platforms, Inc. v. VL Collective IP LLC
September 10, 2024

Statutory Grounds for Challenges

Ground	Assertion	Petitioned Claims
1	Obvious under §103 over Zhao	1-16
2	Obvious under §103 over Steffens in view of Zhao	1-16
3	Obvious under §103 over Ozer in view of Zhao	1-16
4	Obvious under §103 over TLP	1-16

'980 Patent

(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 Ozar et al. 382/173
(75) Inventors: Leonid Sigal , Providence, RI (US); Ying Zhu , Monmouth Junction, NJ (US); Dorin Comaniciu , Princeton, NJ (US)	FOREIGN PATENT DOCUMENTS EP 1 217 572 A2 6/2002
(73) Assignee: Siemens Corporate Research, Inc. , Princeton, NJ (US)	OTHER PUBLICATIONS M. Isard, P. Papan: 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. Biler, W. Freeman and A. Willsky: Nonparametric belief propagation, MIT Laboratory for Information and Decision Systems Technical Report P2-551 (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	<i>Primary Examiner</i> —Tom Y. Lu
(22) Filed: May 23, 2005	(74) <i>Attorney, Agent, or Firm</i> —Donald B. Paschburg, F. Chau & Associates, LLC
(65) Prior Publication Data US 2005/0286738 A1 Dec. 29, 2005	(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.
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[501 PROVIDING A SPATIO-TEMPORAL MODEL FOR AN OBJECT] --> 502[502 PROVIDING A VIDEO COMPRISING IMAGES INCLUDING THE OBJECT] 502 --> 503[503 MEASURING THE OBJECT AS A COLLECTION OF COMPONENTS IN EACH IMAGE] 503 --> 504[504 DETERMINING A PROBABILITY THAT THE OBJECT IS IN EACH IMAGE] 504 --> 505[505 DETECTING THE OBJECT IN AN IMAGE UPON COMPARING THE PROBABILITIES FOR THE IMAGE TO A THRESHOLD FOR DETECTING THE OBJECT]</pre>	

Title: Graphical Object Models for Detection and Tracking

Inventors: Leonid Sigal, Ying Zhu, Dorin Comaniciu

Original Assignee: Siemens Corporate Research, Inc.,

Priority Date: May 27, 2004

EX1001 ('980 Patent) at 1; Paper 3 (Petition) at 5



'980 Patent

OVERVIEW OF THE '980 PATENT



The '980 Patent Addresses a Problem Related to Object Detection

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. 15 20

EX1001 ('980 Patent) at 1:15-21; Paper 26 (POR) at 1-2



The Improved Object Detection Technique of the '980 Patent

Therefore, a need exists for a system and method of detecting and tracking an object, implementing component detection and performing inference over space and time.

EX1001 ('980 Patent) at 1:22-24; Paper 26 (POR) at 1-2

According to an embodiment of the present disclosure, a probabilistic framework for automatic component-based detection and tracking of objects in images and/or video combines object detection with tracking in a unified framework. Tracking makes use of object detection for initialization and re-initialization during transient failures for occlusions. Object detection considers the consistency of the detection over time. Modeling objects by an arrangement of image-base, and possibly overlapping, components facilitates detection of complex articulated objects as well as helps in handling partial object occlusions or local illumination changes.

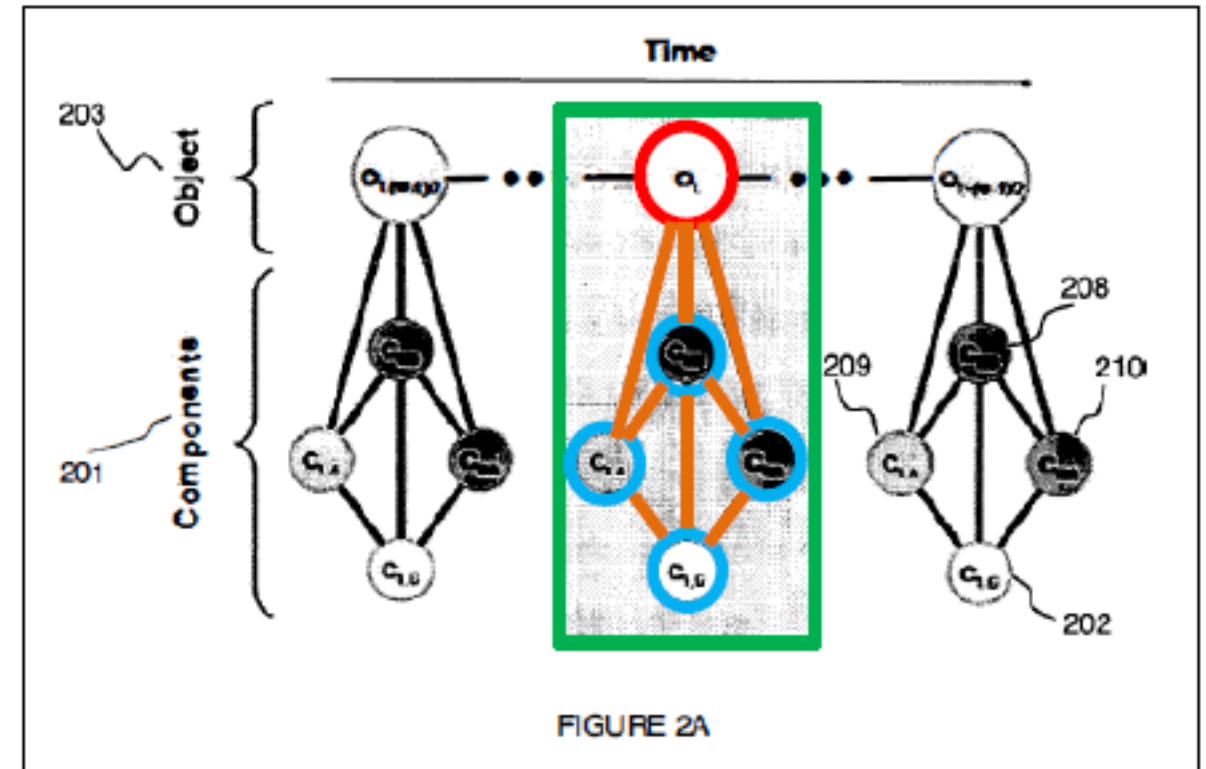
EX1001 ('980 Patent) at 2:43-54; Paper 26 (POR) at 2



The Improved Object Detection Technique of the '980 Patent

Referring to FIGS. 2A and 2B, object detection and tracking is formulated as an inference in a two-layer graphical model in which a coarse layer node represents the whole object and fine layer nodes represent multiple components of the object. Directed edges between nodes represent learned spatial and temporal probabilistic constraints. Each node in the graphical model corresponds to a position and scale of the component or the object as a whole in an image at a given time instant. Each node also has an associated Adaptive Boosting (Adaboost) detector that is used to define the local image likelihood and a proposal process. Typically the likelihoods

EX1001 ('980 Patent) at 2:55-3:2; Paper 26 (POR) at 2

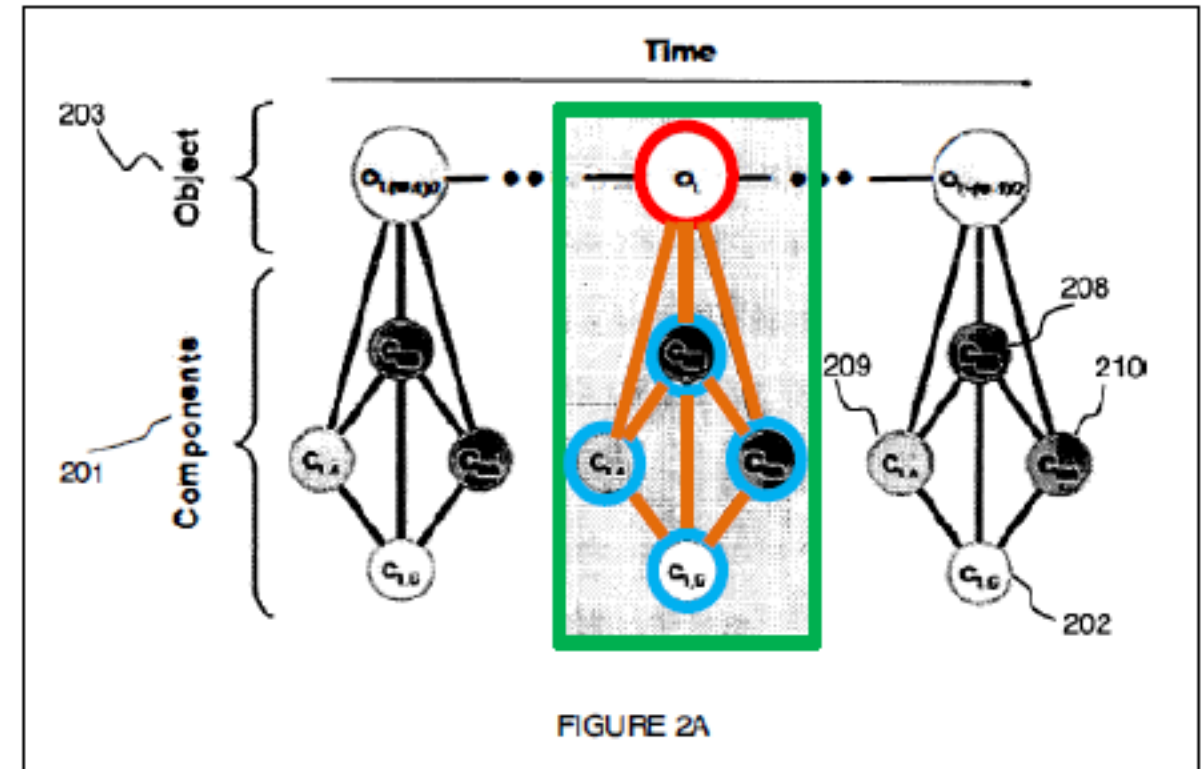


EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 2-3

The Improved Object Detection Technique of the '980 Patent

Building the Graphical Model; For a single frame, objects are represented using a two-layer spatial graphical model, for example as shown in FIGS. 2A and 2B. The fine, component, layer 201 includes a set of loosely connected parts, e.g., 202. The coarse, object, layer 203 corresponds to an entire appearance model of the object and is connected to all constituent components, e.g., 202. Examples of such models for pedes-

EX1001 ('980 Patent) at 4:18-24; Paper 26 (POR) at 2-3

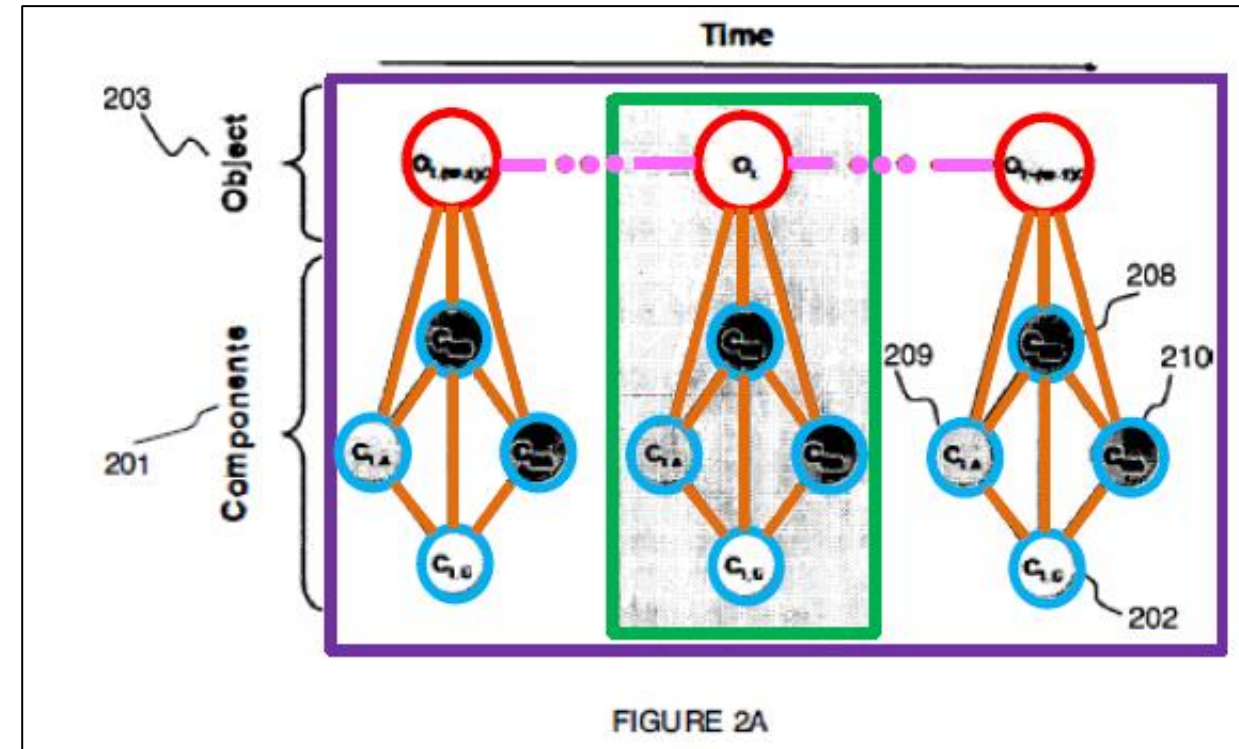


EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 2-3

The Improved Object Detection Technique of the '980 Patent

To integrate temporal constraints the spatial graphical models are extended over time to an arbitrary length temporal window. The resulting spatio-temporal graphical models are shown in FIGS. 2A and 2B. Having a two-layer graphical model allows the inference process to reason explicitly about the object as a whole, e.g., 203, and reduce the complexity of the graphical model by allowing the assumption of the conditional independence of components, e.g., 202 and 208-210, over time given the overall object appearance. Alternatively, a single object layer model can be built with bi-directional temporal constraints.

EX1001 ('980 Patent) at 4:34-45; Paper 26 (POR) at 3-4



EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 3-4

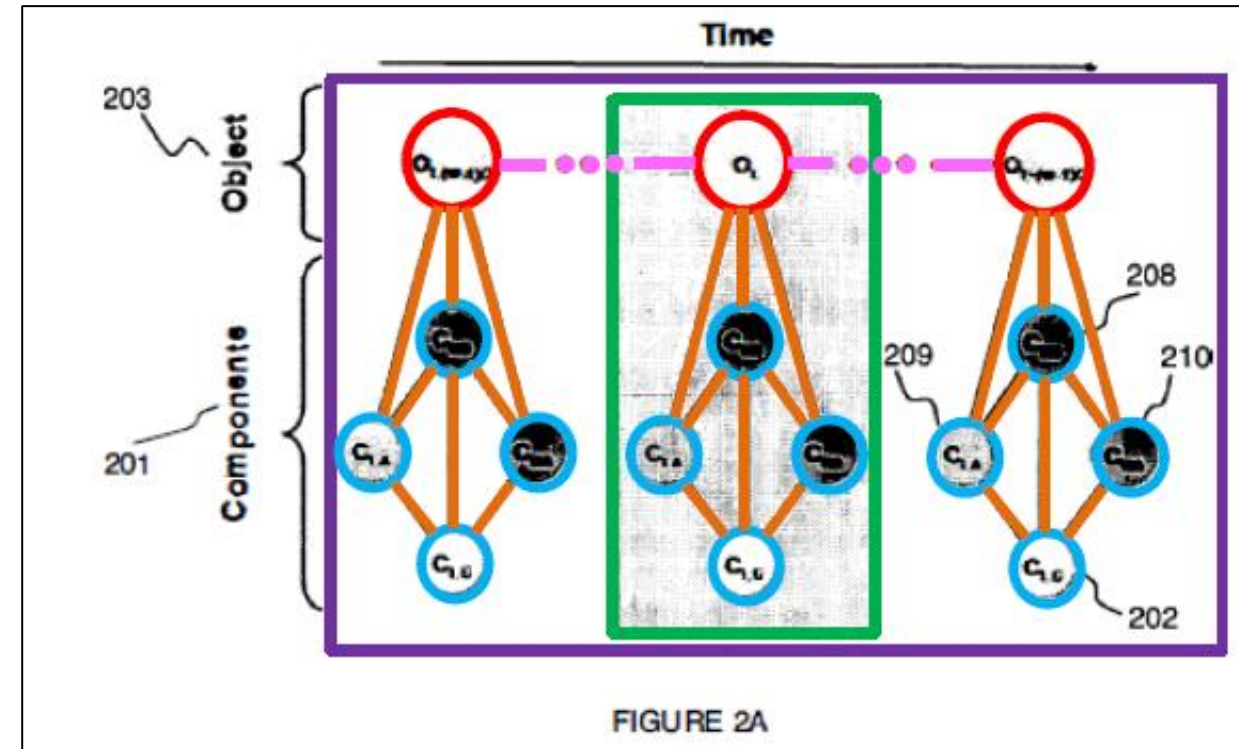
The Improved Object Detection Technique of the '980 Patent

the object. Directed edges between nodes represent learned spatial and temporal probabilistic constraints. Each node in

EX1001 ('980 Patent) at 2:59-60; Paper 26 (POR) at 3-4

According to an embodiment of the present disclosure, a system comprises the following modules: a graphical model modules, an inference module providing the ability to infer a state of each node in the graph, a local evidence distribution (or image likelihood) module, a proposal process module for some or all nodes in a graphical model, and a module comprising a set of spatial and/or temporal constraints corresponding to the edges in a graph.

EX1001 ('980 Patent) at 4:15-17; Paper 26 (POR) at 3-4



EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 3-4

The Improved Object Detection Technique of the '980 Patent

Graphical Object Models; An object, such as an automobile (see FIG. 2B), is modeled as a spatio-temporal directed graphical model. Each node in the graph represents either the object or a component of the object at time t . Nodes have an associated state vector $X^T = (x, y, s)$ defining the component's real-valued position and scale within an image (where $x, y, s, \in \Re$). The joint probability distribution for this spatio-temporal

EX1001 ('980 Patent) at 3:47-53; Paper 26 (POR) at 4

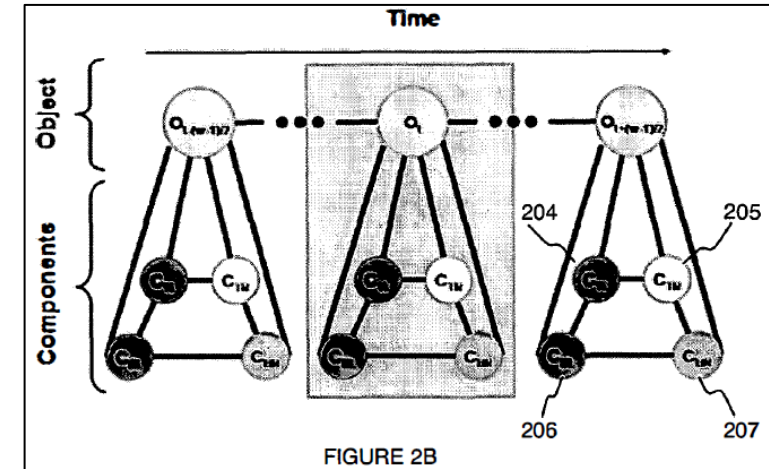
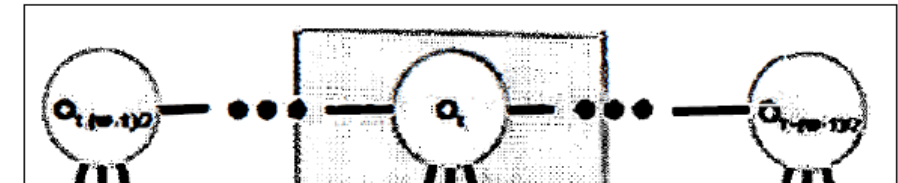


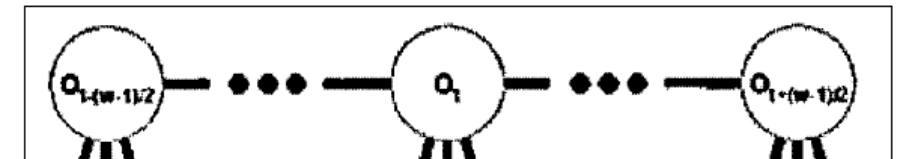
FIGURE 2B

As can be seen in FIGS. 2A and 2B, an object at a current time is denoted as O_t , with an earlier model denoted as $O_{t-(w-1)/2}$ and a later model denoted as $O_{t+(w-1)/2}$.

EX1001 ('980 Patent) at 4:45-48; Paper 26 (POR) at 4



EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 4-5



EX1003 (Provisional) Fig. 2(d); Paper 26 (POR) at 4-5

The Improved Object Detection Technique of the '980 Patent

3). The joint probability distribution for this spatio-temporal graphical object model with N components can be written as:

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

where X_t^O and $X_t^{C_n}$ is the state of the object, O, and object's n-th component, C_n , at time t respectively ($n \in (1, N)$ and $t \in (1, T)$); $\psi_{ik}(X_i^O, X_j^O)$ is the temporal compatibility of object state between frames i and j; $\psi_{ik}(X_i^O, X_i^{C_k})$ is the spatial compatibility of the object and it's components at frame i; $\psi_{ik}(X_i^{C_k}, X_i^{C_l})$ is the spatial compatibility between object components at frame i and $\phi_i(X_i^O)$ and $\phi_i(X_i^{C_k})$ denote the local evidence for the object and component states respectively.

$\psi_{ik}(X_i^O, X_j^O)$ is the temporal compatibility of object state between frames i and j; $\psi_{ik}(X_i^O, X_i^{C_k})$ is the spatial compatibility of the object and it's components at frame i; $\psi_{ik}(X_i^{C_k}, X_i^{C_l})$ is the spatial compatibility between object components at frame i and $\phi_i(X_i^O)$ and $\phi_i(X_i^{C_k})$ denote the local evidence for the object and component states respectively.

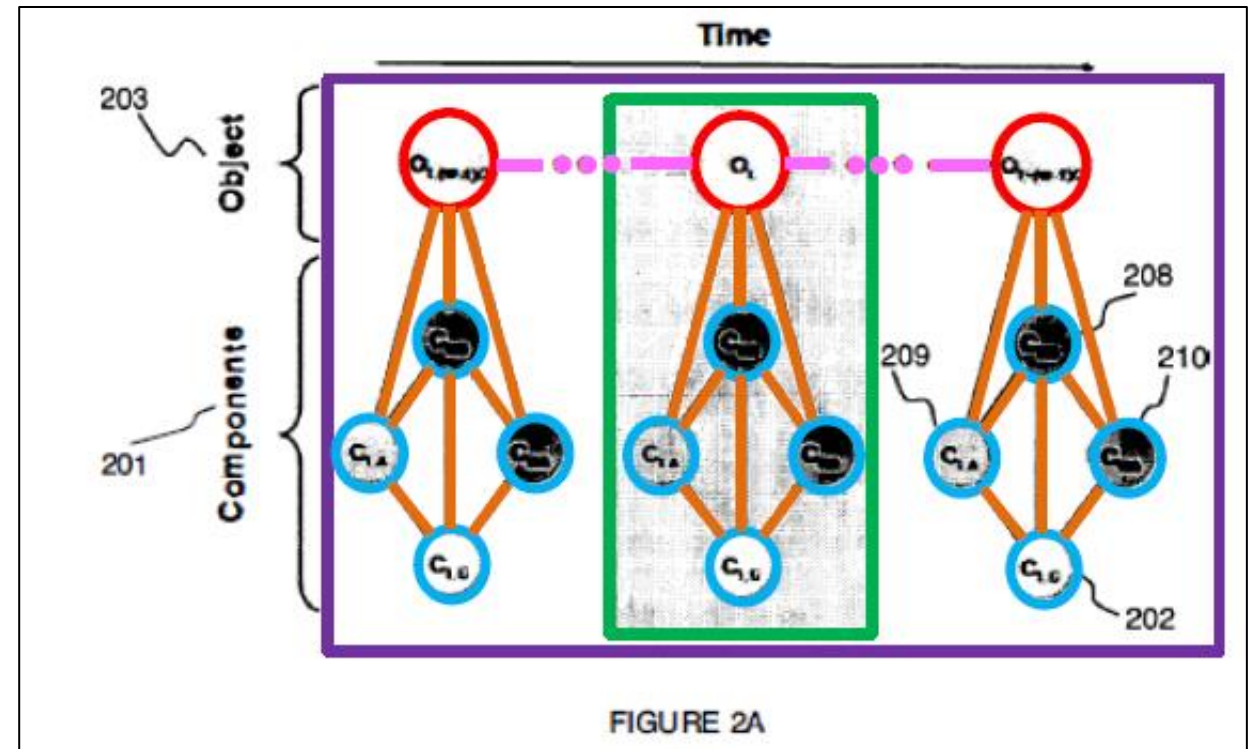


FIGURE 2A

EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 3-4

EX1001 ('980 Patent) at 3:53-4:8; Paper 26 (POR) at 5

The Improved Object Detection Technique of the '980 Patent

AdaBoost Image Likelihoods; A likelihood $\phi_i(X_i)$ models the probability of observing the image conditioned on the state X_i of the node i . A likelihood model according to an embodiment of the present disclosure uses a boosted classifier and is robust to partial occlusions and the variability of image statistics across many different inputs.

EX1001 ('980 Patent) at 5:25-30; Paper 26 (POR) at 5-6

Given a set of labeled patterns the AdaBoost procedure learns a weighted combination of base weak classifiers,

$$h(I) = \sum_{k=1}^K \alpha_k h_k(I),$$

where I is an image pattern, and $h_k(I)$ is the weak classifier chosen for the round k of boosting, and α_k is the corresponding weight. A weak classifier method is implemented: $\alpha_k(I) =$

EX1001 ('980 Patent) at 5:41-57; Paper 26 (POR) at 5-6

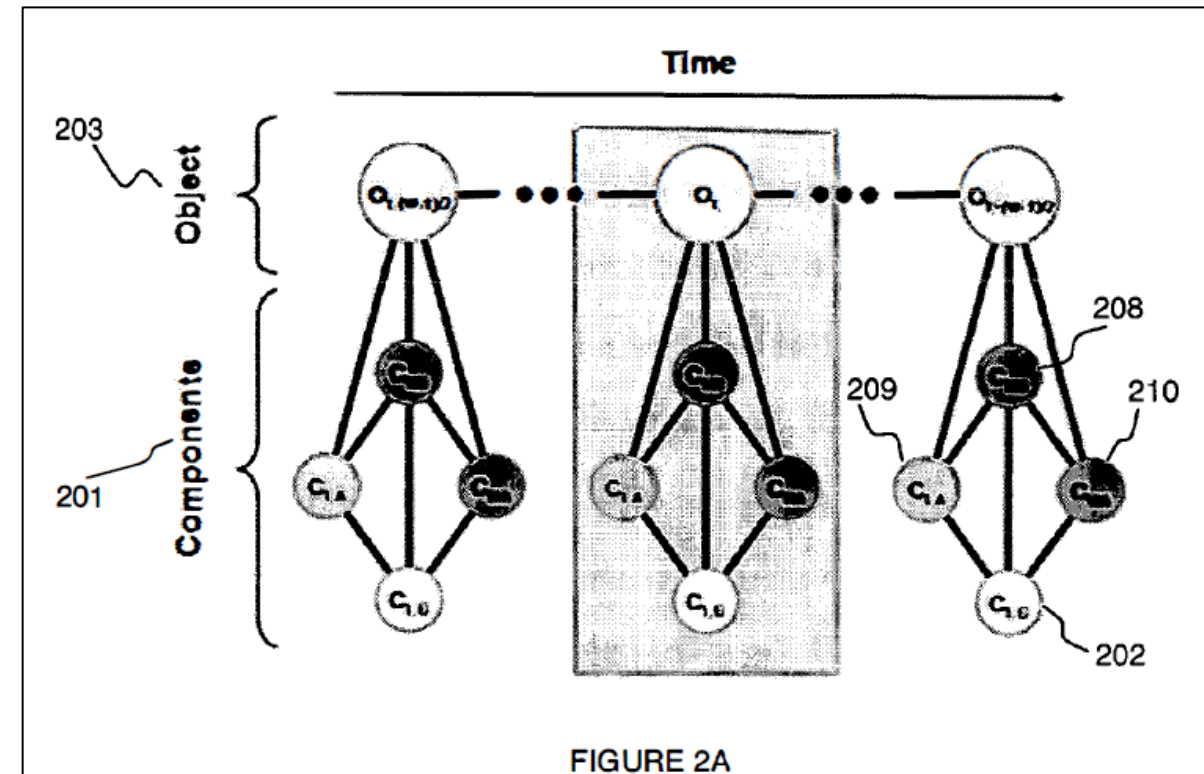
The output of the AdaBoost classifier is a confidence $h_k(I)$ that the given pattern I is of the desired class. An object is considered present if

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

EX1001 ('980 Patent) 5:58-67; Paper 26 (POR) at 5-6

The Improved Object Detection Technique of the '980 Patent

Non-parametric BP; Inferring the state of the object and its components in a framework is defined as estimating belief in a graphical model. A form of non-parametric belief propagation PAMPAS (proposed by M. Isard in "PAMPAS: Real-valued graphical models for computer vision", in the Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, vol. I, pp. 421-428, 2004.) is implemented to deal with this task. The approach is a generalization of particle filtering, which allows inference over arbitrary graphs rather than a simple chain. In this generalization the message used in belief propagation is approximated with a kernel density, formed by propagating a particle set through a mixture of Gaussians density, and the conditional distribution used in standard particle filtering is replaced by product of incoming messages. Most of the computational complexity lies in sampling from a product of kernel densities required for message passing and belief estimation; we use efficient sequential multiscale Gibbs sampling and epsilon-exact-sampling to address this problem.



EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 5-6

EX1001 ('980 Patent) at 6:23-41; Paper 26 (POR) at 5-6

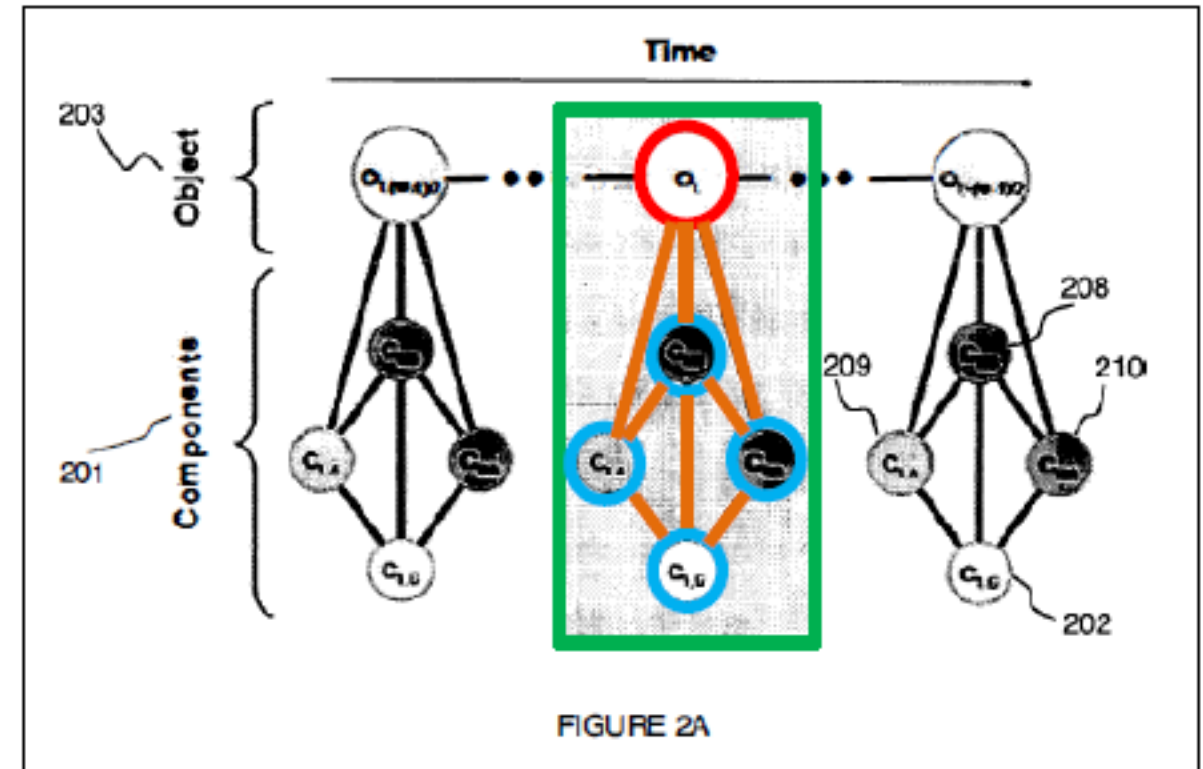
The Improved Object Detection Technique of the '980 Patent

Individual messages may not constrain a node well, however the product over all incoming messages into the node tends to produce a very tight distribution in the state space. For example, any given component of a vehicle is incapable of estimating the height of the vehicle reliably, however once information from all components is intergraded in the object layer node, an estimate for the overall object size can be determined.

More formally a message m_{ij} is written as

$$m_{ij}(X_j) = \int \psi_{ij}(X_i, X_j) \phi_i(X_i) \prod_{k \in \{A_i \setminus j\}} m_{ki}(X_i) dX_i \quad (3)$$

where A_i is the set of neighbors of node i , $\{A_i \setminus j\}$ is the set of neighbors of node i except node j , and $\phi_i(X_i)$ is the local evidence (or likelihood) associated with the node i , and $\psi_{ij}(X_i, X_j)$ is the potential designating the compatibility between the states of node i and j . The details of how the message updates can be carried out by stratified sampling from belief and proposal function is illustrated in FIG. 4.



EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 2-3

EX1001 ('980 Patent) 6:41-64; Paper 26 (POR) at 5-6

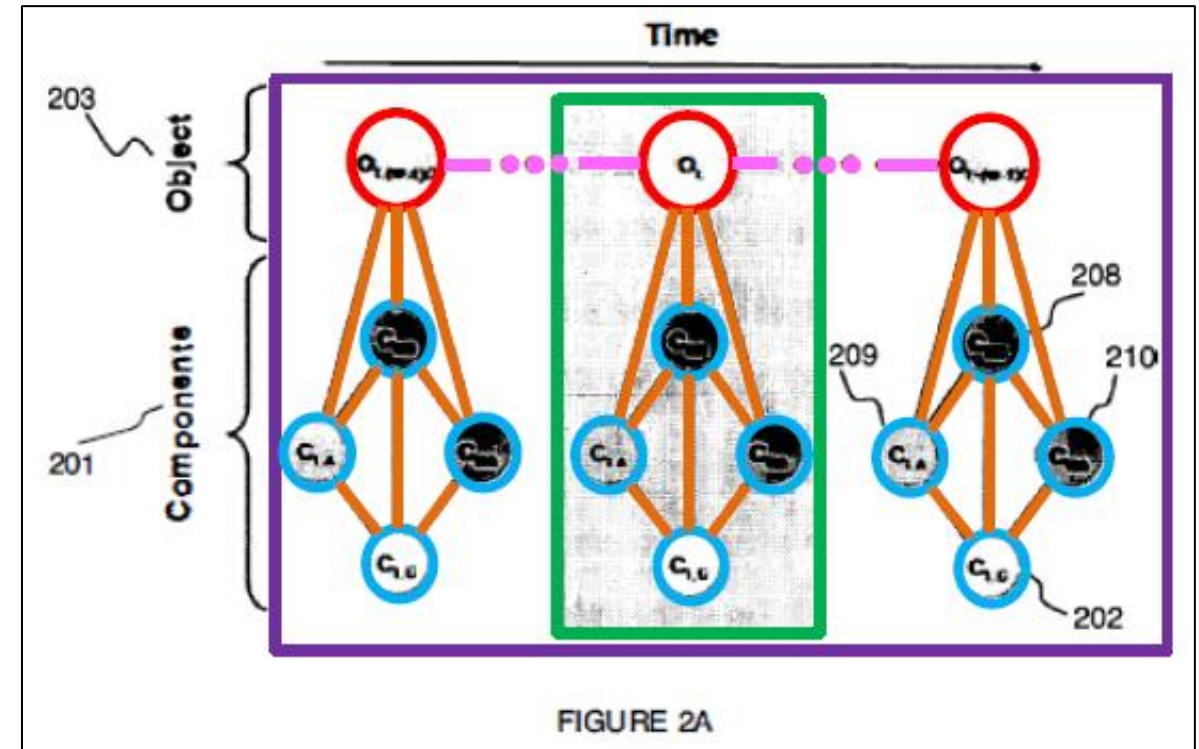
The Improved Object Detection Technique of the '980 Patent

Individual messages may not constrain a node well, however the product over all incoming messages into the node tends to produce a very tight distribution in the state space. For example, any given component of a vehicle is incapable of estimating the height of the vehicle reliably, however once information from all components is intergraded in the object layer node, an estimate for the overall object size can be determined.

More formally a message m_{ij} is written as

$$m_{ij}(X_j) = \int \psi_{ij}(X_i, X_j) \phi_i(X_i) \prod_{k \in \{A_i \setminus j\}} m_{ki}(X_i) dX_i \quad (3)$$

where A_i is the set of neighbors of node i , $\{A_i \setminus j\}$ is the set of neighbors of node i except node j , and $\phi_i(X_i)$ is the local evidence (or likelihood) associated with the node i , and $\psi_{ij}(X_i, X_j)$ is the potential designating the compatibility between the states of node i and j . The details of how the message updates can be carried out by stratified sampling from belief and proposal function is illustrated in FIG. 4.



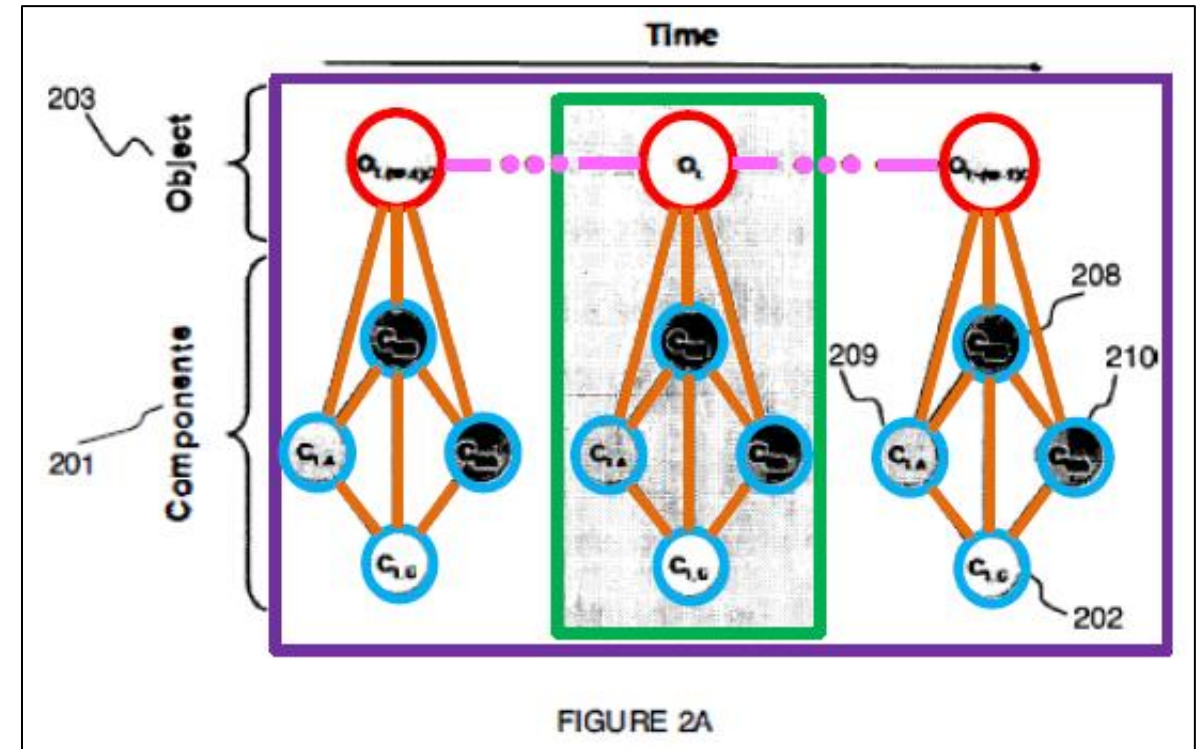
EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 3-4

EX1001 ('980 Patent) 6:41-64; Paper 26 (POR) at 5-6

The Improved Object Detection Technique of the '980 Patent

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.

EX1001 ('980 Patent) 7:6-22; Paper 26 (POR) at 6-7

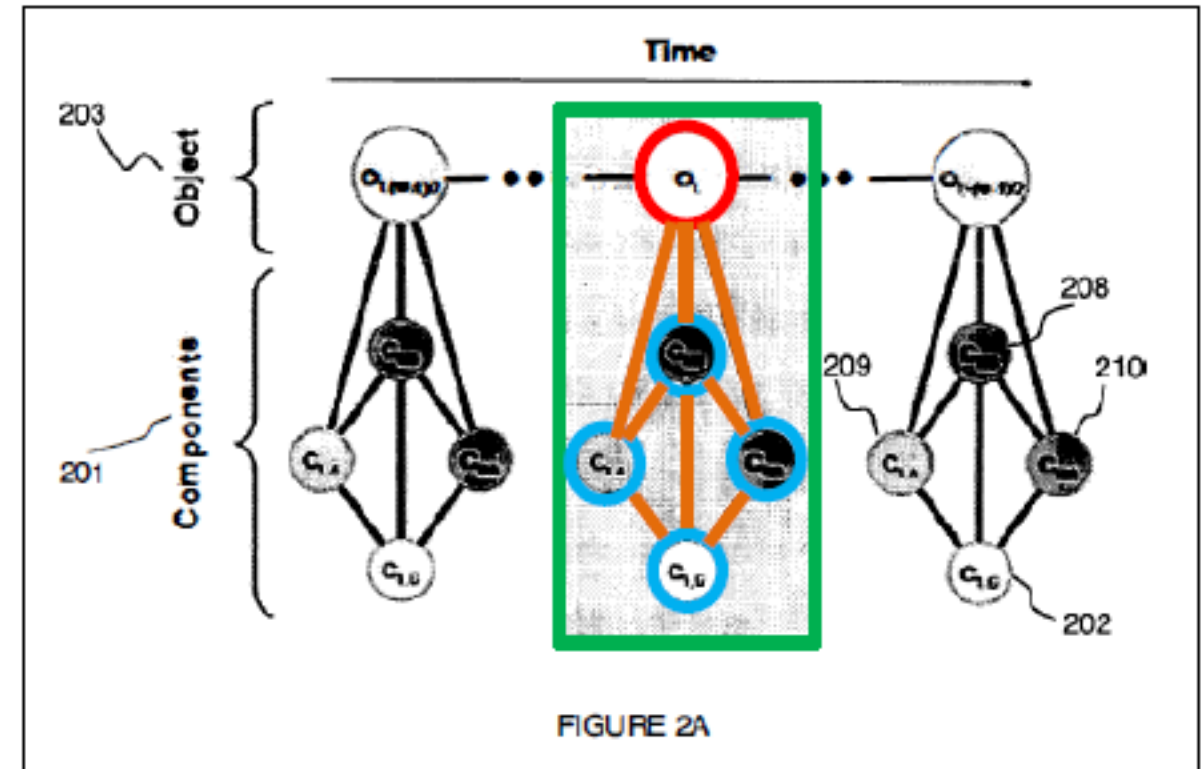


EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 3-4

The Improved Object Detection Technique of the '980 Patent

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.

EX1001 ('980 Patent) 7:6-22; Paper 26 (POR) at 6-7



EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 2-3

CLAIM CONSTRUCTION

Claim Construction

Claim 1

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

EX1001 ('980 patent), at claim 1



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

Term	Patent Owner’s Construction	Petitioners’ Construction ¹
“detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object”	detecting the object in any image based on comparing the probabilities for each image of the plurality of images to a threshold for detecting the object	detecting the collection of components in any image upon comparing the collection of component probabilities for each image to a threshold for detecting the collection of components

Paper 26 (POR) at 18-25; Paper 44 (Sur-Reply) at 3-15

¹ Petitioners do not formally propose this construction but apply it in their invalidity analysis. See Paper 32 (Reply) at 2, 4-5, 8-11, 25-27.

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

Element 1[e] requires a multi-image analysis.

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

EX1001 ('980 patent), at claim 1; Paper 26 (POR) at 18-20



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

Element 1[e] requires a multi-image analysis.

“We have explained that a patent’s express purpose of the invention ‘informs the proper construction of claim terms.’”

Sequoia Tech., LLC v. Dell, Inc., 66 F.4th 1317, 1326 (Fed. Cir. 2023)

“As we stated in *Renishaw PLC v. Marposs Societa’ per Azioni*, 158 F.3d 1243, 1250 (Fed.Cir.1998), ‘ultimately, the interpretation to be given a term can only be determined and confirmed with a full understanding of what the inventors actually invented and intended to envelop with the claim.’ It is therefore entirely proper to consider the functions of an invention in seeking to determine the meaning of particular claim language.”

Medrad, Inc. v. MRI Devices Corp., 401 F.3d 1313, 1319 (Fed. Cir. 2005)



Paper 26 (POR) at 20-21, 23; Paper 44 (Sur-Reply) at 18-19

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

Element 1[e] requires a multi-image analysis.

According to an embodiment of the present disclosure, a probabilistic framework for automatic component-based detection and tracking of objects in images and/or video combines object detection with tracking in a unified framework. Tracking makes use of object detection for initialization and re-initialization during transient failures for occlusions. Object detection considers the consistency of the detection over time. Modeling objects by an arrangement of image-base, and possibly overlapping, components facilitates detection of complex articulated objects as well as helps in handling partial object occlusions or local illumination changes.

EX1001 ('980 patent) at 2:43-54; Paper 26 (POR) at 20-21;
Paper 44 (Sur-Reply) at 18-19

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.

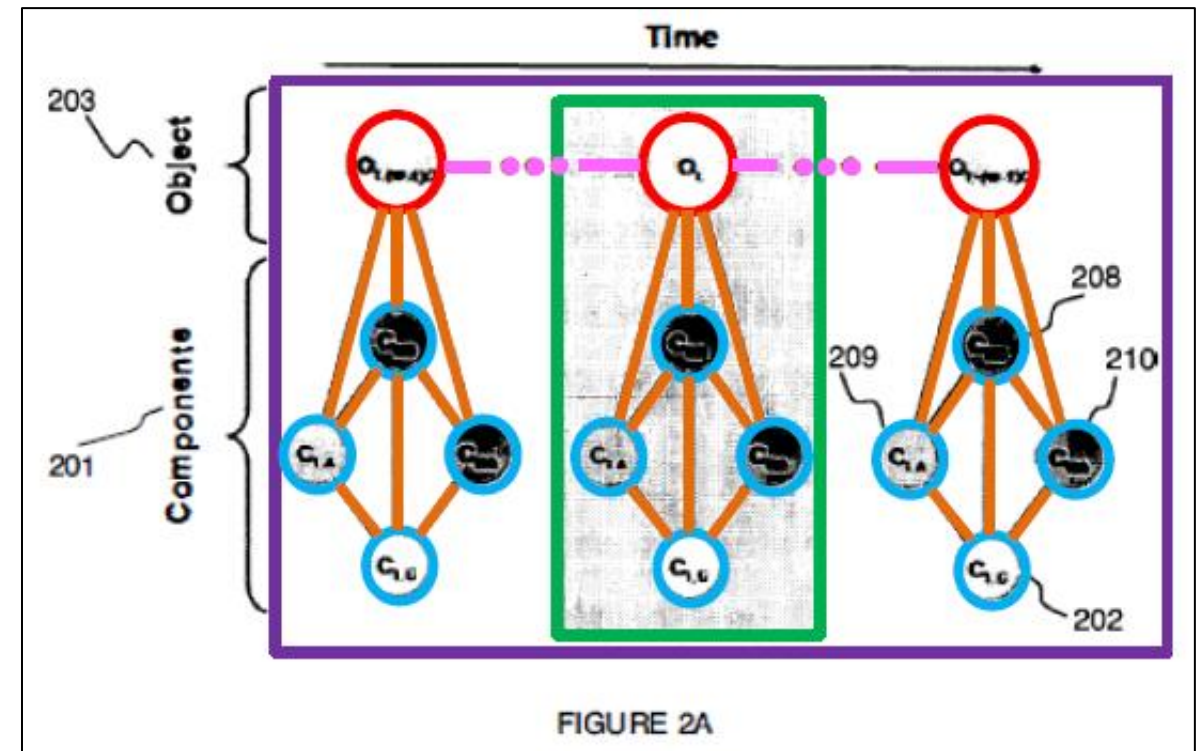
EX1001 ('980 patent) at 7:6-22; Paper 26 (POR) at 20-21; Paper 44 (Sur-Reply) at 18-19

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

Element 1[e] requires a multi-image analysis.

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.

EX1001 ('980 patent) at 7:6-22; Paper 26 (POR) at 20-21; Paper 44 (Sur-Reply) at 18-19



EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 3-4

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

Element 1[e] requires a multi-image analysis.

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 .

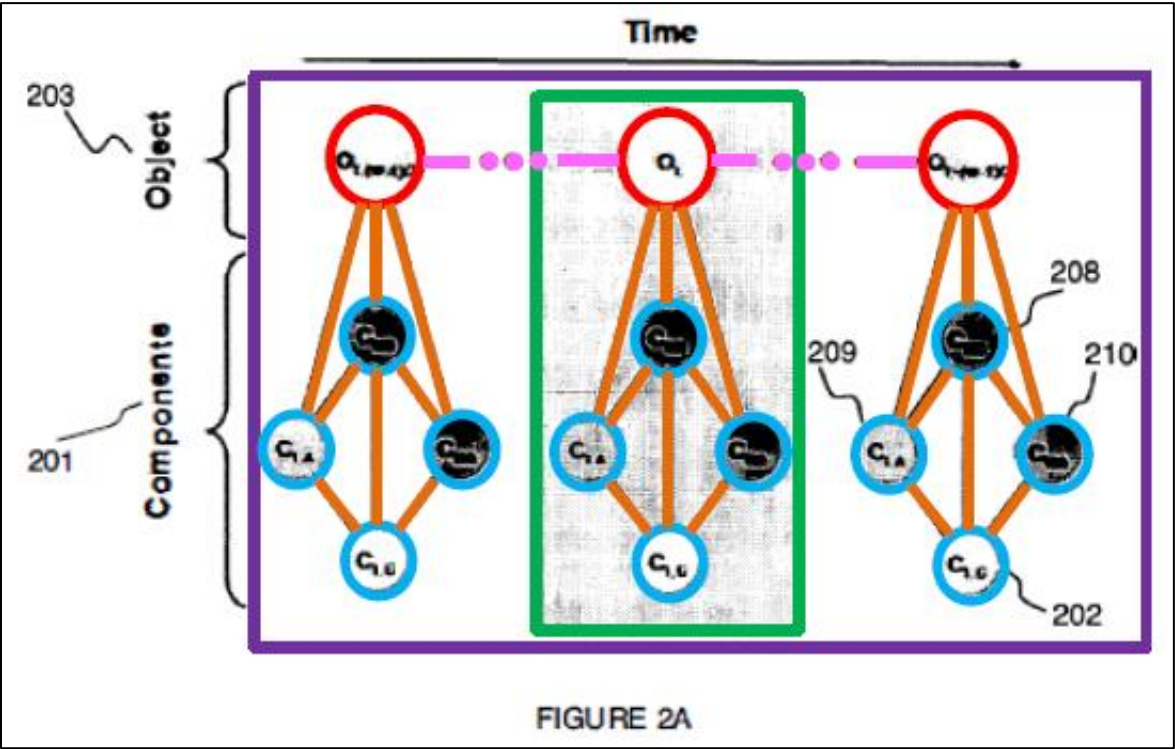
EX1001 ('980 patent) at 7:6-22; Paper 26 (POR) at 20-21; Paper 44 (Sur-Reply) at 18-19

Window size	Temporal distance backward: $t-(w-1)/2$	Temporal distance forward: $t+(w-1)/2$
W=3	$t-(3-1)/2 = t-(2)/2 = t-1$	$t+(3-1)/2 = t-(2)/2 = t+1$

Paper 26 (POR) at 24

As can be seen in FIGS. 2A and 2B, an object at a current time is denoted as o_t with an earlier model denoted as $o_{t-(w-1)/2}$ and a later model denoted as $o_{t+(w-1)/2}$.

EX1001 ('980 Patent) 4:45-48; Paper 26 (POR) at 24



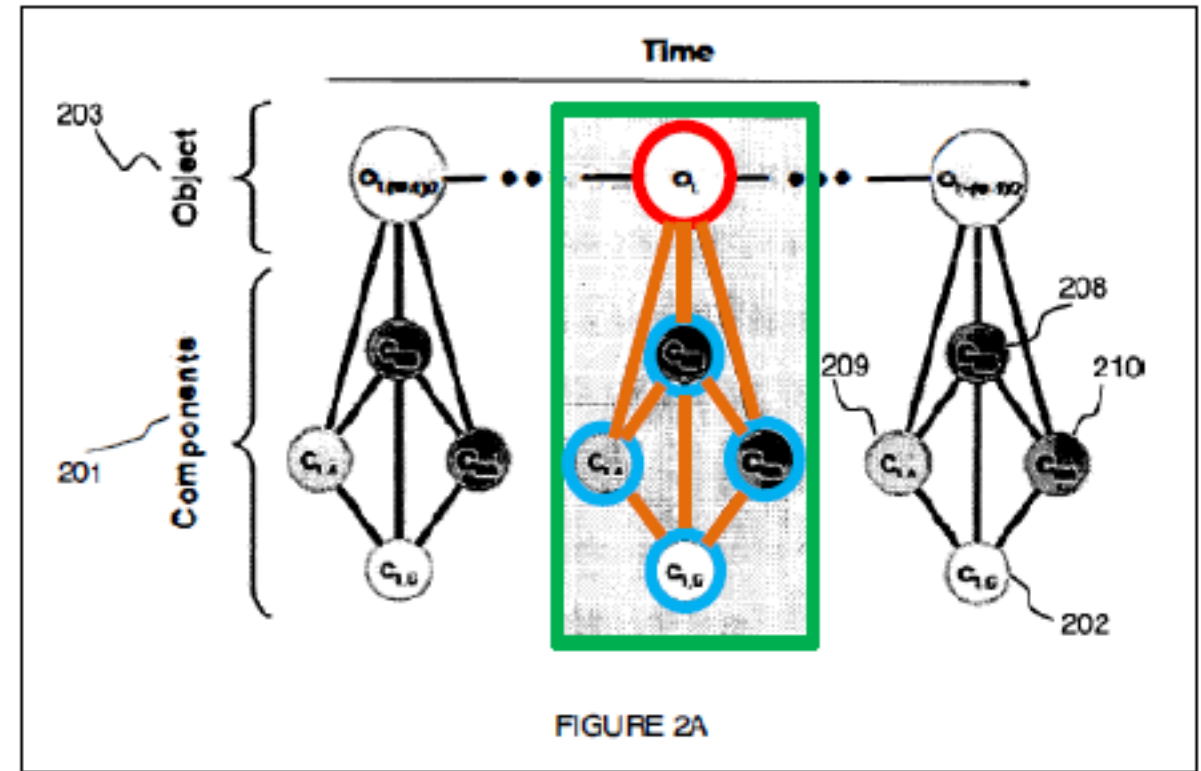
EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 3-4

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

“w=1” is not a multi-image analysis; it is not covered by claim 1

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.

EX1001 ('980 Patent) 7:6-22; Paper 26 (POR) at 22-23



EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 2-3

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

“w=1” is not a multi-image analysis; it is not covered by claim 1

Window size	Temporal distance backward: $t-(w-1)/2$	Temporal distance forward: $t+(w-1)/2$
W=1	$t-(1-1)/2 = t-(0)/2 = t-0$	$t+(1-1)/2 = t-(0)/2 = t+0$

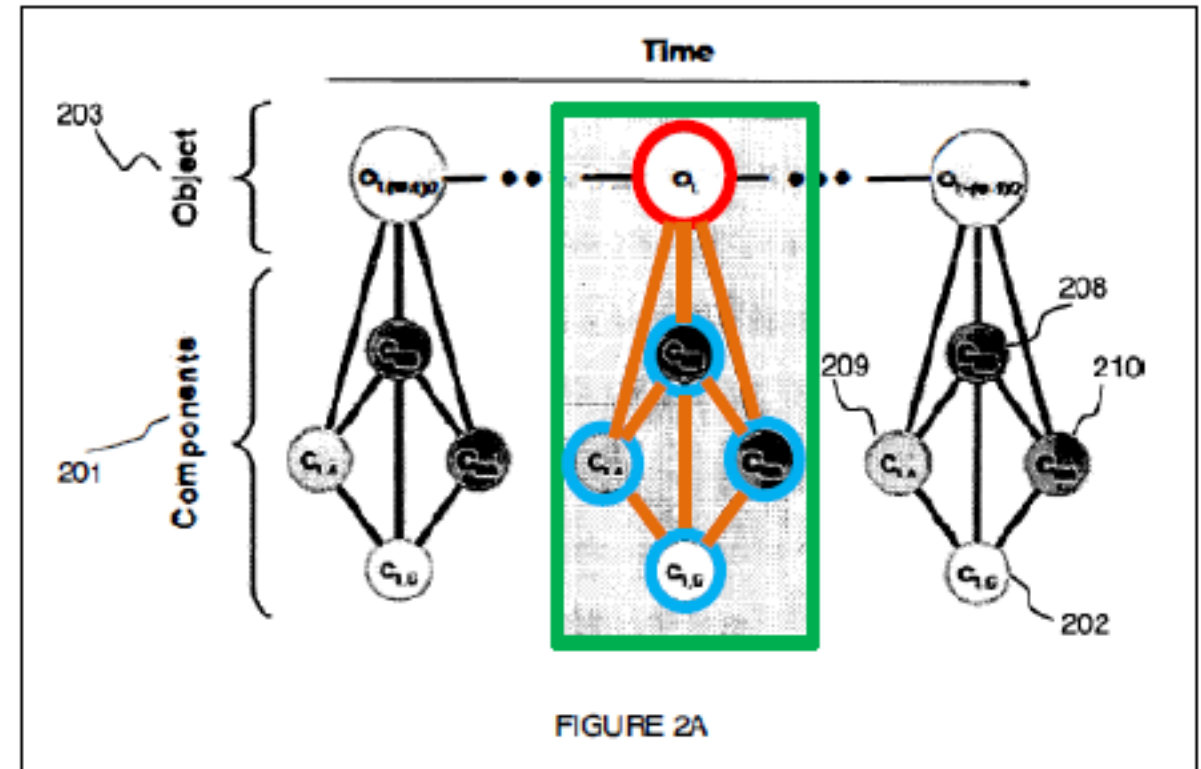
Paper 26 (POR) at 24

As can be seen in FIGS. 2A and 2B, an object at a current time is denoted as o_t with an earlier model denoted as $o_{t-(w-1)/2}$ and a later model denoted as $o_{t+(w-1)/2}$.

EX1001 ('980 Patent) 4:45-48; Paper 26 (POR) at 24

objects entering and leaving the scene. Note that with $w=1$, the algorithm resembles single frame component-based fusion.

EX1001 ('980 Patent) 7:6-22; Paper 26 (POR) at 23-24

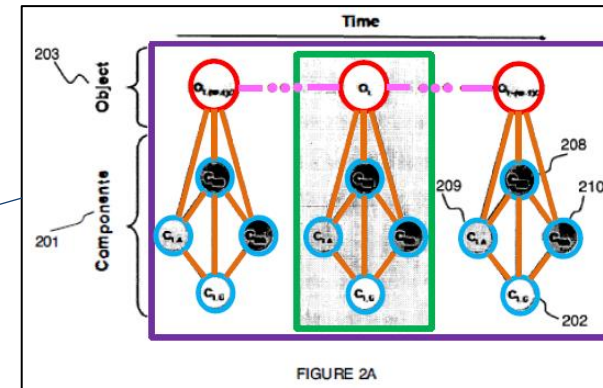


EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 2-3

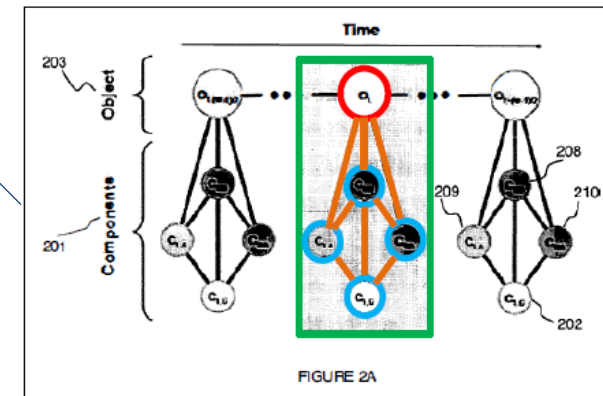
Claim Construction: 1/9[e] “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object”

The '980 patent distinguishes $w=1$ from the claimed invention

Experiments; Tests were performed using a set of images collected with a single car-mounted grayscale camera. The result of vehicle detection and tracking over a sequence of 55 consecutive frames was evaluated. A 3-frame spatio-temporal object model was used and was shifted in a tracking centric way over time. BP was run with 30 particles for 10 iterations at every frame. For comparison a simple fusion scheme was implemented that averages the best detection result from each of the four components to produce an estimate for the vehicle position and scale independently at every frame. The performance of the simple fusion detection is poor suggesting that the noisy component detectors often do not have the global maximum at the correct position and scale. In contrast, the spatio-temporal object model consistently combines the evidence for accurate estimates throughout the sequence.



EX1001 ('980 Patent) Fig. 2A;
Paper 26 (POR) at 3-4



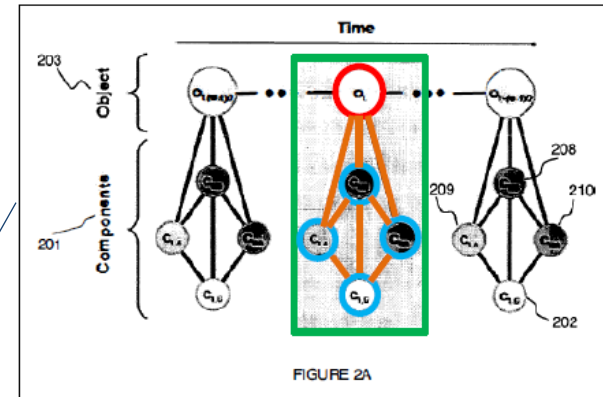
EX1001 ('980 Patent) Fig. 2A;
Paper 26 (POR) at 2-3

EX1001 ('980 Patent) 7:56-8:3; Paper 26 (POR) at 22-23

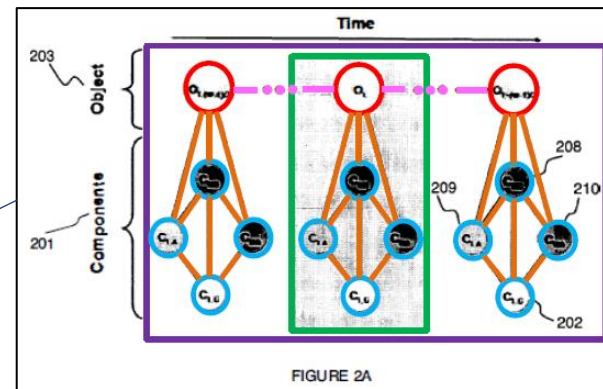
Claim Construction: 1/9[e] “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object”

The '980 patent distinguishes $w=1$ from the claimed invention

Experiments; Tests were performed using a set of images collected with a single car-mounted grayscale camera. The result of vehicle detection and tracking over a sequence of 55 consecutive frames was evaluated. A 3-frame spatio-temporal object model was used and was shifted in a tracking centric way over time. BP was run with 30 particles for 10 iterations at every frame. For comparison a simple fusion scheme was implemented that averages the best detection result from each of the four components to produce an estimate for the vehicle position and scale independently at every frame. The performance of the simple fusion detection is poor suggesting that the noisy component detectors often do not have the global maximum at the correct position and scale. In contrast, the spatio-temporal object model consistently combines the evidence for accurate estimates throughout the sequence.



EX1001 ('980 Patent) Fig. 2A;
Paper 26 (POR) at 2-3



EX1001 ('980 Patent) Fig. 2A;
Paper 26 (POR) at 3-4

EX1001 ('980 Patent) 7:56-8:3; Paper 26 (POR) at 22-23

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

Figure 5’s typo does not alter the construction.

“The actual words of the claim are the controlling focus.”

Digital Biometrics, Inc. v. Identix, Inc., 149 F.3d 1335, 1344 (Fed. Cir. 1998)

“[One] cannot alter the plain meaning of the claim language by referring to parts of the specification out of context.”

Seal-Flex, Inc. v. Athletic Track & Ct. Const., 172 F.3d 836, 845 (Fed. Cir. 1999)



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

Figure 5’s typo does not alter the construction.

detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object.

EX1001 ('980 patent) at 8:40-43 (claim 1[e]); Paper 26 (POR) at 21-22; Paper 44 (Sur-Reply) at 13

Referring to FIG. 5, a computer implemented method for object detection includes providing a spatio-temporal model, e.g., see FIGS. 2A and 2B, for an object 501, providing a video including a plurality of images including the object 502, and measuring the object as a collection of components in each image of the video 503. The method further includes determining a probability that the object is in each image 504 by using message passing between components represented as nodes of the spatio-temporal model, and detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object 505.

EX1001 ('980 patent) at 7:53-55; Paper 26 (POR) at 21-22; Paper 44 (Sur-Reply) at 13

505

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

FIGURE 5

EX1001 ('980 patent) at Fig. 5; Paper 26 (POR) at 20-21; Paper 44 (Sur-Reply) at 18-19

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

“upon” means “based on”

“Based on the full teaching of the specification, we conclude that Wi-Fi’s proposed construction of claim 15 is unreasonable. It would allow an erroneous sequence number length field to be present without an erroneous sequence number field, which the specification indicates would not work, while requiring all erroneous sequence number fields to be associated with erroneous sequence number length fields, which the patent teaches is not necessary. The Board’s construction, on the other hand, comports with what the patent teaches about the number and length fields. Even though the language of claim 15, standing alone, provides some support for Wi-Fi’s interpretation, we hold that in the end the claim must be read as the Board construed it in order to be faithful to the invention disclosed in the specification.”

Wi-Fi One, LLC v. Broadcom Corp., 887 F.3d 1329, 1346 (Fed. Cir. 2018)



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

“upon” means “based on”

According to an embodiment of the present disclosure, a probabilistic framework for automatic component-based detection and tracking of objects in images and/or video combines object detection with tracking in a unified framework. Tracking makes use of object detection for initialization and re-initialization during transient failures for occlusions. Object detection considers the consistency of the detection over time. Modeling objects by an arrangement of

EX1001 ('980 patent) at 2:43-50; Paper 26 (POR) at 24-25

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 $w=1$,

EX1001 ('980 patent) at 7:6-22; Paper 26 (POR) at 24-25

71. In light of the specification, a POSITA would have also understood that detecting the object in any image “**upon**” comparing the probabilities for each image to a threshold for detecting the object means detecting the object “**based on**” comparing the probabilities to a threshold. A POSITA would have understood the claim this way because the patent teaches basing the detection of an object on whether the object is determined to be present in other images, which is how it achieves “consistency of the detection over time” and produces “temporal consistency” over multiple image frames and “consistent results over time.” EX1001 at 2:43-54; 7:6-22. For example, the patent teaches that “object detection”

EX2004 (Saber Decl.) ¶71; Paper 26 (POR) at 24-25

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

“upon” means “based on”

Referring to FIGS. 2A and 2B, object detection and tracking is formulated as an inference in a two-layer graphical model in which a coarse layer node represents the whole object and fine layer nodes represent multiple components of the object. Directed edges between nodes represent learned spatial and temporal probabilistic constraints. Each node in the graphical model corresponds to a position and scale of the component or the object as a whole in an image at a given time instant. Each node also has an associated Adaptive Boosting

EX1001 ('980 patent) at 2:55-63; Paper 26 (POR) at 24-25

To integrate temporal constraints the spatial graphical models are extended over time to an arbitrary length temporal window. The resulting spatio-temporal graphical models are shown in FIGS. 2A and 2B. Having a two-layer graphical model allows the inference process to reason explicitly about the object as a whole, e.g., 203, and reduce the complexity of the graphical model by allowing the assumption of the conditional independence of components, e.g., 202 and 208-210, over time given the overall object appearance. Alternatively, a single object layer model can be built with bi-directional temporal constraints.

EX1001 ('980 patent) at 4:34-44; Paper 26 (POR) at 24-25

EX1001 at 2:43-54; 7:6-22. For example, the patent teaches that “object detection” is “formulated as an inference” in a model. EX1001 at 2:55-3:2. The model is a “spatio-temporal model” for object detection that covers more than “a single frame” because it is “extended over time” to multiple frames, thus it includes multiple instances of the object at a “current time” as well as at “earlier” or “later” times.

EX1001 at 4:18-48, Figs. 2A and 2B; see § VI of my analysis above. The patent

EX2004 (Saber Decl.) ¶71; Paper 26 (POR) at 24-25

As can be seen in FIGS. 2A and 2B, an object at a current time is denoted as o_t with an earlier model denoted as $o_{t-(w-1)/2}$ and a later model denoted as $o_{t+(w-1)/2}$.

EX1001 ('980 patent) at 4:45-48; Paper 26 (POR) at 24-25

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

“upon” means “based on”

AdaBoost Image Likelihoods; A likelihood $\phi_i(X_i)$ models the probability of observing the image conditioned on the state X_i of the node i . A likelihood model according to an embodiment of the present disclosure uses a boosted classifier and is robust to partial occlusions and the variability of image statistics across many different inputs.

EX1001 ('980 Patent) at 5:25-30; Paper 26 (POR) at 24-25

The output of the AdaBoost classifier is a confidence $h_k(I)$ that the given pattern I is of the desired class. An object is considered present if

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

EX1001 ('980 Patent) 5:58-67; Paper 26 (POR) at 24-25

Proposal Process; To reliably detect and track the object non-parametric BP makes use of the bottom-up proposal process, that constantly looks for and suggests alternative hypothesis for the state of the object and components. Proposal distribution is modeled using a weighted particle set. To form a proposal particle set for a component, a corresponding AdaBoost detector is run over an image at a number of scales to produce a set of detection results that score above the

$$\frac{1}{2} \sum_{k=1}^K \alpha_k$$

threshold. While this set tends to be manageable for the entire

EX1001 ('980 Patent) 7:23-44; Paper 26 (POR) at 24-25

EX1001 at 4:18-48, Figs. 2A and 2B; see § VI of my analysis above. The patent teaches that the detection inference is defined as an estimation of “belief” in the model over a w -frame window, performed by comparing the likelihood of object presence in each image to a threshold and sending “messages” based on that likelihood for each image (e.g., when it scores above the threshold) to the object in the image under consideration for detecting the object. EX1001 at 5:25-7:44; see § VI of my analysis above. A POSITA would have understood that by this process, the detection of an object in an image is **based on** the probabilities of the object being in other images (as determined by comparing the probabilities to a threshold).

EX2004 (Saber Decl.) ¶71; Paper 26 (POR) at 24-25

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

Non-parametric BP; Inferring the state of the object and its components in a framework is defined as estimating belief in a graphical model. A form of non-parametric belief propaga-

EX1001 ('980 Patent) at 6:23-41; Paper 26 (POR) at 24-25

implemented to deal with this task. The approach is a generalization of particle filtering, which allows inference over arbitrary graphs rather than a simple chain. In this generalization the message used in belief propagation is approximated with a kernel density, formed by propagating a particle set through a mixture of Gaussians density, and the conditional distribution used in standard particle filtering is replaced by product of incoming messages. Most of the com-

EX1001 ('980 Patent) at 6:23-41; Paper 26 (POR) at 24-25

More formally a message m_{ij} is written as

$$m_{ij}(X_j) = \int \psi_{ij}(X_i, X_j) \phi_i(X_i) \prod_{k \in \{A_i \setminus j\}} m_{ki}(X_i) dX_i \quad (3)$$

where A_i is the set of neighbors of node i , $\{A_i \setminus j\}$ is the set of neighbors of node i except node j , and $\phi_i(X_i)$ is the local evidence (or likelihood) associated with the node i , and $\psi_{ij}(X_i, X_j)$ is the potential designating the compatibility between the states of node i and j . The details of how the message updates can be carried out by stratified sampling from belief and proposal function is illustrated in FIG. 4.

EX1001 ('980 Patent) at 6:42-64; Paper 26 (POR) at 24-25

“upon” means “based on”

EX1001 at 4:18-48, Figs. 2A and 2B; see § VI of my analysis above. The patent teaches that the detection inference is defined as an estimation of “belief” in the model over a w -frame window, performed by comparing the likelihood of object presence in each image to a threshold and sending “messages” based on that likelihood for each image (e.g., when it scores above the threshold) to the object in the image under consideration for detecting the object. EX1001 at 5:25-7:44; see § VI of my analysis above. A POSITA would have understood that by this process, the detection of an object in an image is **based on** the probabilities of the object being in other images (as determined by comparing the probabilities to a threshold).

EX2004 (Saber Decl.) ¶71; Paper 26 (POR) at 24-25

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

“upon” means “based on”

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.

EX1001 ('980 patent) at 4:34-44; Paper 26 (POR) at 24-25

EX1001 at 4:18-48, Figs. 2A and 2B; see § VI of my analysis above. The patent teaches that the detection inference is defined as an estimation of “belief” in the model over a w-frame window, performed by comparing the likelihood of object presence in each image to a threshold and sending “messages” based on that likelihood for each image (e.g., when it scores above the threshold) to the object in the image under consideration for detecting the object. EX1001 at 5:25-7:44; see § VI of my analysis above. A POSITA would have understood that by this process, the detection of an object in an image is **based on** the probabilities of the object being in other images (as determined by comparing the probabilities to a threshold).

EX2004 (Saber Decl.) ¶71; Paper 26 (POR) at 24-25

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

“upon” means “based on”

Petitioners do not contest PO’s evidence.

Furthermore, because of the **strict prohibition against the presentation of new evidence or arguments at a hearing**, it is strongly recommended that each demonstrative include a citation to a paper in the record, which allows the Board to easily ascertain whether a given demonstrative contains “new” argument or evidence or, instead, contains only that which is developed in the existing record.

Paper 45 at 4



“We need not reach the merits of these arguments because Bradium waived them by failing to present them to the Board.”

Bradium Techs. LLC v. Iancu, 923 F.3d 1032, 1048 (Fed. Cir. 2019)



Paper 44 (Sur-Reply) at 14-15

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

Petitioners’ construction rewrites “object” as “collection of components”

13. These claim 1 and 9 limitations clearly specify image processing steps, and on an image frame by image frame basis, consisting of measuring the **object as a collection of components**, determining a probability that the object is present, and **detecting the (presence of the) object in any image by comparing the (collection of component) probabilities for each image to a threshold.** The prior limitations recite:

1[a]/9[a]: providing a spatio-temporal model for an object to be detected;

1[b]/9[b]: providing a video comprising a plurality of images including the object;

EX1052 (Bajaj Supp. Decl.) ¶13; Paper 44 (Sur-Reply) at 3-4

22 So objects are always a collection
1 of components. **And then determining a**
2 **probability that the object -- you can always put**
3 **parentheses again -- of collection of components**
4 **because the object is always a collection of**
5 **components.** And then detecting that object which
6 is again a collection of components is upon
7 comparing the probabilities for each image to a
8 threshold for detecting the object.

EX2008 (Bajaj Dep. Tr.) at 18:2-19:5-21; Paper 44 (Sur-Reply) at 3-4

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

Petitioners’ construction rewrites “object” as “collection of components”

13. These claim 1 and 9 limitations clearly specify image processing steps, and on an image frame by image frame basis, consisting of measuring the **object as a collection of components**, determining a probability that the object is present, and **detecting the (presence of the) object in any image by comparing the (collection of component) probabilities for each image to a threshold.** The prior limitations recite:

1[a]/9[a]: providing a spatio-temporal model for an object to be detected;

1[b]/9[b]: providing a video comprising a plurality of images including the object;

EX1052 (Bajaj Supp. Decl.) ¶13; Paper 44 (Sur-Reply) at 3-4

15 The object is the collection of
16 components in each image. You determine a
17 probability that the object is in each image and
18 detect the object in any image upon comparing the
19 probabilities. **Since the object is a collection**
20 **of components, there is a probability for each of**
21 **the components.**
22 And the joint probability, as you
1 learn later on, as far as that claim is
2 concerned, **it's comparing the probabilities of**
3 **each of the collection of components in that**
4 **image to a threshold.**

EX2008 (Bajaj Dep. Tr.) at 21:8-22:4; Paper 44 (Sur-Reply) at 3-4

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

Petitioners’ construction rewrites “object” as “collection of components”

13. These claim 1 and 9 limitations clearly specify image processing steps, and on an image frame by image frame basis, consisting of measuring the **object as a collection of components**, determining a probability that the object is present, and **detecting the (presence of the) object in any image by comparing the (collection of component) probabilities for each image to a threshold.** The prior limitations recite:

1[a]/9[a]: providing a spatio-temporal model for an object to be detected;

1[b]/9[b]: providing a video comprising a plurality of images including the object;

EX1052 (Bajaj Supp. Decl.) ¶13; Paper 44 (Sur-Reply) at 3-4

5 Q. And when you said just now
6 "comparing the probabilities to a threshold,"
7 you're saying your interpretation of the claim is
8 that the probabilities are probabilities for
9 components, right?

10 A. Comparing the probabilities for the
11 collection of components because that's what
12 constitutes the object in each image.

EX2008 (Bajaj Dep. Tr.) at 24:5-12; Paper 44 (Sur-Reply) at 3-4

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

Petitioners’ construction rewrites “object” as “collection of components”

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.

EX1052 (Bajaj Supp. Decl.) ¶16; Paper 44 (Sur-Reply) at 4

17. According to the '980 specification (at 5:32), boosted detectors are trained for each component N using AdaBoost. Each component N has an associated weak classifier and the AdaBoost procedure learns a weighted combination of the base weak classifiers. The output of the AdaBoost classifier is a confidence that the object is present in an image where an object is considered present if the determined confidence is greater than or equal to the threshold recited at 5:65 of the '980 Patent. The confidence is converted to a likelihood function that the object is detected in each image. See '980 Patent at 6:1-3. Thus, 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].

EX1052 (Bajaj Supp. Decl.) ¶17; Paper 44 (Sur-Reply) at 4



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

Petitioners’ construction rewrites “object” as “collection of components”

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

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

EX1001 ('980 patent), claim 8; Paper 26 (POR) at 5-6, 38-39

7 Q. Okay. And so since we're talking
8 about a potential function in the superscript
9 "O," we're just talking about the object, not the
10 ones that are superscript "C," right?
11 A. The object is the sum of its parts.

EX2008 (Bajaj Dep. Tr.) at 90:7-11; Paper 44 (Sur-Reply) at 3-4

19 Q. So the difference between the object
20 state XO and the component state XC, those are
21 different states, right?
22 A. Not at any point in time. The
1 object is the collection of components.

EX2008 (Bajaj Dep. Tr.) at 92:19-93:1; Paper 44 (Sur-Reply) at 3-4

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

Petitioners’ constructions are improperly incorporated by reference

“Thus, parties that incorporate expert testimony by reference in their petitions, motions, or replies without providing explanation of such testimony risk having the testimony not considered by the Board. *See Cisco Systems, Inc. v. C-Cation Techs., LLC*, Case IPR2014-00454 (PTAB Aug. 29, 2014) (Paper 12) (informative).”

PTAB Consolidated Trial Practice Guide (Nov. 2019) at 35-36.



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

Petitioners’ constructions are improperly incorporated by reference

Contrary to PO and its expert, Dr. Saber, the '980 Patent is *not* “distinguish[ing] the w=1 scenario from the invention” (POR, 22), but generally describing ways of “performing inference over space and time” with and without “lengthy off-line processing.” Ex1001, 1:22-24, 7:6-9; Ex1052, ¶¶23-35; Paper-

Paper 32 (Reply) at 2

model.” Ex1002, 163. Therefore, the Examiner understood Cham to disclose element 1/9[e] as claimed, and that the Examiner therefore understood 1/9[e] to require a frame by frame comparison of a threshold to the probability that the object is in the respective frame. Ex1052, ¶¶6-22; cf. *Neville v. Found.*

Paper 32 (Reply) at 4

generate redefined contours. *Id.*, 31-32. Not so, and PO ignores Zhao’s teachings as well as the knowledge of a POSITA. Ex1052, ¶¶6-22, 23-35.

Paper 32 (Reply) at 9

specification, where w is an odd integer ≥ 1 (greater than *or equal to* 1). It is therefore incorrect that PO argues a scenario where w=1 is excluded from the purported invention. Dr. Bajaj’s opinion, unlike PO’s, is not based on an incorrect interpretation of the w=1 embodiment. Ex1052, ¶¶23-35.

Paper 32 (Reply) at 11

Paper 44 (Sur-Reply) at 4 (citing Reply, 2, 4-5, 8-11, 25-27)



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

Petitioners’ constructions are forfeited

“Petitioner is required to identify—in the Petition—how the challenged claims are to be construed. 37 C.F.R. § 42.104(b)(3). Presenting a new proposed claim construction for the first time in a Reply does not fulfill this requirement.”

GCE Gas Control Equipment Inc. v. VBOX, Inc., IPR2023-00326, Paper 27 at 11 (PTAB Jun. 18, 2024).

“In its Reply, Petitioner changed its position, arguing for the first time that ‘the claims do not require safety; they require only ‘treating’ Cushing’s syndrome or its symptoms.’ Reply 14. Not only is this an inappropriate new argument (see Pet. 23 (arguing that the claims should be given their plain and ordinary meaning and that “no further claim construction is required for purposes of this PGR”))), it contradicts the obviousness rationale set forth in the Petition.”

Teva Pharms. USA, Inc. v. Corcept Therapeutics, Inc., PGR2019-00048, Paper 51 at 17 (PTAB Nov. 18, 2020)



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

Petitioners’ constructions are forfeited

IX. CLAIM CONSTRUCTION

For purposes of this petition only, Petitioners present no terms for construction. Although the parties have not reached the claim construction stage in

Paper 3 (Petition) at 12

13. These claim 1 and 9 limitations clearly specify image processing steps, and on an image frame by image frame basis, consisting of measuring the **object as a collection of components**, determining a probability that the object is present, and detecting the (presence of the) object in any image by comparing the **(collection of component) probabilities** for each image to a threshold. The prior

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.**

Paper 44 (Sur-Reply) at 5

EX1052, ¶¶13-14 (cited in Paper 32 (Reply) at 2, 4-5, 8-9, 25-27)



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

Petitioners’ construction is inconsistent with the claims; it renders a term superfluous

1. A computer implemented method for **object detection** comprising:

[1a] providing a spatio-temporal model **for an object to be detected**;

[1b] providing a video comprising a plurality of images including **the object**;

[1c] **measuring the object** as a **collection of components** in each image;

[1d] determining a **probability that the object** is in each image; and

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

“There is no dispute that the ‘pressurized collection vessel’ receives the dispersion; that function is clearly contemplated by the surrounding claim language. But allowing ‘collection’ to mean ‘receive’ would render ‘collection’ **entirely superfluous** and allow any pressurized vessel to constitute a ‘pressurized collection vessel’; **such a result is disfavored**.”

Akzo Nobel Coatings, Inc. v. Dow Chem. Co., 811 F.3d 1334, 1340 (Fed. Cir. 2016) (cleaned up).

“[W]hile **interpretations that render some portion of the claim language superfluous are disfavored**, where neither the plain meaning nor the patent itself commands a difference in scope between two terms, they may be construed identically.”

Power Mosfet Techs., L.L.C. v. Siemens AG, 378 F.3d 1396, 1410 (Fed. Cir. 2004).



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

Petitioners’ construction is inconsistent with the claims; it renders a term superfluous

1. A computer implemented method for **object detection** comprising:

[1a] providing a spatio-temporal model **for an object to be detected**;

[1b] providing a video comprising a plurality of images including **the object**;

[1c] **measuring the object** as a **collection of components** in each image;

[1d] determining a **probability that the object** is in each image; and

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

“Starting with the language of claim 21, the terms ‘engaging’ and ‘sealing’ are **both expressly recited** in the claim and **therefore ‘engaging’ cannot mean the same thing as ‘sealing’**; if it did, one of the terms would be superfluous.”

Primos, Inc. v. Hunter’s Specialties, Inc., 451 F.3d 841, 848 (Fed. Cir. 2006).

“In the absence of any evidence to the contrary, **we must presume that the use of these different terms in the claims connotes different meanings.**”

CAE Screenplates, Inc. v. Heinrich Fiedler GmbH & Co. KG, 224 F.3d 1308, 1317 (Fed. Cir. 2000)



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

Petitioners’ construction is inconsistent with the claims; it renders terms indistinct

1. A computer implemented method for **object detection** comprising:

[1a] providing a spatio-temporal model **for an object to be detected**;

[1b] providing a video comprising a plurality of images including **the object**;

[1c] **measuring the object** as a **collection of components** in each image;

[1d] determining a **probability that the object** is in each image; and

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

“If the terms ‘pusher assembly’ and ‘pusher bar’ described a single element, one would expect the claim to consistently refer to this element as either a ‘pusher bar’ or a ‘pusher assembly,’ but not both, especially **not within the same clause**. Therefore, in our view, the plain meaning of the claim will not bear a reading that ‘pusher assembly’ and ‘pusher bar’ are synonyms.”

Ethicon Endo-Surgery, Inc. v. United States Surgical Corp., 93 F.3d 1572, 1579 (Fed. Cir. 1996).

“In these recitations, the ‘user’ and the ‘computer’ are **distinct entities**. To construe the term ‘user’ to mean a ‘computer’ would result in the claim being interpreted to recite, for example, ‘enabling the software on a computer for use by a [computer].’ **The language of the claims does not reasonably or logically permit such a construction.**”

z4 Technologies, Inc. v. Microsoft Corp., 507 F.3d 1340, 1348 (Fed. Cir. 2007)



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

Petitioners’ construction is inconsistent with the specification

“Likewise, the written description makes clear that the terms ‘users’ and ‘computers’ are distinct and used to describe different things.”

z4 Technologies, Inc. v. Microsoft Corp., 507 F.3d 1340, 1348 (Fed. Cir. 2007)



To integrate temporal constraints the spatial graphical models are extended over time to an arbitrary length temporal window. The resulting spatio-temporal graphical models are shown in FIGS. 2A and 2B. Having a two-layer graphical model allows the inference process to reason explicitly about the object as a whole, e.g., 203, and reduce the complexity of the graphical model by allowing the assumption of the conditional independence of components, e.g., 202 and 208-210 over time given the overall object appearance. Alternatively, a single object layer model can be built with bi-directional temporal constraints.

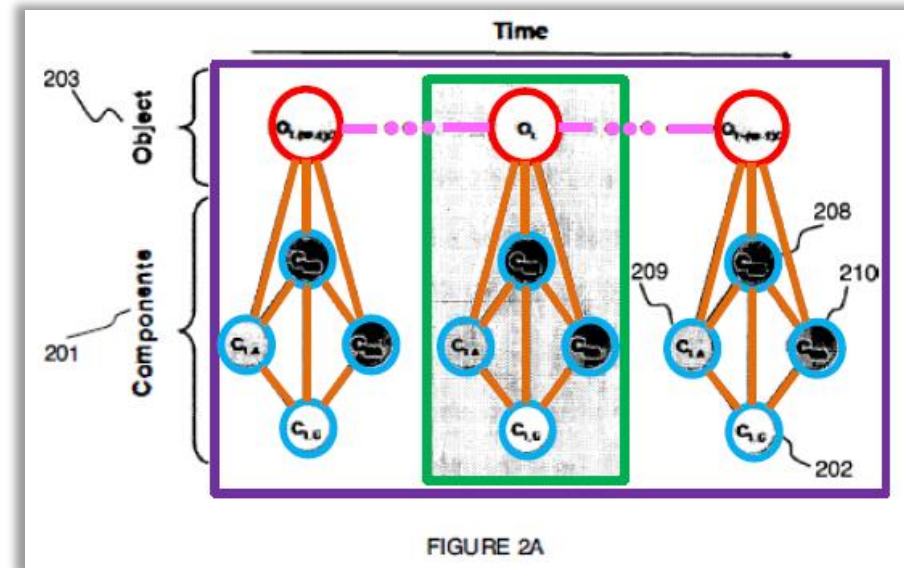


FIGURE 2A

EX1001 ('980 Patent) 4:34-45; Paper 44 (Sur-Reply) at 6-7

EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 3-4

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

Petitioners’ construction is inconsistent with the specification

“Likewise, the written description makes clear that the terms ‘users’ and ‘computers’ are **distinct and used to describe different things.**”

z4 Technologies, Inc. v. Microsoft Corp., 507 F.3d 1340, 1348 (Fed. Cir. 2007)



3). The joint probability distribution for this spatio-temporal graphical object model with N components can be written as:

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

where X_t^O and $X_t^{C_n}$ is the state of the object, O, and object’s n-th component, C_n, at time t respectively ($n \in (1, N)$ and t

$\in (1, T)$; $\psi_{ik}(X_i^O, X_j^O)$ is the temporal compatibility of object state between frames i and j; $\psi_{ik}(X_i^O, X_i^{C_k})$ is the spatial compatibility of the object and its components at frame i; $\psi_{ik}(X_i^{C_k}, X_i^{C_l})$ is the spatial compatibility between object components at frame i and $\phi_i(X_i^O)$ and $\phi_i(X_i^{C_n})$ denote the local evidence for the object and component states respectively.

EX1001 ('980 Patent) 3:47-4:7; Paper 44 (Sur-Reply) at 6-7

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

Petitioners’ construction reads out all preferred embodiments (w>1) in favor of a disfavored embodiment (w=1)

“We are mindful to not limit claims to a preferred embodiment. See *Teleflex, Inc. v. Ficosa N. Am. Corp.*, 299 F.3d 1313, 1328 (Fed. Cir. 2002). But we also recognize that “[a] claim construction exclud[ing] a preferred embodiment is rarely, if ever correct.” *Kaufman v. Microsoft Corp.*, 34 F.4th 1360, 1372 (Fed. Cir. 2022) (cleaned up). Here, we do not limit the claim language based on the preferred embodiment. Instead, we recognize that it aligns with, and thus bolsters, what the plain claim language indicates: that the disk partition is a logical volume's minimum construction unit.”

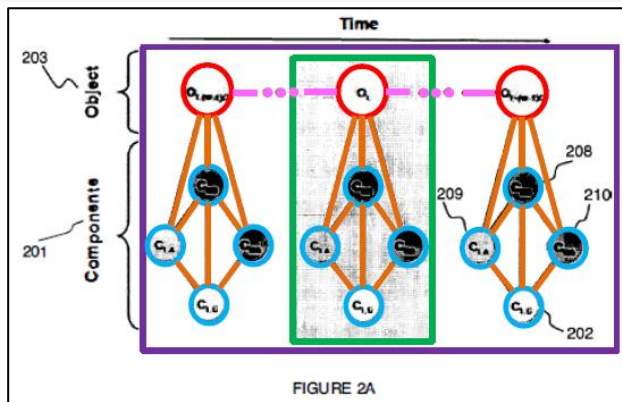
Sequoia Tech., LLC v. Dell, Inc., 66 F.4th 1317, 1327 (Fed. Cir. 2023)



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

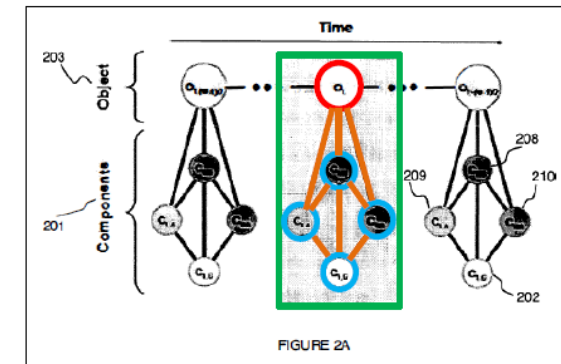
Petitioners’ construction reads out all preferred embodiments (w>1) in favor of a disfavored embodiment (w=1)

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



EX1001 ('980 Patent) Fig. 2A;
Paper 26 (POR) at 3-4

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



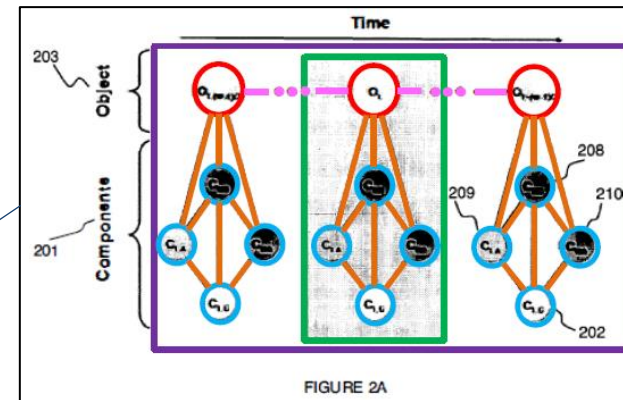
EX1001 ('980 Patent) Fig. 2A;
Paper 26 (POR) at 2-3

Paper 44 (Sur-Reply) at 7-8

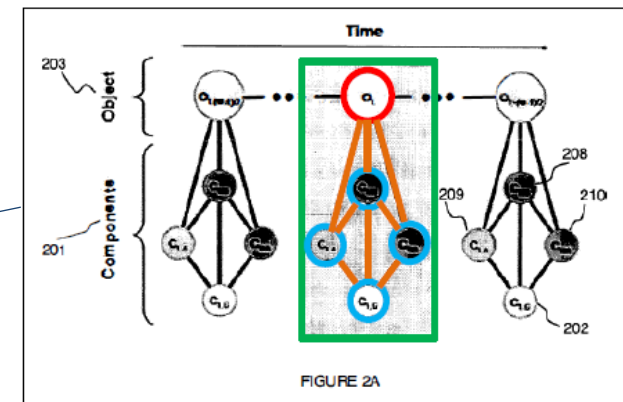
Claim Construction: 1/9[e] “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object”

Petitioners’ construction reads out all preferred embodiments ($w > 1$) in favor of a disfavored embodiment ($w = 1$)

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.



EX1001 ('980 Patent) Fig. 2A;
Paper 26 (POR) at 3-4



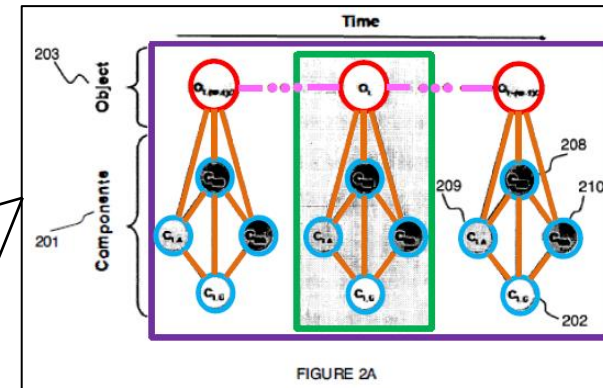
EX1001 ('980 Patent) Fig. 2A;
Paper 26 (POR) at 2-3

EX1001 ('980 patent) at 7:6-22; Paper 44 (Sur-Reply) at 8

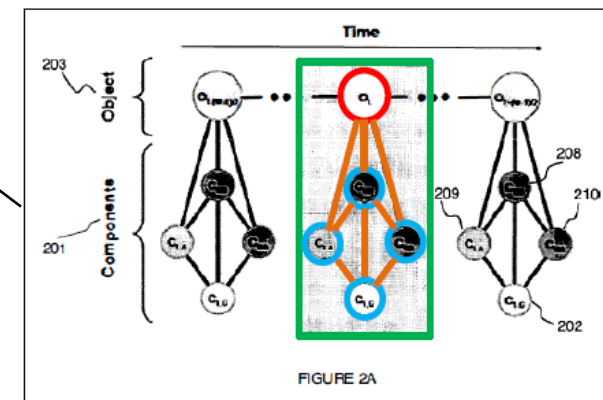
Claim Construction: 1/9[e] “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object”

Petitioners’ construction reads out all preferred embodiments (w>1) in favor of a disfavored embodiment (w=1)

Experiments; Tests were performed using a set of images collected with a single car-mounted grayscale camera. The result of vehicle detection and tracking over a sequence of 55 consecutive frames was evaluated. A 3-frame spatio-temporal object model was used and was shifted in a tracking centric way over time. BP was run with 30 particles for 10 iterations at every frame. For comparison a simple fusion scheme was implemented that averages the best detection result from each of the four components to produce an estimate for the vehicle position and scale independently at every frame. The performance of the simple fusion detection is poor suggesting that the noisy component detectors often do not have the global maximum at the correct position and scale. In contrast, the spatio-temporal object model consistently combines the evidence for accurate estimates throughout the sequence.



EX1001 ('980 Patent) Fig. 2A;
Paper 26 (POR) at 3-4



EX1001 ('980 Patent) Fig. 2A;
Paper 26 (POR) at 2-3

EX1001 ('980 Patent) 7:56-8:3; Paper 44 (Sur-Reply) at 8

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

Petitioners’ construction renders the claims indefinite

“If, after applying all other available tools of claim construction, a claim is ambiguous, it should be construed to preserve its validity. ... The canon favoring constructions that preserve claim validity therefore counsels against construing ‘communications path’ to include wireless communications.”

Ruckus Wireless, Inc. v. Innovative Wireless Sols., LLC, 824 F.3d 999, 1004 (Fed. Cir. 2016)

“This reading of the claim is indefinite. ... This court therefore rejects this proposed construction.”

Geneva Pharms., Inc. v. GlaxoSmithKline PLC, 349 F.3d 1373, 1384 (Fed. Cir. 2003)



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

Petitioners’ construction removes clarity and creates confusion

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

“Further, the **mismatch between** the various **plural** IP addresses provided earlier in the claim **and the singular** ‘said different IP Address’ suggests a term that is entirely untethered from the words around it. It is true that, as a general rule, a singular noun encompasses the plural for claim construction purposes. This claim construction principle, however, does not render the mismatch irrelevant. Where, as here, a singular/plural mismatch further confuses an already confused claim, **it is proper to consider the mismatch in discerning whether a POSA could understand the claim with reasonable certainty.**”

Bushnell Hawthorne, LLC v. Cisco Sys., Inc., 813 F. App’x 522, 526 (Fed. Cir. 2020) (cleaned up)

Antecedent Basis?

- P(**object**) in each image of the plurality?
- P(**collection of components**) in a single image?



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

Petitioners’ construction requires rewriting the claims

1. A computer implemented method for object detection comprising:

[1a] providing a spatio-temporal model for an object to be detected;

[1b] providing a video comprising a plurality of images including the object;

[1c] measuring the object as a collection of components in each image;

[1d] determining ~~a-probability~~ probabilities that the ~~object~~ collection of components is in each image; and

[1e] detecting the object in any image upon comparing the collection of component probabilities for each image to a threshold for detecting the ~~object~~ collection of components.

“In construing claims, **the analytical focus must begin and remain centered on the language of the claims themselves**, for it is that language that the patentee chose to use to particularly point out and distinctly claim the subject matter which the patentee regards as his invention.”

Interactive Gift Exp., Inc. v. Compuserve Inc., 256 F.3d 1323, 1331 (Fed. Cir. 2001)

“The actual words of the claim are the controlling focus.”

Digital Biometrics, Inc. v. Identix, Inc., 149 F.3d 1335, 1344 (Fed. Cir. 1998)

“Claims mean precisely what they say.”

Cent. Admixture Pharm. Servs., Inc. v. Advanced Cardiac Sols., P.C., 482 F.3d 1347, 1355 (Fed. Cir. 2007)



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

Petitioners’ construction renders claim 1 narrower than dependent claims

“By definition, an independent claim is broader than a claim that depends from it, so if a dependent claim reads on a particular embodiment of the claimed invention, the corresponding independent claim must cover that embodiment as well.”

Bd. of Regents of the Univ. of Texas Sys. v. BENQ Am. Corp., 533 F.3d 1362, 1371 (Fed. Cir. 2008).

“Under the doctrine of claim differentiation, dependent claims are presumed to be of narrower scope than the independent claims from which they depend.”

AK Steel Corp. v. Sollac & Ugine, 344 F.3d 1234, 1242 (Fed. Cir. 2003).



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

Petitioners’ construction renders claim 1 narrower than claim 3

1. A computer implemented method for object detection comprising:

[1a] providing a spatio-temporal model for an **object** to be detected;

[1b] providing a video comprising a plurality of images including the **object**;

[1c] measuring the **object** as a collection of **components** in each image;

[1d] determining a probability that the **object** is in each image; and

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

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**.

“The fact that the claim distinguishes between ‘each pre-programmed code’ and ‘the matched one or more pre-programmed codes’ is significant. Different claim terms are presumed to have different meanings.”

Bd. of Regents of the Univ. of Texas Sys. v. BENQ Am. Corp., 533 F.3d 1362, 1371 (Fed. Cir. 2008).



Paper 44 (Sur-Reply) at 9-10

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

Petitioners’ construction renders claim 1 narrower than claim 4

1. A computer implemented method for object detection comprising:

[1a] providing a spatio-temporal model for an object to be detected;

[1b] providing a video comprising a plurality of images including the object;

[1c] measuring the object as a collection of components in each image;

[1d] **determining a probability that the object is in each image**; and

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

4. The computer implemented method of claim 1, wherein **determining the probability that the object is in each image** comprises detecting the object in a **current image** according to measurements of the object as a collection of components determined from a **prior image** and a **later image** relative to the current image.

Q. Okay. So then at the end of the 15 paragraph, you refer to "I(j-0). “What is -- what does “j-0” mean in that circumstance?

A. That's the condition when “W” is equal to “1” using the way that the patent specifies it as well. So this windowing, it refers to (j-(w-1)/2) and (j+(w-1)/2). And **when “W” is equal to “1,”** that would become zero. So hence, **there’s no previous frame or no future frame.**

Q. Okay. So if there’s no previous frame and no future frame, then the analysis does not extend past the point in time of the current image, right?

A. **The analysis of detection in that image frame is not using any information from a previous frame or a future frame.** The processing is still being done frame by frame. “J” is moving from the beginning to the end of your video processing. The computer method is detecting, measuring, you know, comparing these probabilities.

EX2008 (Bajaj Dep. Tr.) at 51:14-52:21; Paper 44 (Sur-Reply) at 10, 11

Paper 44 (Sur-Reply) at 10



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

Petitioners’ construction covering $w=1$ violates canons of claim construction: plain and unambiguous claim language governs

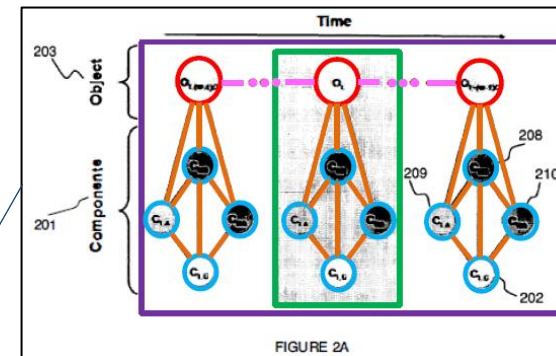
“First, absent contravening evidence from the specification or prosecution history, plain and unambiguous claim language controls the construction analysis.”

DSW, Inc. v. Shoe Pavilion, Inc., 537 F.3d 1342, 1347 (Fed. Cir. 2008).

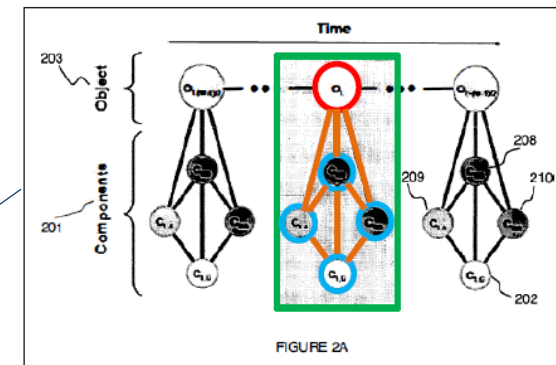


While it is possible to perform inference over the spatio-temporal model defined for the entire image sequence, there are many applications for which this is not an option due to the lengthy off-line processing needed. A w -frame windowed smoothing algorithm is used where w is an odd integer ≥ 1 . There 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.

EX1001 ('980 patent) at 7:6-22; Paper 44 (Sur-Reply) at 11



EX1001 ('980 Patent) Fig. 2A;
Paper 26 (POR) at 3-4



EX1001 ('980 Patent) Fig. 2A;
Paper 26 (POR) at 2-3

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

Petitioners’ construction covering $w=1$ violates canons of claim construction: plain and unambiguous claim language governs

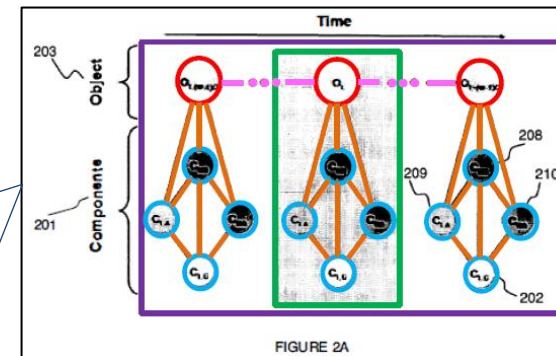
“First, absent contravening evidence from the specification or prosecution history, plain and unambiguous claim language controls the construction analysis.”

DSW, Inc. v. Shoe Pavilion, Inc., 537 F.3d 1342, 1347 (Fed. Cir. 2008).

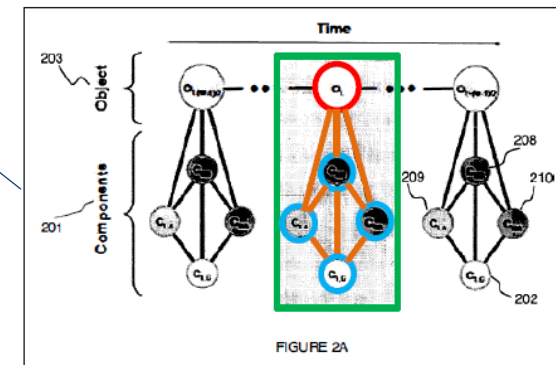


Experiments; Tests were performed using a set of images collected with a single car-mounted grayscale camera. The result of vehicle detection and tracking over a sequence of 55 consecutive frames was evaluated. A 3-frame spatio-temporal object model was used and was shifted in a tracking centric way over time. BP was run with 30 particles for 10 iterations at every frame. For comparison a simple fusion scheme was implemented that averages the best detection result from each of the four components to produce an estimate for the vehicle position and scale independently at every frame. The performance of the simple fusion detection is poor suggesting that the noisy component detectors often do not have the global maximum at the correct position and scale. In contrast, the spatio-temporal object model consistently combines the evidence for accurate estimates throughout the sequence.

EX1001 ('980 patent) at 7:56-8:3; Paper 44 (Sur-Reply) at 11



EX1001 ('980 Patent) Fig. 2A;
Paper 26 (POR) at 3-4



EX1001 ('980 Patent) Fig. 2A;
Paper 26 (POR) at 2-3

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

Petitioners’ construction covering w=1 violates canons of claim construction: claims need not cover every embodiment in the specification

“The mere fact that there is an alternative embodiment disclosed in the asserted patent that is not encompassed by our claim construction does not outweigh the language of the claim, especially when the court’s construction is supported by the intrinsic evidence.”

Aug. Tech. Corp. v. Camtek, Ltd., 655 F.3d 1278, 1285 (Fed. Cir. 2011) (cleaned up)

“[W]e should not infer that any particular embodiment is included in a claim when there is probative evidence that sufficiently indicates the contrary.”

SIMO Holdings Inc. v. Hong Kong uCloudlink Network Tech. Ltd., 983 F.3d 1367, 1378 (Fed. Cir. 2021)

“However, in a case such as this, where the patent describes multiple embodiments, every claim does not need to cover every embodiment. This is particularly true where the plain language of a limitation of the claim does not appear to cover that embodiment.”

Pacing Technologies, LLC v. Garmin Intern., Inc., 778 F.3d 1021, 1026 (Fed. Cir. 2015)



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

Petitioners’ construction covering w=1 violates canons of claim construction: disfavored embodiments may be excluded

“And even more notably, the specification **expressly criticizes** the interweaving practice where securing yarns include yarns of the upper and lower layers—i.e., the construction proposed by Barrday and endorsed by the dissent. ... The **‘repeated description’** of securing yarns as separate and distinct structures from yarns of the upper and lower layers, the **‘extolling of the virtues’** of separate securing yarns, and the **‘criticism’** of the interweaving practice where securing yarns include yarns of the upper and lower layers **‘clearly point to the conclusion’ that the construction** of securing yarns **excludes** yarns from the upper and lower layers.”

Barrday, Inc. v. Lincoln Fabrics Inc., 2023 WL 7871688, at *4-5 (Fed. Cir. Nov. 16, 2023)

“[A] patent applicant need not expressly state ‘my invention does not include X’ to indicate his exclusion of X from the scope of his patent because **the patentee’s choice of preferred embodiments can shed light on the intended scope of the claims.**”

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



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

The prosecution history does not support Petitioners’ construction

Allowable Subject Matter

3. Claims 1-16 are allowed.

The following is an examiner’s statement of reasons for allowance:

Independent claims 1 and 9 both define features of providing a spatio-temporal model for an object to be detected; 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. These features in combination with other features in claims 1 and 9, which are the broadest allowable claims, are not taught or suggested by the art of record.

The close prior art, U.S. Patent No. 6,597,801 B1, Cham et al, teaches providing a kinematic model for an object to be detected (column 11, line 26), providing a video comprising a plurality of images including the object (figure 4B); measuring the object as a collection of components in each image (figure 10A); performing feature search operations for object features by comparing a total probability with a predetermined threshold (column 3, lines 64-65). However, Cham does not teach the claimed “spatio-temporal model”, which is different from the kinematic model (spatial model).

EX1002 at 162-163

Paper 44 (Sur-Reply) at 13-14



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

The prosecution history does not support Petitioners’ construction

Elements 1/9[c]-1/9[e] are performed using the spatiotemporal model of 1/9[a].

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.

EX1052 (Bajaj Supp. Decl.) ¶14

Q. So why is providing a spatio-temporal model for an object to be detected a precursor to performing the Steps [c], 6 [d], and [e]?

A. Obviously. What am I going to detect? I'm going to detect an object, so you have to provide me what type of object are you trying to detect.

...

Q. Okay. So you need the spatio-temporal model to do the other measuring, determining, and detecting steps?

A. You need the video as well.

EX2008 (Bajaj Dep. Tr.) at 32:3-33:2



Claim Construction: 1/9[a] “spatio-temporal model”

Term	Patent Owner’s Construction	Petitioners’ Construction ²
“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

Paper 44 (Sur-Reply) at 15-19

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

Claim Construction: 1/9[a] “spatio-temporal model”

The meaning of “spatio-temporal model” is informed by the surrounding claim language

1. A computer implemented method for object detection comprising:

[1a] providing a ***spatio-temporal model*** for an object to be detected;

[1b] providing a video comprising a plurality of images including the object;

[1c] measuring the object as a collection of components in each image;

[1d] determining a probability that the object is in each image; and

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

“In any event, and decisively, our inquiry is not limited to an analysis of the phrase in isolation.”

Intel Corp. v. Qualcomm Inc., 21 F.4th 784, 791-92 (Fed. Cir. 2021)

“While certain terms may be at the center of the claim construction debate, the context of the surrounding words of the claim also must be considered in determining the ordinary and customary meaning of those terms.”

ACTV, Inc. v. Walt Disney Co., 346 F.3d 1082, 1088 (Fed. Cir. 2003)



Claim Construction: 1/9[a] “spatio-temporal model”

The express function of “spatio-temporal model” is “for an object to be detected”

1. A computer implemented method for object detection comprising:

[1a] providing a *spatio-temporal model* for an object to be detected;

[1b] providing a video comprising a plurality of images including the object;

[1c] measuring the object as a collection of components in each image;

[1d] determining a probability that the object is in each image; and

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

“It is ... entirely proper to consider the functions of an invention in seeking to determine the meaning of particular claim language.”

Medrad, Inc. v. MRI Devices Corp., 401 F.3d 1313, 1319 (Fed. Cir. 2005)

“Here, we conclude that the ‘for decoding’ language in the preamble of claims 1 and 23 is properly construed as a claim limitation, and not merely a statement of purpose or intended use for the invention, because ‘decoding’ is the essence or a fundamental characteristic of the claimed invention.”

Vizio, Inc. v. Int’l Trade Comm’n, 605 F.3d 1330, 1340-41 (Fed. Cir. 2010)



Paper 44 (Sur-Reply) at 16

Claim Construction: 1/9[a] “spatio-temporal model”

Petitioners’ expert agrees that elements 1/9[c]-1/9[e] are performed using the spatiotemporal model of 1/9[a].

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.

EX1052 (Bajaj Supp. Decl.) ¶14

Q. So why is providing a spatio-temporal model for an object to be detected a precursor to performing the Steps [c], [d], and [e]?

A. Obviously. What am I going to detect? I'm going to detect an object, so you have to provide me what type of object are you trying to detect.

...

Q. Okay. So you need the spatio-temporal model to do the other measuring, determining, and detecting steps?

A. You need the video as well.

EX2008 (Bajaj Dep. Tr.) at 32:3-33:2



Claim Construction: 1/9[a] “spatio-temporal model”

The specification repeatedly indicates the function of the spatio-temporal model is object detection based on probabilities in other images

“The person of ordinary skill in the art is deemed to read the claim term not only in the context of the particular claim in which the disputed term appears, but in the context of the entire patent, including the specification. The specification may reveal a special definition given to a claim term by the patentee that differs from the meaning it would otherwise possess. In such cases, the inventor's lexicography governs. The specification need not reveal such a definition explicitly. When a patentee uses a claim term throughout the entire patent specification, in a manner consistent with only a single meaning, he has defined that term by implication.”

AstraZeneca LP v. Apotex, Inc., 633 F.3d 1042, 1051-52 (Fed. Cir. 2010) (cleaned up)



Claim Construction: 1/9[a] “spatio-temporal model”

The specification repeatedly indicates the function of the spatio-temporal model is object detection based on probabilities in other images

According to an embodiment of the present disclosure, a computer implemented method for object detection comprises providing a spatio-temporal model for an object to be detected, providing a video comprising a plurality of images including the object, and measuring the object as a collection of components in each image. The method further comprises

EX1001 ('980 patent) at 1:28-33

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

EX1001 ('980 patent) at 1:37-39

Providing the spatio-temporal model for the object to be detected further comprises providing a temporal window defining a plurality of images in which measurements of components detected therein are passed to components detected in the current image.

EX1001 ('980 patent) at 1:48-52

According to an embodiment of the present disclosure, a program storage device is provided readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for object detection. The method steps comprising providing a spatio-temporal model for an object to be detected, providing a video comprising a plurality of images including the object, and measuring the object as a collection of components in each image.

EX1001 ('980 patent) at 2:7-15

the object. Directed edges between nodes represent learned spatial and temporal probabilistic constraints. Each node in

EX1001 ('980 patent) at 2:59-60

Graphical Object Models; An object, such as an automobile (see FIG. 2B), is modeled as a spatio-temporal directed graphical model. Each node in the graph represents either the

EX1001 ('980 patent) at 3:47-48

Paper 44 (Sur-Reply) at 17

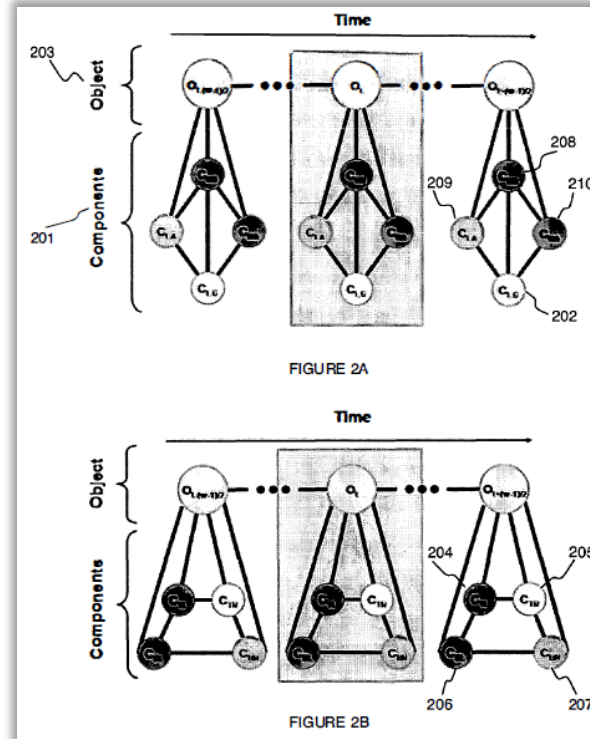
Claim Construction: 1/9[a] “spatio-temporal model”

The specification repeatedly indicates the function of the spatio-temporal model is object detection based on probabilities in other images

To integrate temporal constraints the spatial graphical models are extended over time to an arbitrary length temporal window. The resulting spatio-temporal graphical models are shown in FIGS. 2A and 2B. Having a two-layer graphical model allows the inference process to reason explicitly about the object as a whole, e.g., 203, and reduce the complexity of the graphical model by allowing the assumption of the conditional independence of components, e.g., 202 and 208-210, over time given the overall object appearance. Alternatively, a single object layer model can be built with bi-directional temporal constraints.

As can be seen in FIGS. 2A and 2B, an object at a current time is denoted as o_t with an earlier model denoted as $o_{t-(w-1)/2}$ and a later model denoted as $o_{t+(w-1)/2}$.

EX1001 ('980 patent) at 1:48-52



EX1001 ('980 patent) Figs. 2A-2B

The output of the AdaBoost classifier is a confidence $h_k(I)$ that the given pattern I is of the desired class. An object is considered present if

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

EX1001 ('980 patent) at 5:58-67

Referring to FIG. 5, a computer implemented method for object detection includes providing a spatio-temporal model, e.g., see FIGS. 2A and 2B, for an object 501, providing a video including a plurality of images including the object 502, and measuring the object as a collection of components in each image of the video 503. The method further includes determining a probability that the object is in each image 504 by using message passing between components represented as nodes of the spatio-temporal model, and detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object 505.

EX1001 ('980 patent) at 7:45-55

Paper 44 (Sur-Reply) at 17

Claim Construction: 1/9[a] “spatio-temporal model”

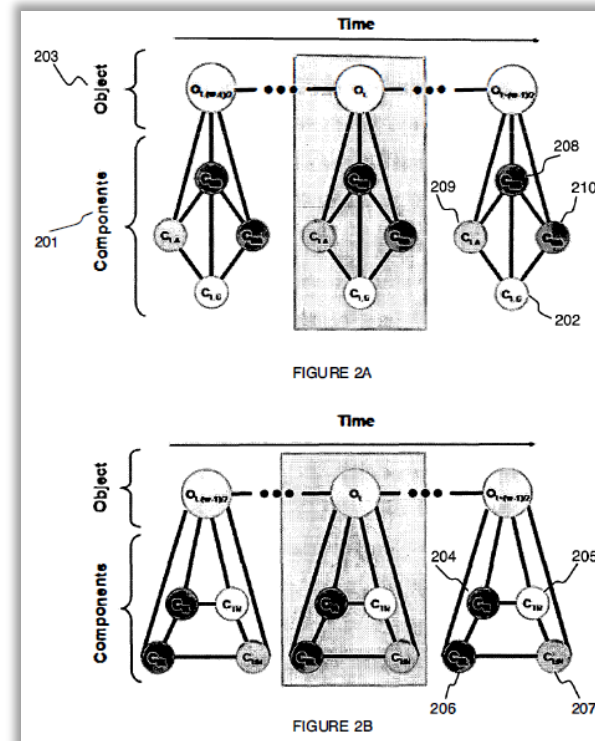
The specification repeatedly indicates the function of the spatio-temporal model is object detection based on probabilities in other images

Experiments; Tests were performed using a set of images collected with a single car-mounted grayscale camera. The result of vehicle detection and tracking over a sequence of 55 consecutive frames was evaluated. A 3-frame spatio-temporal object model was used and was shifted in a tracking centric way over time. BP was run with 30 particles for 10 iterations at every frame. For comparison a simple fusion scheme was implemented that averages the best detection result from each of the four components to produce an estimate for the vehicle position and scale independently at every frame. The performance of the simple fusion detection is poor suggesting that the noisy component detectors often do not have the global

maximum at the correct position and scale. In contrast, the spatio-temporal object model consistently combines the evidence for accurate estimates throughout the sequence.

A 3-frame spatio-temporal object model was run at a single instance in time for two pedestrians in two different scenes. Similar to the vehicle detection BP was run with 30 particles for 10 iterations. For both experiments the temperature of the likelihood is set to $T_0=0.2$.

EX1001 ('980 patent) at 7:56-8:8



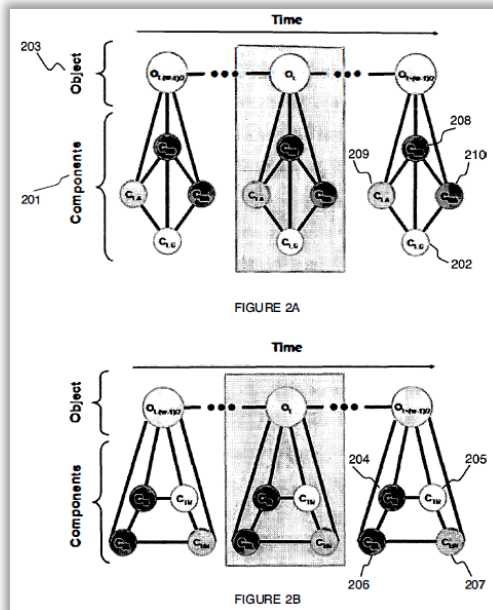
EX1001 ('980 patent) Figs. 2A-2B

Claim Construction: 1/9[a] “spatio-temporal model”

The specification teaches spatial models extended over time

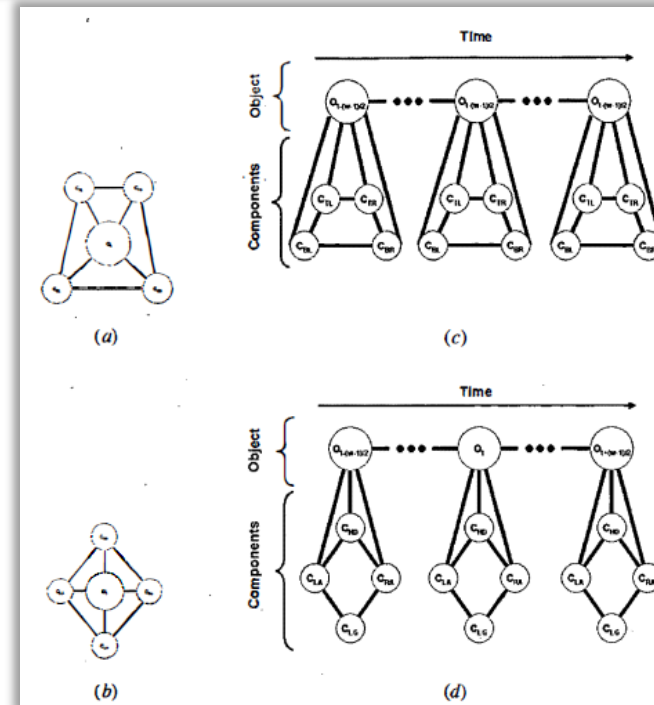
To integrate temporal constraints the spatial graphical models are extended over time to an arbitrary length temporal window. The resulting spatio-temporal graphical models are shown in FIGS. 2A and 2B. Having a two-layer graphical

EX1001 ('980 patent) at 4:34-45



EX1001 ('980 patent) Figs. 2A-2B

Figure 2: Spatial graphical models for the vehicle (a) and pedestrian (b), plus their spatio-temporal counterparts (c) and (d) respectively. Spatio-temporal models are obtained by replicating a spatial model along the temporal domain to a w -length window and then connecting the second-layer nodes across time.



EX1003 (provisional) at 7

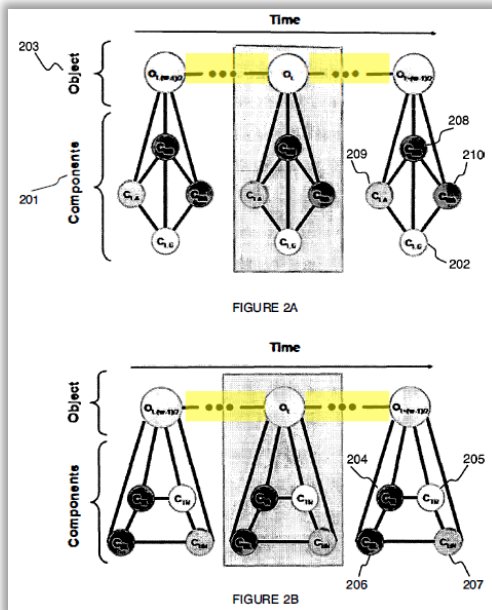
Paper 44 (Sur-Reply) at 17

Claim Construction: 1/9[a] “spatio-temporal model”

The specification teaches temporal probabilistic connectivity in the spatio-temporal model

the object. Directed edges between nodes represent learned spatial and temporal probabilistic constraints. Each node in

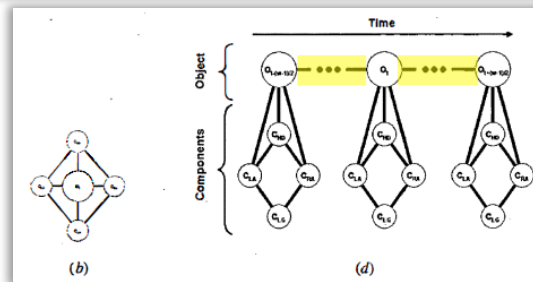
EX1001 ('980 patent) at 2:59-60



EX1001 ('980 patent) Figs. 2A-2B

ical 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

The framework presented here can be viewed as having five distinct components: (i) a graphical model, (ii) an inference algorithm that provides ability to infer a state of each node in a graph, (iii) a local evidence distribution (or image likelihood), (iv) a proposal process for some or all nodes in a graphical model, and a set of (v) spatial and/or temporal constraints corresponding to the edges in a graph. We will now discuss each one of these in term.



EX1003 (provisional) at 5-7

9). The joint probability distribution for this spatio-temporal graphical object model with N components can be written as:

$$P(X_0^O, X_0^{C_1}, X_0^{C_2}, \dots, X_0^{C_N}, \dots, X_T^O, X_T^{C_1}, X_T^{C_2}, \dots, X_T^{C_N}) = \frac{1}{Z} \prod_{ij} \psi_{ij}(X_i^O, X_j^O) \prod_{ik} \psi_{ik}(X_i^O, X_i^{C_k}) \prod_{ikl} \psi_{ikl}(X_i^{C_k}, X_i^{C_l}) \prod_i \phi_i(X_i^O) \prod_{ik} \phi_i(X_i^{C_k})$$

where X_t^O and $X_t^{C_n}$ is the state of the object, O, and object's n-th component, C_n , at time t respectively ($n \in (1, N)$ and $t \in (1, T)$; $\psi_{ik}(X_i^O, X_j^O)$ is the temporal compatibility of object state between frames i and j; $\psi_{ik}(X_i^O, X_i^{C_k})$ is the spatial compatibility of the object and its components at frame i; $\psi_{ikl}(X_i^{C_k}, X_i^{C_l})$ is the spatial compatibility between object components at frame i and $\phi_i(X_i^O)$ and $\phi_i(X_i^{C_k})$ denote the local evidence for the object and component states respectively.

EX1001 ('980 patent) at 3:47-4:8

Paper 44 (Sur-Reply) at 17

Claim Construction: 1/9[a] “spatio-temporal model”

The spatio-temporal model is used to achieve the function and purpose of the invention

To integrate temporal constraints the spatial graphical models are extended over time to an arbitrary length temporal window. The resulting spatio-temporal graphical models are shown in FIGS. 2A and 2B. Having a two-layer graphical model allows the inference process to reason explicitly about the object as a whole, e.g., 203, and reduce the complexity of the graphical model by allowing the assumption of the conditional independence of components, e.g., 202 and 208-210, over time given the overall object appearance. Alternatively, a single object layer model can be built with bi-directional temporal constraints.

EX1001 ('980 patent) at 1:48-52;
Paper 44 (Sur-Reply) at 18-19

According to an embodiment of the present disclosure, a probabilistic framework for automatic component-based detection and tracking of objects in images and/or video combines object detection with tracking in a unified framework. Tracking makes use of object detection for initialization and re-initialization during transient failures for occlusions. Object detection considers the consistency of the detection over time. Modeling objects by an arrangement of image-base, and possibly overlapping, components facilitates detection of complex articulated objects as well as helps in handling partial object occlusions or local illumination changes.

EX1001 ('980 patent) at 2:43-54;
Paper 44 (Sur-Reply) at 18-19

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.

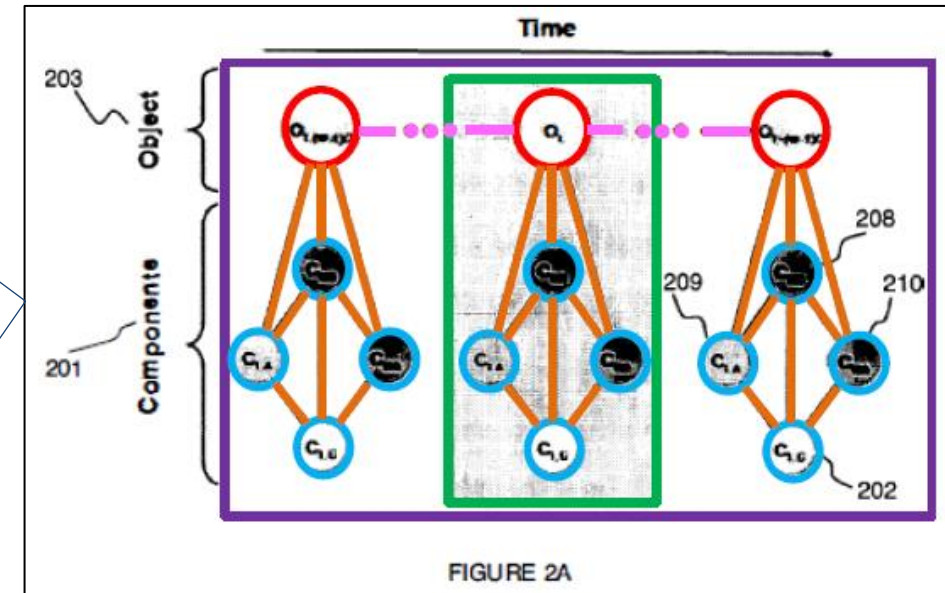
EX1001 ('980 patent) at 7:6-22;
Paper 44 (Sur-Reply) at 18-19

Claim Construction: 1/9[a] “spatio-temporal model”

The spatio-temporal model is used to achieve the function and purpose of the invention

Experiments; Tests were performed using a set of images collected with a single car-mounted grayscale camera. The result of vehicle detection and tracking over a sequence of 55 consecutive frames was evaluated. A 3-frame spatio-temporal object model was used and was shifted in a tracking centric way over time. BP was run with 30 particles for 10 iterations at every frame. For comparison a simple fusion scheme was implemented that averages the best detection result from each of the four components to produce an estimate for the vehicle position and scale independently at every frame. The performance of the simple fusion detection is poor suggesting that the noisy component detectors often do not have the global maximum at the correct position and scale. In contrast, the spatio-temporal object model consistently combines the evidence for accurate estimates throughout the sequence.

EX1001 ('980 patent) at 7:56-8:3; Paper 44 (Sur-Reply) at 19



EX1001 ('980 Patent) Fig. 2A; Paper 26 (POR) at 3-4

Claim Construction: 1/9[a] “spatio-temporal model”

Petitioners’ construction is divorced from the intrinsic record and covers purely spatial models

“[A] common meaning, such as one expressed in a relevant dictionary, that flies in the face of the patent disclosure is undeserving of fealty. ... Ultimately, the interpretation to be given a term can only be determined and confirmed with a full understanding of what the inventors actually invented and intended to envelop with the claim. ... A claim construction is persuasive, not because it follows a certain rule, but because it defines terms in the context of the whole patent.”

Renishaw PLC v. Marposs Societa’ per Azioni, 158 F.3d 1243, 1250 (Fed. Cir. 1998).



Claim Construction: 1/9[a] “spatio-temporal model”

Petitioners’ construction is divorced from the intrinsic record and covers purely spatial models

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.

EX1052 (Bajaj Supp. Decl.) ¶14

Q. ... And I'd like to know what you mean by "change configuration in time."

A. ... And with this ability that the component collection model will change configuration; that's exactly what a spatio-temporal model that's provided, saying I'm giving you a model -- giving "you" being the computer program -- that's going to change in time. It's a moving model.

But all I'm asking you for you to do is look at each image, and capture and measure each of the components.

So the model changing configuration in time is implicit that you know that the model can move in time. Just like when you're filming it. The model is moving and you're giving me a video which is capturing a moving model.

That's what I mean by the "component collection model to change configuration in time."

Q. Okay. So then the ability for the model to change configuration in time, is that what provides the temporal element of it being a spatio-temporal model?

A. That's correct. Obviously.

'980 Patent

PRIORITY DATE



Priority Date

Petitioners' argument is waived

A. Prior Art

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

Paper 3 (Petition) at 5

In the Decision to Institute we acknowledged that a Patent Owner “is not presumed to be entitled to the earlier filing dates of ancestral applications which do not share the same disclosure.” Dec. on Inst. 8.

However, we required, as other panels have required, that:

Petitioner, however, must first raise the issue of whether Patent Owner is entitled to an effective filing date by “identifying, specifically, the features, claims, and ancestral applications allegedly lacking written description support for the claims based on the identified features.” *Lupin*, IPR2015–01775, Paper 15 at 11 (citing *Focal Therapeutics*, IPR2014–00116, Paper 8 at 10). Once the initial production burden is satisfied by a petitioner, the patent owner must “make a sufficient showing of entitlement to earlier filing date(s), in a manner that is commensurate in scope with the specific points and contentions raised by the petitioner.” *Id.* (emphasis added).

Id. No Federal Circuit case is cited in our Decision to support this proposition.

MaxLite, Inc. v. Jiaxing Super Lighting Elec. Appliance Co., Ltd., IPR2020-00208, Paper 14 at 6 (PTAB Jun. 1, 2021)

proceeding to respond. Petitioner waived this argument because it was not presented in the Petition. See Consolidated Trial Practice Guide, 73 (Nov. 2019) (“Petitioner may not submit new evidence or argument in reply that it could have presented earlier, e.g. to make out a prima facie case of unpatentability.”).

Xilinx, Inc. v. Analog Devices, Inc., IPR2020-01559, Paper 31 at 42 (PTAB Mar. 11, 2022)

Priority Date

Petitioners' written description argument is limited to 1/9[d] and 1/9[e]

Claim 1

1. A computer implemented method for object detection comprising:

[1a] providing a spatio-temporal model for an object to be detected;

[1b] providing a video comprising a plurality of images including the object;

[1c] measuring the object as a collection of components in each image;

[1d] determining a probability that the object is in each image; and

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

PO argues as of the filing date of the provisional application (Ex1003) on May 27, 2004, the earliest priority date of the '980 Patent, "the inventors had possession of the claimed subject matter." POR, 7. Petitioners reserved the right to challenge the May 27, 2004 priority date or any evidence of prior invention. Pet., 5. Under PO's interpretation and claim construction of 1/9[d]-[e], the provisional does not provide written description under pre-AIA §112. *Parus Holdings, Inc. v. Google LLC*, 70 F.4th 1365, 1374 (Fed. Cir. 2023) (patentee had failed to prove priority claim because its published patent application did not have adequate written description support for challenged claims). Likewise, PO cannot demonstrate that the provisional discloses 1/9[e] or 1/9[d]. POR, 8-9; Ex2004, ¶¶35-36; Ex1051, 39:17-66:14.

Paper 32 (Reply) at 5-6

Paper 44 (Sur-Reply) at 20-21



Priority Date

Petitioners' sole written description argument is contrary to law

39:17-40:2; Ex1003, 9-10. The only discussion about “threshold” in the provisional is with regards to AdaBoost, but without further disclosure of “detecting the object in an image upon comparing the probabilities for the image to a threshold for detecting the object.” Ex1003, 9-10; Ex1001, Figure 5. Similarly, the additional disclosures in the provisional that PO and Dr. Saber identify also do not have sufficient written description support for 1/9[e] or 1/9[d]. POR, 8-9; Ex1003, 8-10, Ex2004, ¶¶35-36. Given the provisional lacks sufficient written

Paper 32 (Reply) at 5-6

“Even so, when examining the written description for support for the claimed invention, we have held that the exact terms appearing in the claim “need not be used *in haec verba*.” *Lockwood v. Am. Airlines, Inc.*, 107 F.3d 1565, 1572 (Fed.Cir.1997). We are therefore troubled by the fact that the Board did not cite any evidence other than the fact that the terms were not present in the specification to support its finding.”

Blue Calypso, LLC v. Groupon, Inc., 815 F.3d 1331, 1346 (Fed. Cir. 2016).



Paper 44 (Sur-Reply) at 21

Priority Date

Patent Owner identified ample, un rebutted evidence

Claim 1

1. A computer implemented method for object detection comprising:

[1a] providing a spatio-temporal model for an object to be detected;

[1b] providing a video comprising a plurality of images including the object;

[1c] measuring the object as a collection of components in each image;

[1d] determining a probability that the object is in each image; and

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

32. Petitioners' argument that the provisional does not include the '980 patent's Figure 5 flowchart or the exact text of element 1/9[e] ignores the evidence that I cited. Reply, 6. I am informed that written description support does not require the exact claim language to appear in the provisional. Based on that understanding, I identified portions of the provisional application that indicate to the POSITA that the inventors possessed the claimed invention. For example, as I explained, the provisional describes an "inference algorithm" that uses a "probabilistic image likelihood model $\phi_i(X_i)$," which is "the likelihood of observing the [] image conditioned on the state X_i of node i ." EX1003, 9. It further states that "[t]he output of the AdaBoost classifier is a confidence $h_k(I)$ " that is converted to a likelihood function, *i.e.*, a probability, where "[i]t is customary to consider object present if $h_k(I) \geq \frac{1}{2} \sum_{k=1}^K \alpha_k$." EX1003, 10. Based on these disclosures (and the others identified in my initial declaration), the POSITA would understand that that the inventors were in possession of element 1/9[d].

Priority Date

Patent Owner identified ample, un rebutted evidence

Claim 1

1. A computer implemented method for object detection comprising:

[1a] providing a spatio-temporal model for an object to be detected;

[1b] providing a video comprising a plurality of images including the object;

[1c] measuring the object as a collection of components in each image;

[1d] determining a probability that the object is in each image; and

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

33. As also explained in my initial declaration, the provisional application further states that “[o]bject detection and tracking are formulated as inference in a two-layer graphical model,” where “[i]nferring the state of the object and its components in our framework is defined as estimating a belief in a graphical model” using “a form of non-parametric belief propagation.” EX1003, 4, 8. “The approach is a generalization of particle filtering” where “the conditional distribution used in standard particle filtering is replaced by product of incoming messages,” where “we integrate information from all components in the second-layer node,” (*i.e.*, in the object) and send messages between “the set of neighbors of node *i*” (as well as between frames) based on “the local evidence (or likelihood) associated with the node *i*” ($\phi_i(X_i)$) and “the potential designating the compatibility between the states of node *i* and *j*” (which can be in different frames). EX1003, 8-9. In other words, the object is detected based on object probabilities from multiple frames. Furthermore, a “w-frame windowed smoothing algorithm” can be used for object detection, “where *w* is an odd integer ≥ 1 (see Figs. 2c and 2d),” to achieve “temporal integration” and “temporal consistency” between frames.” *Id.* Based on these disclosures (and the others identified in my initial declaration), the POSITA would understand that the inventors were in possession of element 1/9[e].

'980 Patent

TLP IS NOT PRIOR ART



Priority Date

Petitioners' sole public accessibility theory is pre-conference dissemination to peer reviewers

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

Paper 3 at 6

66. TLP was presented at the 2004 CVPR for Poster Session 2. See EX1010 at 4. In fact, TLP was listed next to my own paper in the 2004 CVPR program. *Id.* at 4 (listing TLP next to “A Segmentation-Free Approach for Skeletonization of Gray-Scale Images via Anisotropic Vector Diffusion” by Z. Yu and C. Bajaj). When submitting a paper for the 2004 CVPR, paper submissions are done well in advance of the actual conference. For the 2004 CVPR, the deadline for full paper submission was November 19, 2003. EX1010 at 1. After submission, the papers are widely disseminated and made publicly available to academics for peer review. After that process, authors are notified as to whether the papers have been accepted. For the 2004 CVPR, notification to the authors was March 2, 2004. *Id.* For papers that are accepted, authors submit manuscripts for the conference. The deadline for manuscript submission for the 2004 CVPR was April 9, 2004. *Id.* Since TLP was a paper presented at the 2004 CVPR, it is my opinion that it underwent the submission process outlined above, just as my own paper was. Accordingly, it is my opinion that TLP was widely disseminated and made publicly available through the submission process, and no later than April 9, 2004.

Paper 26 at 54-57; Paper 44 (Sur-Reply) at 22-23, 25-27



Priority Date

Petitioners' public accessibility theory fails as a matter of law

"Live Traffic" paper in SRI ³		TLP
"was still subject to pre-publication review"	✓	Alleged dissemination was "for peer review" before 2004 CVPR. EX1004 ¶¶66.
was posted "solely to facilitate peer review in preparation for later publication"	✓	Alleged dissemination was "for peer review" before 2004 CVPR. EX1004 ¶¶66.
"was not publicized or placed in front of the interested public"	✓	No evidence TLP was publicized or placed in front of public before May 27, 2004. <i>See</i> EX1004 ¶¶66; EX2007 ¶¶61-64.
"this record does not evince that the Live Traffic paper was accessible to anyone other than the peer-review committee, thus further suggesting an absence of actual public accessibility"	✓	No evidence TLP was accessible outside the CVPR peer review committee. <i>See</i> EX1004 ¶¶66; EX2007 ¶¶61-64.
"the peer-review feature also suggests no intent to publicize"	✓	No evidence of intent to publicize during peer review. <i>See</i> EX1004 ¶¶66; EX2007 ¶¶61-64.

³*SRI Int'l, Inc. v. Internet Sec. Sys.*, 511 F.3d 1186, 1197-98 (Fed. Cir. 2008)

Priority Date

Petitioners' third-party discovery confirms their public accessibility theory is incorrect

----- Forwarded message from reynolds@cc.gatech.edu -----

Date: Fri, 13 Feb 2004 19:37:42 GMT
From: reynolds@cc.gatech.edu
Reply-To: reynolds@cc.gatech.edu
Subject: Reviews for Paper #1109
To: ls@cs.brown.edu, cvprcl@cc.gatech.edu

Dear Leonid,

In preparation for the author response period, please find below the reviews we have received for the following paper:

Paper ID: 1109
Title: Tracking Loose-limbed People

These reviews were written during the initial reviewing phase and then discussed by the Program Committee Members and Area Chair(s) in charge of your paper.

Please note that the set of attached reviews is not final. While the content of the reviews below will not change, your paper will receive a summary written by an Area Chair and may receive additional reviews before final decisions are made. Papers for which only two reviews are currently available will receive at least one more review.

EX2009

Instructions for the Author Response Period

1. The author response period will begin on Feb 12, 2004 and will end on Feb 17th at 10pm EST.
2. Responses are limited to 4000 characters of plain text.
3. Author responses must be entered directly on the website. Go to the 'My Papers' page after logging in and click on the appropriate link for each of your papers.
4. Your response will be read only by the Area Chair(s) in charge of your paper and by the Program Chairs. Your response will NOT be communicated verbatim to the reviewers of your paper, but may result in further discussion between the Area Chair and those reviewers.

IMPORTANT:

- AUTHOR RESPONSES ARE ENTIRELY VOLUNTARY. FAILURE TO SUBMIT A RESPONSE WILL NOT BE CONSIDERED NEGATIVELY DURING FINAL DECISION MAKING
- **DO NOT INCLUDE ANY INFORMATION IN YOUR RESPONSE THAT MAY REVEAL AUTHOR IDENTITIES!!!** The entire decision process will be conducted double-blind. IF THE RESPONSE CONTAINS ANY IDENTIFYING INFORMATION (WEBSITES, EMAILS, ETC) IT WILL BE DELETED IN ITS ENTIRETY.

Paper 44 (Sur-Reply) at 26-27



Priority Date

Dr. Saber opines that TLP discovery is consistent with non-disclosure

63. These emails are consistent with my experience as an author of many conference papers as well as a member of several conference committee where I have served in roles such as Technical Program Chair, Area Chair, Special Sessions Chair among others where technical papers are submitted and peer reviewed prior to acceptance. In my experience, papers submitted to technical conferences are managed by the Technical Program Chair(s) charged with facilitating peer review under strict protocols of confidentiality and in some instances “blind” review (*i.e.*, the reviewers do not know whose paper they are reviewing). It is customary and standard practice that papers submitted to a conference for review and potential acceptance are kept strictly confidential during that process, including to protect papers that are not accepted from public disclosure or public association with the authors. Indeed, disclosing a conference submission to the public or to anyone outside of the peer review committee personnel assigned to that submission would be considered unethical and a serious breach of trust that could result in disciplinary

actions, not to mention reputational harm to the violator among professional and academic peers. The blind review process, which ensures that the authors of a submission remain unknown, would also prevent peer reviewers from knowing (or divulging to others) a relationship between an author and a submission. In my experience, which I believe to be representative of academic and professional norms, papers submitted for acceptance to a conference are not made available to the public, not included in a public catalog, public database, or public search tool, and are not publicized or disseminated to the public before they are officially published by the conference (if they are accepted, otherwise, not at all).

EX2010

Paper 44 (Sur-Reply) at 26-27



Priority Date

Disclosure to PTO during prosecution does not make TLP prior art

“In *Abbott Laboratories v. Baxter Pharmaceutical Products, Inc.*, 334 F.3d 1274, 1279 (Fed.Cir.2003), the court explained that ‘mere submission of an IDS to the USPTO does not constitute the patent applicant’s admission that any reference in the IDS is material prior art.’ We agree that ResQNet did not convert these manuals into printed publication prior art by including them with the IDS submitted to the PTO.”

ResQNet.com, Inc. v. Lansa, Inc., 594 F.3d 860, 866 (Fed. Cir. 2010)



Priority Date

Petitioners new theory that TLP was publicly accessible on “ResearchGate”

106:22-107:14. Petitioners’ evidence of record shows that not only was TLP publicly available at least as of April 9, 2004 (Ex1004, ¶¶65-66; Ex1010), TLP was available even earlier as of January 31, 2004 (Ex1050, ¶¶40-48; Paper-28).

TLP is accordingly prior art to the ’980 Patent.

Paper 32 (Reply) at 7

41. The Sigal conference paper is also available from the *IEEE Computer Society Digital Library*,¹⁵ and the digital repositories *ResearchGate*¹⁶ and *Semantic Scholar*.¹⁷ Author Michael J. Black contributed the Sigal article to *ResearchGate* and the entry indicates a January 2004 date.

45. Michael J. Black posted the Sigal conference paper (Exhibit 1009) and indicated a January 2004 date (see Attachment 1D). The date on the Sigal conference

EX1050 (Hall-Ellis Decl.)

17 Q. So, are you saying that TLP was
18 published on ResearchGate in January 2004?

19 A. Yes, I am.

EX2011 (Hall-Ellis Dep. Tr. at 12:17-19)

Paper 44 (Sur-Reply) at 24-35



Priority Date

Petitioners “ResearchGate” argument is a new theory

“Before the Board, Samsung did not argue that the WD4 reference was publicly accessible because it was discussed at the Torino meeting or disseminated at the Geneva meeting. ... Thus, Samsung waived any separate argument on appeal that the WD4 reference was publicly accessible because it was disseminated at the Torino or Geneva meetings.”

Samsung Elecs. Co. v. Infobridge Pte. Ltd., 929 F.3d 1363, 1369-70 (Fed. Cir. 2019)



“We agree with Patent Owner that Petitioner’s argument in Section II.C. of its opening brief is a new argument or theory of public disclosure that was not raised in the Petition and, therefore, is waived. See *Ariosa*, 805 F.3d at 1367. As the Federal Circuit found, Samsung argued in the Petition that WD4 was publicly accessible because it was uploaded to the JCT-VC and MPEG websites and emailed to the JCT-VC listserv. See *Samsung*, 929 F.3d at 1369–1373. ... We determine that this is a new argument or theory because the argued dissemination is not based on the listserv email, but on post-listserv email ‘secondary distribution’ that may have occurred with respect not only to WD4, but also ‘other versions of Working Draft 4.’ Accordingly, Petitioner’s new ‘public dissemination’ argument is waived.”

Samsung Elecs. Co. v. Infobridge Pte., Ltd., IPR2017-00099, Paper 43 at 42 (PTAB Nov. 13, 2020)



Priority Date

Petitioners new “ResearchGate” theory should be ignored

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.

of this case warrant granting the Motion. We caution Petitioner that granting the Motion does not mean we will entertain impermissible new theories in the Reply.

Paper 28 at 4

Paper 44 (Sur-Reply) at 24-25



Priority Date

Petitioners new “ResearchGate” theory fails

Dr. Hall-Ellis’ concession that TLP was not on ResearchGate in 2004 is dispositive

30. Founded in 2008 by virologist and computer scientist Dr. Ijad Madisch with physician Dr. Sören Hofmayer, and computer scientist Horst Fickenscher, ResearchGate restricts its user accounts to people at recognized institutions and published researchers. Most of ResearchGate’s users are involved

EX1050 (Hall-Ellis Decl.)

17 Q. So, are you saying that TLP was
18 published on ResearchGate in January 2004?

19 A. Yes, I am.

EX2011 (Hall-Ellis Dep. Tr. at 12:17-19)

1 Q. TLP was not available on
2 ResearchGate prior to 2008, correct?

3 A. That is true.

EX2011 (Hall-Ellis Dep. Tr. at 16:1-3)

Paper 44 (Sur-Reply) at 24-25



Priority Date

Petitioners new “ResearchGate” theory fails

Dr. Hall-Ellis admits TLP was not catalogued or indexed before August 30, 2004

15 Q. Okay. So, in your view, when was
16 the 2004 IEEE Proceedings, when was it catalogued
17 and indexed?

18 A. It was catalogued on August 30,
19 2004. Which means that all of the bibliographic
20 data were transcribed which also means it was, as
21 lawyers call it, indexed.

EX2011 (Hall-Ellis Dep. Tr. at 45:15-21)

21 Q. Okay. And so there was not a, the
22 2004 IEEE proceeding were not catalogued before
1 August 30, 2004?

2 A. No, I do not find evidence of that.

EX2011 (Hall-Ellis Dep. Tr. at 46:21-47:2)

Paper 44 (Sur-Reply) at 24-25



Priority Date

Petitioners new “ResearchGate” theory fails

No evidence that “Jan 2004” is a publication date

5 Q. And it says January 2004, but it
6 doesn't say **published** on January 2004. Correct?
7 A. **It does not have that word, that is**
8 **correct.**
9 Q. Okay. And so, that is your -- that
10 it was published on January 2004, that is an
11 **inference** you are making based on the date and
12 the year, correct?
13 A. It is not the only piece of
14 information. But, definitely is a piece of
15 information. **Yes.**

EX2011 (Hall-Ellis Dep. Tr. at 22:5-15)

1 Q. And so many authors, they might say
2 something was published on a date.
3 A. Uh-huh.
4 Q. If that is the date they intended to
5 say, that -- that is **not necessarily meaning it**
6 **was a printed publication** for purposes of patent
7 laws, correct?
8 A. **It might not be.** And, we also have
9 to know if the documents themselves were
10 distributed at the conference. Sometimes they
11 are, sometimes they are not.
12 **So, we can't just make these**
13 **generalizations about anything** except to say we
14 need to look at all of the pieces of information.

EX2011 (Hall-Ellis Dep. Tr. at 26:1-14)

1 Q. Okay. And do you have any reason to
2 believe that **the TLP authors** that put the TLP
3 reference on ResearchGate **were familiar with the**
4 **patent requirements for something being a printed**
5 **publication** when they uploaded the document to
6 ResearchGate?
7 A. **That is outside the scope of**
8 **anything I would know.** I don't know these
9 people.

EX2011 (Hall-Ellis Dep. Tr. at 27:1-9)

Paper 44 (Sur-Reply) at 24-25



Priority Date

Petitioners new “ResearchGate” theory fails

Dr. Hall-Ellis lacks personal knowledge of the ResearchGate date and speculates based on unreliable evidence

“However, **absent any evidence pertaining to the general practices** of the Copyright Office, Westlaw, and Dialog, or the typical time that elapses between copyright registration, inclusion in the Copyright Office's automated catalog, and subsequent incorporation into one of the commercial databases, **any presumption along those lines would be pure speculation.**”

In re Lister, 583 F.3d 1307, 1317 (Fed. Cir. 2009)



18 So, to say the date is accurate or
19 it is close, or it is not close, I think that is
20 asking a lot from the authors who are very busy,
21 who really do what they do when they do it.

EX2011 (Hall-Ellis Dep. Tr. at 54:6-56-18)

15 A. I would never take these solely on
16 face value, never. Everything has to be
17 investigated more fully.

EX2011 (Hall-Ellis Dep. Tr. at 58:10-17)

12 How carefully we transcribe this, I
13 think varies from person to person.

EX2011 (Hall-Ellis Dep. Tr. at 59:6-13)

Paper 44 (Sur-Reply) at 24-25



Ground 1: Obviousness over Zhao

Ground	Assertion	Petitioned Claims
1	Obvious under §103 over Zhao	1-16
2	Obvious under §103 over Steffens in view of Zhao	1-16
3	Obvious under §103 over Ozer in view of Zhao	1-16
4	Obvious under §103 over TLP	1-16

Zhao: Overview

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

1.2 Challenges and Strategies

1.2.1 Why is it difficult to detect humans

Although humans can easily detect other humans and estimate their body part locations from a single image, these problems are inherently difficult for a computer. The difficulties

EX1006 (Zhao) at 20

Shape decomposition should result in real/intuitive parts of an object (as shown in Fig. 2.1). However, without prior knowledge of the identification of the object, this goal is hard to achieve. Many shape decomposition approaches tend to produce over-segmented parti-

EX1006 (Zhao) at 23

4.1 Why Do We Need Contextual Information ?

High performance object detection depends on reliable contour extraction, but contour extraction is an under-constrained problem without knowledge about the objects to be detected. The experimental results given in Chapters 2 and 3 also illustrate that higher level

EX1006 (Zhao) at 69

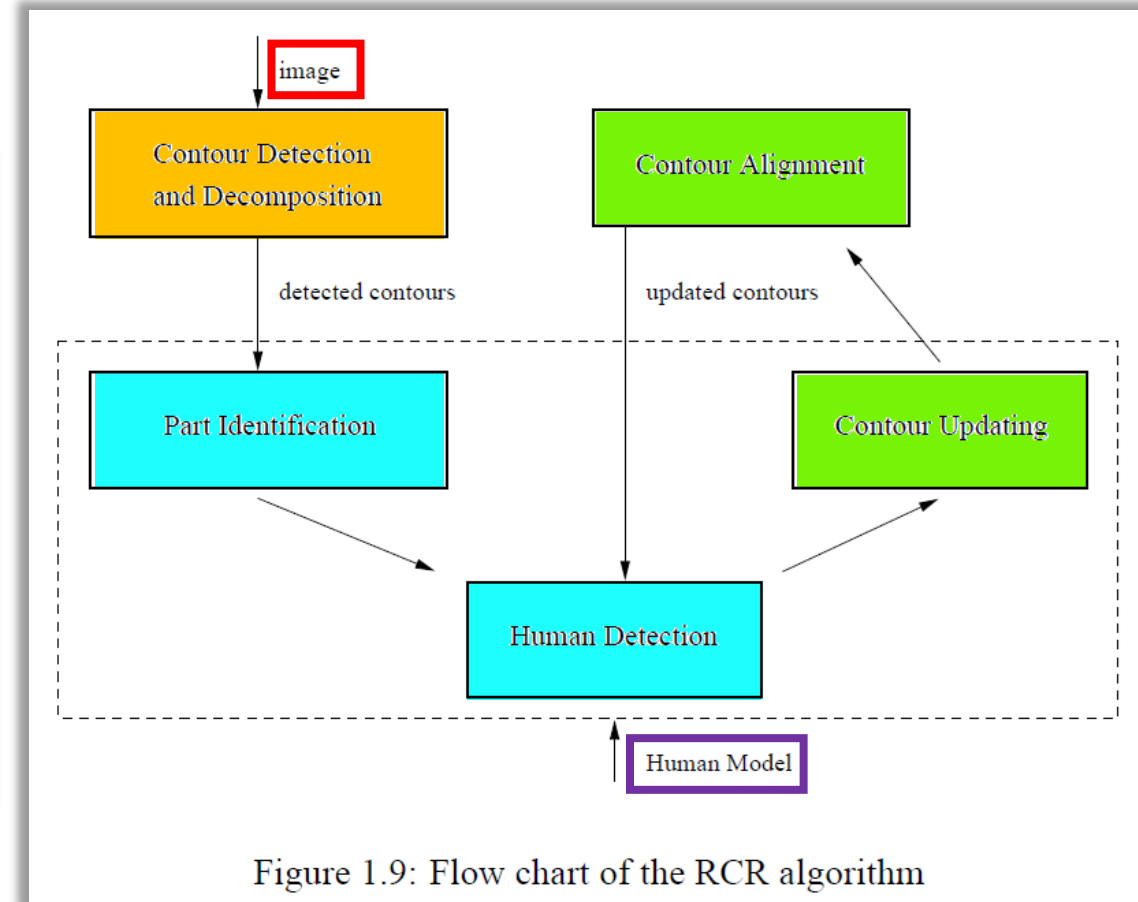
Paper 26 (POR) at 10-11

Zhao: Overview

Recursive Context Reasoning

This thesis proposes a recursive context reasoning (RCR) algorithm to solve the above dilemma. A TRS-invariant probabilistic model is designed to encode the shapes of the body parts and the context information — the size and spatial relationships between body parts. A Bayesian framework is developed to perform human detection and part identification under partial occlusion. A contour updating procedure is introduced to integrate the human model and the identified body parts to predict the shapes and locations of the parts missed by the contour detector; the refined contours are used to reevaluate the Bayesian similarity measure and determine if the detected contour is a person or not. Therefore, contour extraction, body part localization, and human detection are improved by combining the context constraints from the identified body parts and the human model.

EX1006 (Zhao) at 69



EX1006 (Zhao) at 26

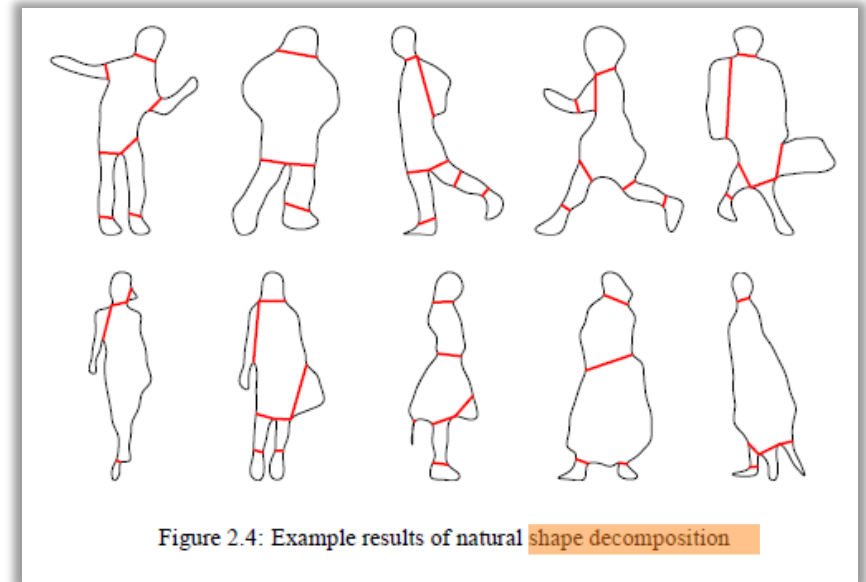
Paper 26 (POR) at 10-11

Zhao: Overview

Recursive Context Reasoning

This thesis proposes a recursive context reasoning (RCR) algorithm to solve the above dilemma. A TRS-invariant probabilistic model is designed to encode the shapes of the body parts and the context information — the size and spatial relationships between body parts. A Bayesian framework is developed to perform human detection and part identification under partial occlusion. A contour updating procedure is introduced to integrate the human model and the identified body parts to predict the shapes and locations of the parts missed by the contour detector; the refined contours are used to reevaluate the Bayesian similarity measure and determine if the detected contour is a person or not. Therefore, contour extraction, body part localization, and human detection are improved by combining the context constraints from the identified body parts and the human model.

EX1006 (Zhao) at 69



EX1006 (Zhao) at 40

to consider what features to use for recognition. Contours (the silhouettes of objects) are commonly used to overcome variable texture. Therefore, I employ the occluding contours or the silhouettes of the human body parts as features to detect humans. Parts-based approaches [21, 25, 31, 58, 78, 80] can handle occlusion and articulated motion effectively. The importance and relevance of significant parts for object recognition has been verified by numerous psychophysical experiments [48]. My recognition scheme is thus based on the shapes of body parts and the relationships between them. Then the questions left are how

EX1006 (Zhao) at 22

Zhao: Overview

Recursive Context Reasoning

This thesis proposes a recursive context reasoning (RCR) algorithm to solve the above dilemma. A TRS-invariant probabilistic model is designed to encode the shapes of the body parts and the context information — the size and spatial relationships between body parts. A Bayesian framework is developed to perform human detection and part identification under partial occlusion. A contour updating procedure is introduced to integrate the human model and the identified body parts to predict the shapes and locations of the parts missed by the contour detector; the refined contours are used to reevaluate the Bayesian similarity measure and determine if the detected contour is a person or not. Therefore, contour extraction, body part localization, and human detection are improved by combining the context constraints from the identified body parts and the human model.

EX1006 (Zhao) at 69

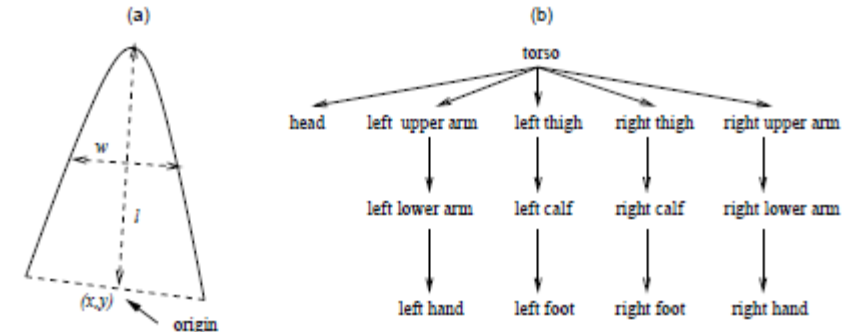


Figure 2.6: (a) Body part model (b) “connect-to” hierarchy

2.4.1 TRS-Invariant Body Model

Each 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. The body parts are modeled with ribbons¹ [114, 115] as shown in Fig. 2.6(a). The width (w) of a ribbon is defined as the average width along the ribbon, and its length (l) and major axis come from the ribbon spine. The aspect ratio $a = w/l$ is invariant under similarity transforms, and it captures the global shape of a ribbon while ignoring small local shape deformations. Thus the aspect ratio is appropriate for the purpose of recognition.

EX1006 (Zhao) at 42

Zhao: Overview

Recursive Context Reasoning

This thesis proposes a recursive context reasoning (RCR) algorithm to solve the above dilemma. A TRS-invariant probabilistic model is designed to encode the shapes of the body parts and the context information — the size and spatial relationships between body parts. A Bayesian framework is developed to perform human detection and part identification under partial occlusion. A contour updating procedure is introduced to integrate the human model and the identified body parts to predict the shapes and locations of the parts missed by the contour detector; the refined contours are used to reevaluate the Bayesian similarity measure and determine if the detected contour is a person or not. Therefore, contour extraction, body part localization, and human detection are improved by combining the context constraints from the identified body parts and the human model.

EX1006 (Zhao) at 69

3.4 Bayesian Similarity Measure for Human Detection

3.4.1 Decision Rule

The human detection problem is: given a detected contour C , determine whether this contour represents the silhouette of a person or not.

A simple decision rule to perform human detection is: the contour C corresponds to a person if

$$P(\text{person}|C) > \text{threshold} \quad (3.11)$$

Otherwise, C is not a person. The threshold controls the receiver operating characteristics (ROC) of the detector. According to Bayes' rule,

$$P(\text{person}|C) = \frac{P(C|\text{person})P(\text{person})}{P(C)} \quad (3.12)$$

Assume that all contours have the same prior, then $P(\text{person})/P(C)$ can be considered as a constant, and the decision rule becomes: the contour C corresponds to a person if

$$P(C|\text{person}) > \text{threshold}. \quad (3.13)$$

The likelihood $P(C|\text{person})$ can be calculated by conditioning on the hypotheses for body part identification:

$$P(C|\text{person}) = \sum_{H \in \mathcal{H}} P(C|H, \text{person})P(H|\text{person}), \quad (3.14)$$

EX1006 (Zhao) at 64

Zhao: Overview

Recursive Context Reasoning

This thesis proposes a recursive context reasoning (RCR) algorithm to solve the above dilemma. A TRS-invariant probabilistic model is designed to encode the shapes of the body parts and the context information — the size and spatial relationships between body parts. A Bayesian framework is developed to perform human detection and part identification under partial occlusion. A contour updating procedure is introduced to integrate the human model and the identified body parts to predict the shapes and locations of the parts missed by the contour detector; the refined contours are used to reevaluate the Bayesian similarity measure and determine if the detected contour is a person or not. Therefore, contour extraction, body part localization, and human detection are improved by combining the context constraints from the identified body parts and the human model.

EX1006 (Zhao) at 69

4.3 Update the Shapes and Locations of the Body Parts

The goal here is to demonstrate that the identified human body parts and the human model can be combined to refine the shapes and locations of the identified body parts and to predict the shapes and locations of the missed body parts.

EX1006 (Zhao) at 72

Note that the RCR algorithm is different from the EM algorithm [111], a widely used approach to learning in the presence of unobserved variables. In the EM algorithm, if some variable is sometimes observable and sometimes not, then we can use the cases for which it has been observed to learn to predict its values when it is not. However, in the RCR algorithm we use the detected body parts and the geometrical relationships between the detected parts and undetected parts to predict the parameters of undetected parts.

EX1006 (Zhao) at 71

Third, the locations of the joints are adjusted to achieve consistency with the modeled spatial and size relationships between the body parts. The locations and the sizes of the missed body parts are inferred from the extended body parts and the detected body parts (shown in Fig. 5.4(d)). Fourth, the predicted outlines of the body parts are aligned with the

EX1006 (Zhao) at 91

Zhao Does Not Disclose or Render Obvious Element 1/9[a] or 1/9[e]

Missing Claim Elements	Claims
1/9[a] “providing a spatio-temporal model for an object to be detected”	1 and 9
1/9[e] “detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object”	1 and 9



Zhao Does Not Disclose or Render Obvious Element 1/9[a] or 1/9[e]

Term	Patent Owner's Construction
“detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object”	detecting the object in any image <u>based on comparing the probabilities for each image of the plurality of images</u> to a threshold for detecting the object
“spatio-temporal model”	spatial model where an object is detected in an image <u>based on the probabilities that the object is in preceding and/or subsequent images</u>

Zhao’s probabilistic analysis is confined to a single image

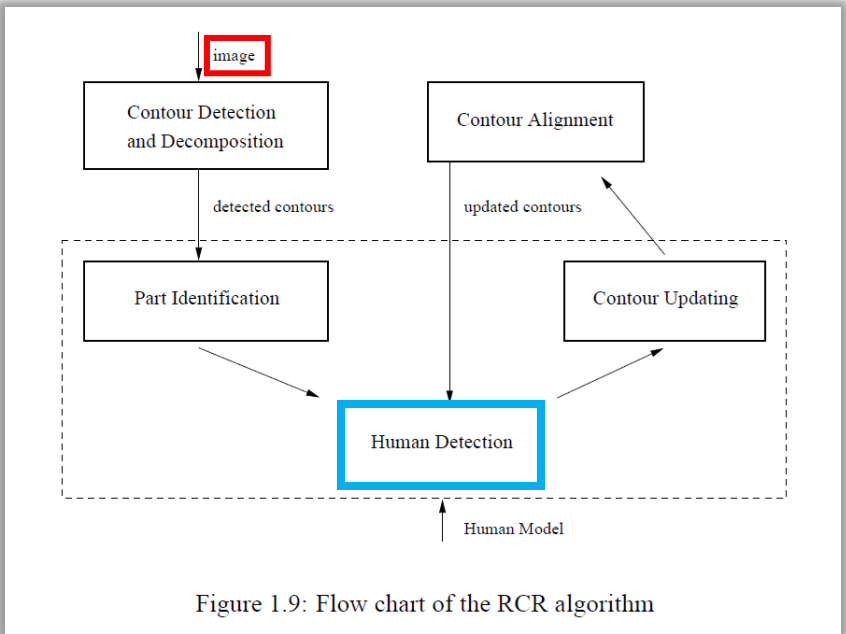


Figure 1.9: Flow chart of the RCR algorithm

EX1006 (Zhao) at 26

I develop a recursive context reasoning (RCR) algorithm (see Fig. 1.9) to solve the above dilemma. Specifically, a contour updating procedure is introduced to integrate the detected body parts and the human model using the weighted Least Squares method to predict the contours of the missed parts. The predicted contours are aligned with edge features, and the refined contours are used to reevaluate the likelihood of a person being present in the image. Therefore, both contour extraction, human detection, and body part localization are

EX1006 (Zhao) at 26

in subsection 3.3.3. This approximation works well in the situation of low noise and unambiguous data. Because the recursive context reasoning algorithm explained in Chapter 4 will improve the accuracy of contour extraction iteratively, the above approximation will get more accurate accordingly.

EX1006 (Zhao) at 64-65

Zhao Does Not Disclose or Render Obvious Element 1/9[a] or 1/9[e]

Term	Patent Owner's Construction
“detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object”	detecting the object in any image <u>based on comparing the probabilities for each image of the plurality of images</u> to a threshold for detecting the object
“spatio-temporal model”	spatial model where an object is detected in an image <u>based on the probabilities that the object is in preceding and/or subsequent images</u>

Zhao's contour updating and alignment does not involve probabilities from other images

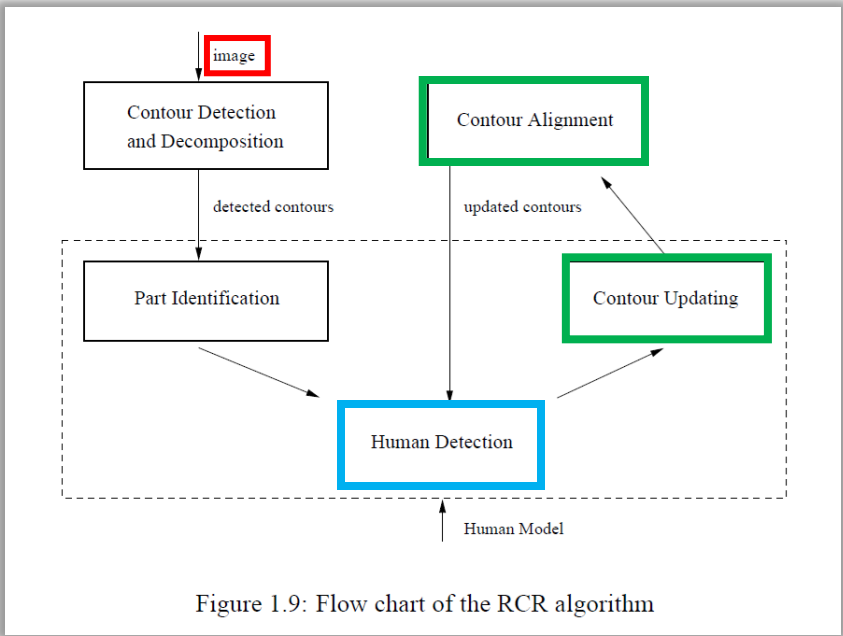


Figure 1.9: Flow chart of the RCR algorithm

EX1006 (Zhao) at 26

4.3 Update the Shapes and Locations of the Body Parts

The goal here is to demonstrate that the identified human body parts and the human model can be combined to refine the shapes and locations of the identified body parts and to predict the shapes and locations of the missed body parts.

EX1006 (Zhao) at 72

it has been observed to learn to predict its values when it is not. However, in the RCR algorithm we use the detected body parts and the geometrical relationships between the detected parts and undetected parts to predict the parameters of undetected parts.

EX1006 (Zhao) at 71

Third, the locations of the joints are adjusted to achieve consistency with the modeled spatial and size relationships between the body parts. The locations and the sizes of the

EX1006 (Zhao) at 91

Zhao Does Not Disclose or Render Obvious Element 1/9[a] or 1/9[e]

Term	Patent Owner's Construction
“detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object”	detecting the object in any image <u>based on comparing the probabilities for each image of the plurality of images</u> to a threshold for detecting the object
“spatio-temporal model”	spatial model where an object is detected in an image <u>based on the probabilities that the object is in preceding and/or subsequent images</u>

Zhao's body model does not combine probabilities from multiple images

37. Petitioners' argument that “Zhao is directed to much more than a ‘simple fusion scheme’” is incorrect. Reply, 8. Zhao's RCR algorithm does not fuse or combine *object* probabilities from *multiple* frames. The probabilistic analysis is confined to the *single frame* under consideration, as I explained in my previous declaration. EX2004, ¶¶81-85, 107-112. Zhao's body model also does not combine object probabilities from multiple frames—the elements of the *body* model have no temporal components at all, as indicated by the lack of temporal indicia. See EX1006 at 43-45; EX2004, ¶¶75-78, 111, 126.

The POSITA would understand that because Zhao's body model and RCR algorithm do not include temporal probabilistic connectivity, Zhao does not disclose a spatio-temporal model. Petitioners' argument that it does rests on Dr. Bajaj's incorrect construction of “spatio-temporal model,” which encompasses a purely spatial model with no temporal connectivity (not spatio-temporal) that is used on each single frame independently. Reply, 8 (citing EX1052, ¶¶6-35). As I explained above in Section VI, Dr. Bajaj's construction is incorrect.

EX2007 (Saber Supp. Decl.) ¶ 37

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

EX1006 (Zhao) at 43

independent of each other. Based on the above assumptions, the probability distributions of the three model matrices A, S, U can be estimated separately:

$$A \sim \mathcal{N}(\bar{A}, \Sigma_A) \tag{2.4}$$
$$S \sim \mathcal{N}(S, \Sigma_S) \tag{2.5}$$
$$U \sim \mathcal{N}(\bar{U}, \Sigma_U). \tag{2.6}$$

Assuming that the variations of the aspect ratios of the body parts are statistically independent², then

$$N(\bar{A}, \Sigma_A) = \prod_i N(a_i; \sigma_{a_i}^2) \tag{2.7}$$

Assuming that the variations of the ratios of the lengths between the body parts and a reference body part f_i are independent³, then

$$N(S_i, \Sigma_{S_i}) = \prod_{j \neq i} N(\bar{s}_{ji}; \sigma_{s_{ji}}^2) \tag{2.8}$$

EX1006 (Zhao) at 43-44

Paper 44 (Sur-Reply) at 28

Zhao Does Not Disclose or Render Obvious Element 1/9[a] or 1/9[e]

Term	Patent Owner's Construction
"detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object"	detecting the object in any image <u>based on comparing the probabilities for each image of the plurality of images</u> to a threshold for detecting the object
"spatio-temporal model"	spatial model where an object is detected in an image <u>based on the probabilities that the object is in preceding and/or subsequent images</u>

Zhao's motion information does not include probabilities

operates on a single frame at a time. And while the RCR process can use some limited information from previous frames, Zhao indicates that such information is "motion information." EX1006, 91-92. Such motion information does not include probabilities and creates no temporal probabilistic connectivity. The POSITA would therefore understand that Zhao's "motion information," e.g., "how the left knee angle changes with time" (i.e., angular velocity) does not include object or component probabilities from other frames. *Id.* Nor did I "acknowledge" that Zhao

EX2007 (Saber Supp. Decl.) ¶ 37

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.

EX1006 (Zhao) at 91-92

Paper 44 (Sur-Reply) at 28



Zhao Does Not Disclose or Render Obvious Claim 4/12

4. The computer implemented method of claim 1, wherein determining the probability that the object is in each image comprises detecting the object in a current image according to measurements of the object as a collection of components determined from a prior image and a later image relative to the current image.

12. The method of claim 9, wherein determining the probability that the object is in each image comprises detecting the object in a current image according to measurements of the object as a collection of components determined from a prior image and a later image relative to the current image.

EX1001 ('980 patent) at claims 4, 12



Zhao Does Not Disclose or Render Obvious Claim 4/12

Zhao does not disclose or suggest the use of later images

Zhao's MAP estimation does not involve later images

118. Zhao's description of its MAP estimation, however, **does not disclose or suggest the use of information from a later image**. EX1006, 74-75. The transformation in Zhao's MAP estimation simply **transforms from the model's coordinate system to the coordinate system of a given image**. EX1006, 74. Claims

EX2004 (Saber Decl.) ¶ 118

The location of the main body part f_j is estimated from the locations of the other main body parts. If more than two main body parts have been identified, then **the transformation T that transforms the coordinates from the human model coordinate system to the image coordinate system** can be estimated by the Least Squares Method. Let $\hat{X}_I = (p_1, p_2, \dots, p_i)$ be the coordinates of the identified parts in the image coordinate system, and let $\bar{U}_I = (q_1, q_2, \dots, q_i)$ be the coordinates of the identified parts in the model coordinate system, where $p_i = (\hat{x}_i, \hat{y}_i, 1)^T$, $q_i = (\bar{u}_i, \bar{v}_i, 1)^T$. Then

$$\hat{X}_I = T\bar{U}_I \quad (4.8)$$

$$\hat{T} = \hat{X}_I \bar{U}_I^T (\bar{U}_I \bar{U}_I^T)^{-1} \quad (4.9)$$

And the position of the unidentified body part f_j is estimated as

$$(\hat{x}_j, \hat{y}_j, 1)^T = \hat{T}(\bar{u}_j, \bar{v}_j, 1)^T. \quad (4.10)$$

EX1006 (Zhao) at 74-75

Zhao Does Not Disclose or Render Obvious Claim 4/12

Zhao does not disclose or suggest the use of later images

Zhao's motion information only involves previous images

119. Zhao's use of "[m]otion information obtained from **previous frames ... to predict the orientations of the limbs and to solve the ambiguity**" also does not disclose or suggest the use of measurements from "a **later image relative to the current image.**" EX1004 ¶107. Dr. Bajaj identifies no disclosure in Zhao of the use of measurements from a later image. EX1004 ¶¶106-110. And although Dr. Bajaj

EX2004 (Saber Decl.) ¶ 118

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.

EX1006 (Zhao) at 91-92 (cited in EX1004 ¶107)

Zhao Does Not Disclose or Render Obvious Claim 4/12

Zhao does not disclose or suggest the use of later images

Zhao does not discuss “adjacent frames”; Zhao’s “adjacent features” are within the same frame

39. Petitioners claim that Patent Owner ignores “Zhao’s disclosure” of using measurements from “adjacent frames,” where “adjacent frames” (according to Petitioners) “necessarily includes previous and later frames.” Reply, 11. Yet Petitioners cite no “disclosure” in Zhao (and I am aware of no such disclosure) regarding “adjacent frames” in Zhao. *Id.*; Pet. 29; EX1004, ¶107; EX1052 ¶¶36-39. Petitioners make a new argument in their Reply based on Zhao’s discussion of “adjacent features,” but Petitioners misinterpret Zhao. The “adjacent features” are “features” within model (analogous to a model’s parts or components) used to perform “feature matching” in the image frame under consideration—there is no discussion of adjacent frames or looking to adjacent frames, let alone previous *and* later frames. EX1006 at 90.

EX2007 (Saber Supp. Decl.) ¶ 39

In contrast, the RCR algorithm proposed in this thesis is a combination of the labeling and alignment methods. The algorithm achieves both efficient and accurate body part localization by making use of the advantages of the above two approaches and overcoming their disadvantages. The combined approach has been used in [109] for 2D-2D feature matching and 2D-3D matching in [85]. The common feature is that the uncertainty information has been used for feature matching and for determining transformation uncertainty with which to predict the positions of adjacent features. The main differences of the RCR algorithm

EX1006 (Zhao) at 90



Zhao Does Not Disclose or Render Obvious Claim 4/12

Petitioners' non-asserted references do not disclose the use of prior and later images for object detection

Baumberg's (EX1011) trained vibration modes and forward extrapolation do not teach the use of prior and later images for object detection

41. EX1011 addresses a technique for “tracking a non-rigid deforming object” to purportedly improve over prior techniques in which “it is not possible to extrapolate *forward* in time.” EX1011, 1-2 (emphasis added). The technique is based on modal analysis (using “vibration modes”), and it generates “trained vibration modes” from a set of training data that do not assume constant elasticity of the object. EX1011, 1-6, 9. The training data includes observations of nodal (point) displacements for the object at multiple points in time. *Id.* at 2-3. It’s use of training data “assume[s] an observed set of matched, aligned shape vectors consisting of nodal (or point) positions” observed at multiple points in time. *Id.* This technique for tuning vibration modes using training data does not teach or suggest determining the probability that an object is in an image, nor doing so by detecting the object in a current image using measurements of the object from prior and later images.

EX2007 (Saber Supp. Decl.) ¶ 41

1 Introduction

We are interested in accurately tracking a non-rigid deforming object. Pentland and

In this paper we describe a method that generates physically based “vibration modes” from a set of training examples of an object deforming, utilising a single assumption of constant (unknown) density. The resulting vibration modes provide a good basis for the types of motion represented in the training set (e.g. walking). We show the results of our method on a real example, modeling the 2D shape of a walking pedestrian using automatically acquired (noisy) training data. The model retains the benefits of conventional modal analysis (e.g. low dimensional parametrisation, decoupled filter mechanism for rapid tracking), whilst utilising the training information to improve accuracy. The training set allows the rejection of physical assumptions being made about the object (e.g. modeling a walking person as a simple lump of elastic “clay”).

A related training based approach to modeling deformable shape is the Point Distribution Model (PDM) described by Cootes and Taylor [4]. The PDM has proven useful in image sequence analysis (e.g. real-time contour tracking [5], robust tracking of deformable models [6]). However, one drawback of this approach is that there is no temporal aspect to the model. Hence it is not possible to extrapolate forward in time to obtain good estimates for the expected shape of the object. Furthermore the purely spatial model can not discrim-

EX1011 (Baumberg) at 1-2

Paper 44 (Sur-Reply) at 29



Zhao Does Not Disclose or Render Obvious Claim 4/12

Petitioners' non-asserted references do not disclose the use of prior and later images for object detection

“Attractive People” (EX1015) teaches single-frame “static” detection; forward-backward smoothing is suggested for tracking

42. His new evidence from EX1015 is a sentence that mentions “forward-backward smoothing” in an article co-authored by one of the inventors of the '980 patent. The forward-backward smoothing mentioned in EX1015 relates to extending a model to track components (*i.e.*, follow detected components across frames), but the component detection is still performed within a single frame on a frame-by-frame basis (“static detection”). EX1015 at 2, 4, 6-7. The authors proposed a “loose-limbed” body model where “each limb is treated quasi-independently with soft constraints between the position and orientation of adjacent parts,” but indicated that detection techniques were “beyond the scope of this paper.” *Id.* at 4. This loose-limbed body model does not teach or suggest determining the probability that an object is in an image or doing so by detecting the object in a current image using measurements of the object from prior and later images.

EX2007 (Saber Supp. Decl.) ¶ 42

We develop the loose-limbed model in detail, formulate the constraints between limbs using mixture models, and outline the inference method. Using images from calibrated cameras we illustrate the inference of 3D human pose with belief propagation. We simulate noisy, bottom-up, feature detectors for the limbs and show how the inference method can resolve ambiguities and cope with clutter. While our focus here is on static detection and pose estimation, the body model can be extended in time to include temporal constraints on the limb motion; we save tracking for future work.

EX1015 (“Attractive People”) at 2

We illustrate the approach by recovering 3D body pose given weak bottom-up information and clutter. The development of bottom-up part detectors is beyond the scope of this paper. Here we exploit a realistic simulation of such detectors in which: 1) the limbs are

EX1015 (“Attractive People”) at 6

It is straightforward to extend the graphical model across time to implement a person 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.

In future work we intend to build automatic body-part detectors. Constructing reliable

EX1015 (“Attractive People”) at 6-7

Paper 44 (Sur-Reply) at 29



Zhao Does Not Disclose or Render Obvious Claim 8/16

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

$$P(X_0^O, X_0^{C0}, X_0^{C1}, \dots, X_0^{CN}, \dots, X_T^O, X_T^{C0}, X_T^{C1}, \dots, X_T^{CN}) = \frac{1}{Z} \prod_{ij} \psi_{ij}(X_i^O, X_j^O) \prod_{ik} \psi_{ik}(X_i^O, X_i^{Ck}) \prod_{ikl} \psi_{kl}(X_i^{Ck}, X_i^{Cl}) \prod_i \phi_i(X_i^O) \prod_{ik} \phi_i(X_i^{Ck}).$$

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

$$P(X_0^O, X_0^{C0}, X_0^{C1}, \dots, X_0^{CN}, \dots, X_T^O, X_T^{C0}, X_T^{C1}, \dots, X_T^{CN}) = \frac{1}{Z} \prod_{ij} \psi_{ij}(X_i^O, X_j^O) \prod_{ik} \psi_{ik}(X_i^O, X_i^{Ck}) \prod_{ikl} \psi_{kl}(X_i^{Ck}, X_i^{Cl}) \prod_i \phi_i(X_i^O) \prod_{ik} \phi_i(X_i^{Ck}).$$

EX1001 ('980 patent) at claims 8, 16

Zhao Does Not Disclose or Render Obvious Claim 8/16

Zhao's probability distribution is not “representable” in the form of claim 8/16

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

The coordinates (x_i, y_i) of a subpart f_i (for example the lower arm) in the coordinate frame of its parent f_j (the upper arm) is simply:

$$(x_i, y_i) = (0, l_j). \quad (2.3)$$

Because the lengths of the body parts are constrained by the length ratio matrix S , only the relative positions of the six main body parts needs to be modeled. Let $X = (x_1, y_1, \dots, x_6, y_6)$. To make this vector TRS-invariant, the coordinates of the joints are represented in a normalized torso coordinate system with the length of the torso normalized to be 1. Then the TRS-invariant relative positions of the six main body parts are $U = (0, 0, u_2, v_2, \dots, u_6, v_6)$, where $(u_i, v_i) = ((x_i, y_i) - (x_1, y_1))/l_1$, (x_1, y_1) are the coordinates of the torso's center, and l_1 is the length of the torso.

EX1006 (Zhao) at 44

independent of each other. Based on the above assumptions, the probability distributions of the three model matrices A, S, U can be estimated separately:

$$A \sim \mathcal{N}(\bar{A}, \Sigma_A) \quad (2.4)$$

$$S \sim \mathcal{N}(S, \Sigma_S) \quad (2.5)$$

$$U \sim \mathcal{N}(U, \Sigma_U). \quad (2.6)$$

Assuming that the variations of the aspect ratios of the body parts are statistically independent², then

$$N(\bar{A}, \Sigma_A) = \prod_i N(a_i; \sigma_{a_i}^2) \quad (2.7)$$

Assuming that the variations of the ratios of the lengths between the body parts and a reference body part f_i are independent³, then

$$N(S_i, \Sigma_{S_i}) = \prod_{j \neq i} N(s_{ji}; \sigma_{s_{ji}}^2) \quad (2.8)$$

where $S_i = \{s_{ji}\}$, $s_{ji} = l_j/l_i$, $j = 1, \dots, m$, $j \neq i$. Thus, S_i is used to constrain the relative lengths of the body parts with respect to the length l_i of a reference part f_i .

EX1006 (Zhao) at 45

Paper 26 (POR) at 11 (citing EX2004, ¶¶75-78, 82), 38-40; Paper 44 (Sur-Reply) at 30

Zhao Does Not Disclose or Render Obvious Claim 8/16

Zhao’s probability distribution is not “representable” in the form of claim 8/16

independent of each other. Based on the above assumptions, the probability distributions of the three model matrices A, S, U can be estimated separately:

$$A \sim \mathcal{N}(\bar{A}, \Sigma_A) \tag{2.4}$$

$$S \sim \mathcal{N}(\bar{S}, \Sigma_S) \tag{2.5}$$

$$U \sim \mathcal{N}(\bar{U}, \Sigma_U). \tag{2.6}$$

Assuming that the variations of the aspect ratios of the body parts are statistically independent², then

$$N(\bar{A}, \Sigma_A) = \prod_i N(\bar{a}_i; \sigma_{a_i}^2) \tag{2.7}$$

Assuming that the variations of the ratios of the lengths between the body parts and a reference body part f_i are independent³, then

$$N(\bar{S}_i, \Sigma_{S_i}) = \prod_{j \neq i} N(\bar{s}_{ji}; \sigma_{s_{ji}}^2) \tag{2.8}$$

where $S_i = \{s_{ji}\}$, $s_{ji} = l_j/l_i$, $j = 1, \dots, m$, $j \neq i$. Thus, S_i is used to constrain the relative lengths of the body parts with respect to the length l_i of a reference part f_i .

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

Claim 8/16		Zhao
$\prod_{ij} \psi_{ij}(X_i^O, X_j^O)$	“temporal compatibility of object state between frames i and j”	✗
$\prod_{ik} \psi_{ik}(X_i^O, X_i^{C_k})$	“spatial compatibility of the object and it’s components at frame i”	✗
$\prod_{ikl} \psi_{kl}(X_i^{C_k}, X_i^{C_l})$	“spatial compatibility between object components at frame i”	✗
$\prod_i \phi_i(X_i^O)$	“local evidence for the object ... state[]”	✗
$\prod_{ik} \phi_i(X_i^{C_k})$	“local evidence for the ... component states”	✗

EX1006 (Zhao) at 45

EX1001 (’980 patent) at 3:47-4:8

Paper 26 (POR) at 11 (citing EX2004, ¶¶75-78, 82), 38-40; Paper 44 (Sur-Reply) at 30



Zhao Does Not Disclose or Render Obvious Claim 8/16

Petitioners’ non-asserted references do not disclose or render obvious claim 8/16

“PAMPAS” (EX1019)

functions $\phi_i(x_i, y_i) \equiv \phi_i(x_i) : \mathcal{L} \rightarrow \mathbb{R}$. Adjacent hidden nodes are related by the correlation functions $\psi_{ij}(x_j, x_i) : \mathcal{L} \times \mathcal{L} \rightarrow \mathbb{R}$. The joint probability over $\mathcal{X}$ is given by

$$P(\mathcal{X}) = \frac{1}{Z} \left(\prod_{(i,j) \in \mathcal{P}} \psi_{ij}(x_j, x_i) \right) \left(\prod_i \phi_i(x_i) \right). \quad (1)$$

The belief propagation algorithm introduces variables such as $m_{ij}(x_j)$ which can intuitively be understood as a *message* from hidden node i to hidden node j about what state node j should be in. Messages are computed iteratively using the update algorithm

$$m_{ij}(x_j) \leftarrow \sum_{x_i \in \mathcal{L}} \left(\phi_i(x_i) \psi_{ij}(x_j, x_i) \prod_{k \in \mathcal{P}(i) \setminus j} m_{ki}(x_i) \right) \quad (2)$$

and the marginal distribution of X_i (the node’s *belief*) is given by

$$b_i(x_i) \propto \phi_i(x_i) \prod_{j \in \mathcal{P}(i)} m_{ji}(x_i). \quad (3)$$

EX1019 (“PAMPAS”) at 2

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

Claim 8/16		“PAMPAS”
$\prod_{ij} \psi_{ij}(X_i^O, X_j^O)$	“temporal compatibility of object state between frames i and j”	✗
$\prod_{ik} \psi_{ik}(X_i^O, X_i^{C_k})$	“spatial compatibility of the object and it’s components at frame i”	✗
$\prod_{ikl} \psi_{kl}(X_i^{C_k}, X_i^{C_l})$	“spatial compatibility between object components at frame i”	✗
$\prod_i \phi_i(X_i^O)$	“local evidence for the object ... state[]”	✗
$\prod_{ik} \phi_i(X_i^{C_k})$	“local evidence for the ... component states”	✗

EX1001 (’980 patent) at 3:47-4:8

Paper 26 (POR) at 11 (citing EX2004, ¶¶75-78, 82), 38-40; Paper 44 (Sur-Reply) at 30



Zhao Does Not Disclose or Render Obvious Claim 8/16

Petitioners’ non-asserted references do not disclose or render obvious claim 8/16

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

“Sudderth” (EX1012)

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

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

However, the nonparametric updates we present may be directly extended to models with higher-order potentials.

EX1012 (Sudderth) at 2

2.1. Belief Propagation

For graphs which are acyclic or tree-structured, the desired conditional distributions $p(x_s \mid y)$ can be directly calculated by a local message-passing algorithm known as *belief propagation* (BP) [16, 24]. At iteration n of the BP algorithm, each node $t \in \mathcal{V}$ calculates a message $m_{ts}^n(x_s)$ to be sent to each neighboring node $s \in \Gamma(t)$:

$$m_{ts}^n(x_s) = \alpha \int_{x_t} \psi_{s,t}(x_s, x_t) \psi_t(x_t, y_t) \times \prod_{u \in \Gamma(t) \setminus s} m_{ut}^{n-1}(x_t) dx_t \quad (2)$$

Here, α denotes an arbitrary proportionality constant. At any iteration, each node can produce an approximation $\hat{p}^n(x_s \mid y)$ to the marginal distribution $p(x_s \mid y)$ by combining the incoming messages with the local observation:

$$\hat{p}^n(x_s \mid y) = \alpha \psi_s(x_s, y_s) \prod_{t \in \Gamma(s)} m_{ts}^n(x_s) \quad (3)$$

Claim 8/16		“Sudderth”
$\prod_{ij} \psi_{ij}(X_i^O, X_j^O)$	“temporal compatibility of object state between frames i and j”	✗
$\prod_{ik} \psi_{ik}(X_i^O, X_i^{C_k})$	“spatial compatibility of the object and it’s components at frame i”	✗
$\prod_{ikl} \psi_{kl}(X_i^{C_k}, X_i^{C_l})$	“spatial compatibility between object components at frame i”	✗
$\prod_i \phi_i(X_i^O)$	“local evidence for the object ... state[]”	✗
$\prod_{ik} \phi_i(X_i^{C_k})$	“local evidence for the ... component states”	✗

EX1001 (’980 patent) at 3:47-4:8

Paper 26 (POR) at 11 (citing EX2004, ¶¶75-78, 82), 38-40; Paper 44 (Sur-Reply) at 30

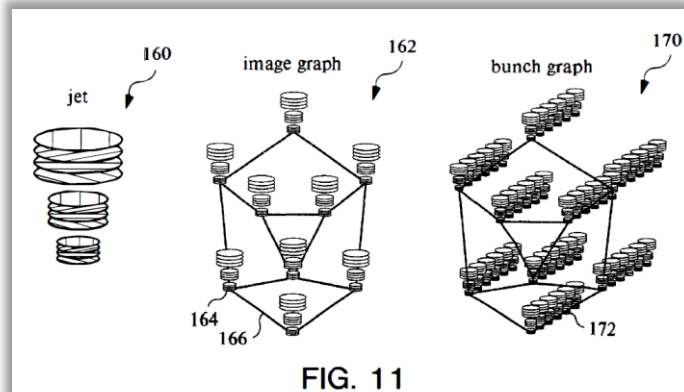


Ground 2: Obviousness over Zhao + Steffens

Ground	Assertion	Petitioned Claims
1	Obvious under §103 over Zhao	1-16
2	Obvious under §103 over Steffens in view of Zhao	1-16
3	Obvious under §103 over Ozer in view of Zhao	1-16
4	Obvious under §103 over TLP	1-16

Steffens: Overview

Steffens teaches a non-probabilistic wavelet-based graph matching approach

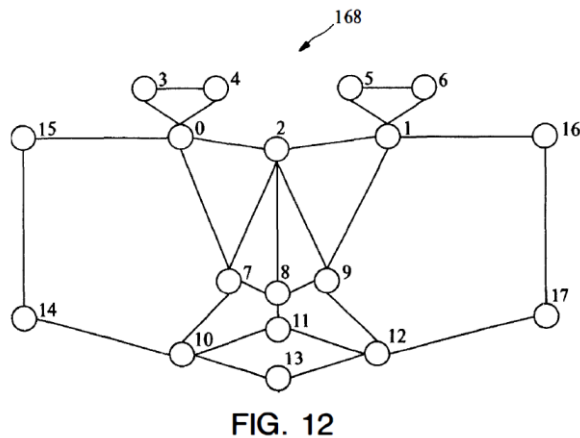


similarity (block 146). The search for the optimum similarity may be performed by positioning the model graph in the upper left hand corner of the image, extracting the jets at the nodes, and determining the similarity between the image graph and the model graph. The search continues by sliding

EX1007 (Steffens) at 7:51-67

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

EX1007 (Steffens) at 3:42-44



The wavelet is a plane wave with wave vector $\mathbf{k}$ restricted by a Gaussian window, the size of which relative to the wavelength is parameterized by σ . The term in the brace removes the DC component. The amplitude of the wavevector $\mathbf{k}$ may be chosen as follows where $\mathbf{v}$ is related to the desired spacial resolutions.

EX1007 (Steffens) at 8:14-67

A labeled image graph 162, as shown in FIG. 11, is used to describe the aspects of an object (in this context, a face). The nodes 164 of the labeled graph refer to points on the object and are labeled by jets 160. Edges 166 of the graph are labeled with distance vectors between the nodes. Nodes and edges define the graph topology. Graphs with equal geometry may be compared. The normalized dot product of the absolute components of two jets defines the jet similarity. This value is independent of the illumination and contrast changes. To compute the similarity between two graphs, the sum is taken over similarities of corresponding jets between the graphs.

EX1007 (Steffens) at 8:14-67

In order to find a face in an image frame, the graph is moved and scaled over the image frame until a place is located at which the graph matches best (the best fitting jets within the bunch jets are most similar to jets extracted from the image at the current positions of the nodes). Since face

EX1007 (Steffens) at 9:37-41

EX1007 (Steffens) at Figs. 11-12

Paper 26 (POR) at 13-14; Paper 44 (Sur-Reply) at 31

It Would Not Have Been Obvious to Combine Steffens and Zhao

Steffens' wavelet-based approach is incompatible with Zhao's contour-based approach

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

EX2004 (Saber Decl.) ¶129

harder, because these are hidden by muscle, skin, and clothing. Segmentation or contour extraction errors also pose a significant challenge to accurate joint localization. Although

EX1006 (Zhao) at 20

The context for human detection is the general object detection problem. There are two approaches to object detection. One approach is to search the whole image at multi-scales for objects. This is a time consuming procedure and may result in multiple responses from a single object. Another approach is to first segment foreground objects from the background, then classify each segmented object as human or non-human. In this dissertation, I employ the second approach to human detection. Classifying only segmented objects rather than

EX1006 (Zhao) at 20

detection system is faster and more accurate than MIT's system. The advantage of the NN-based and wavelet-based methods is that no contours are required to extract from the image first. The disadvantage of the NN-based method is that it may generate many false alarms

EX1006 (Zhao) at 88

Paper 26 (POR) at 42; Paper 44 (Sur-Reply) at 31



It Would Not Have Been Obvious to Combine Steffens and Zhao

Zhao teaches away from Steffens' wavelet-based approach

130. Furthermore, Zhao teaches away from combining Steffens' wavelet-based bunch graph approach with Zhao. Zhao observes that a wavelet-based approach "has to search the whole image at multi-scales for humans" and thus "would be an extremely computationally expensive procedure, and it may cause multiple responses from a single human." EX1006 at 82-83. Zhao further observes that another wavelet-based approach "has to search for pedestrians in different scales in the whole image, so it is very slow." EX1006 at 88. Like the wavelet-based approaches disparaged by Zhao, in Steffens' approach, "[t]he jets are composed of wavelet transforms," and "[i]n order to find a face in an image frame, the graph is moved and scaled over the image frame until a place is located at which the graph matches best (the best fitting jets within the bunch jets are most similar to jets extracted from the image at the current positions of the nodes). EX1007 at 3:42-44, 9:37-41. Thus, a POSITA would not have been motivated to combine Zhao and Steffens and would have been led away from the combination.

EX2004 (Saber Decl.) ¶130

mans walking together as a vehicle. Papageorgious and Poggio [37] present a more robust human detection system based on wavelet analysis and the support vector machine (SVM) technique. However, the system has to search the whole image at multi-scales for humans. This would be an extremely computationally expensive procedure, and it may cause multiple responses from a single human. More recently, Gavrilu and Philomin [69] developed a

EX1006 (Zhao) at 82-83

correctly as shown in Fig. 5.3 using the RCR algorithm. The disadvantage of the wavelet-based method is that it has to search for pedestrians in different scales in the whole image, so it is very slow. Another limitation with these two approaches is that they are global shape based and cannot deal with partial occlusion very well. However, for the application

EX1006 (Zhao) at 88

Paper 26 (POR) at 42-43; Paper 44 (Sur-Reply) at 31



It Would Not Have Been Obvious to Combine Steffens and Zhao

Steffens' deterministic (non-probabilistic) approach is incompatible with Zhao's probabilistic approach

incompatible with Zhao's. The key error in Dr. Bajaj's analysis is that he incorrectly characterizes "the normalized scalar product of the amplitude vector" of two jets in Steffens as a probability. EX1052, ¶¶52-53. As the POSITA would know from basic linear algebra, however, normalizing the absolutes of two vectors does not yield a probability. See EX1007 at 8:56-67.

EX2007 (Saber Supp. Decl.) ¶48

50. As the POSITA would understand, the vectors that Steffens normalizes—which are filter responses of a Gabor wavelet—are not probabilities, and the result of normalizing them does not make them probabilities. See EX1007 at 8:14-67. Moreover, Steffens includes no discussion of a probabilistic similarity measure or any mention of probability. The word "probability" does not even appear in Steffens. Steffens' approach is clearly deterministic and has no probabilistic modeling whatsoever. Thus, Steffens does not simply use different terminology than Zhao or the '980 patent as Dr. Bajaj suggests—it employs a fundamentally different mathematical framework for a similarity measure that is incompatible with Zhao's formulation framework.

EX2007 (Saber Supp. Decl.) ¶50

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

EX1007 (Steffens) at 3:42-44

The wavelet is a plane wave with wave vector $\vec{k}$ restricted by a Gaussian window, the size of which relative to the wavelength is parameterized by σ . The term in the brace removes the DC component. The amplitude of the wavevector k may be chosen as follows where v is related to the desired spacial resolutions.

EX1007 (Steffens) at 8:14-67

A labeled image graph 162, as shown in FIG. 11, is used to describe the aspects of an object (in this context, a face). The nodes 164 of the labeled graph refer to points on the object and are labeled by jets 160. Edges 166 of the graph are labeled with distance vectors between the nodes. Nodes and edges define the graph topology. Graphs with equal geometry may be compared. The normalized dot product of the absolute components of two jets defines the jet similarity. This value is independent of the illumination and contrast changes. To compute the similarity between two graphs, the sum is taken over similarities of corresponding jets between the graphs.

EX1007 (Steffens) at 8:14-67

Paper 44 (Sur-Reply) at 31 (citing EX2007 ¶¶48-50)

Steffens + Zhao Does Not Disclose or Render Obvious Element 1/9[a] or 1/9[e]

Term	Patent Owner's Construction
“detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object”	detecting the object in any image <u>based on comparing the probabilities for each image of the plurality of images</u> to a threshold for detecting the object
“spatio-temporal model”	spatial model where an object is detected in an image <u>based on the probabilities that the object is in preceding and/or subsequent images</u>

Steffens + Zhao fails for the same reasons as Zhao

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

Paper 3 (Petition) at 45

Paper 26 (POR) at 45-46; Paper 44 (Sur-Reply) at 32



Steffens + Zhao Does Not Disclose or Render Obvious Claim 4/12

4. The computer implemented method of claim 1, wherein determining the probability that the object is in each image comprises detecting the object in a current image according to measurements of the object as a collection of components determined from a prior image and a later image relative to the current image.

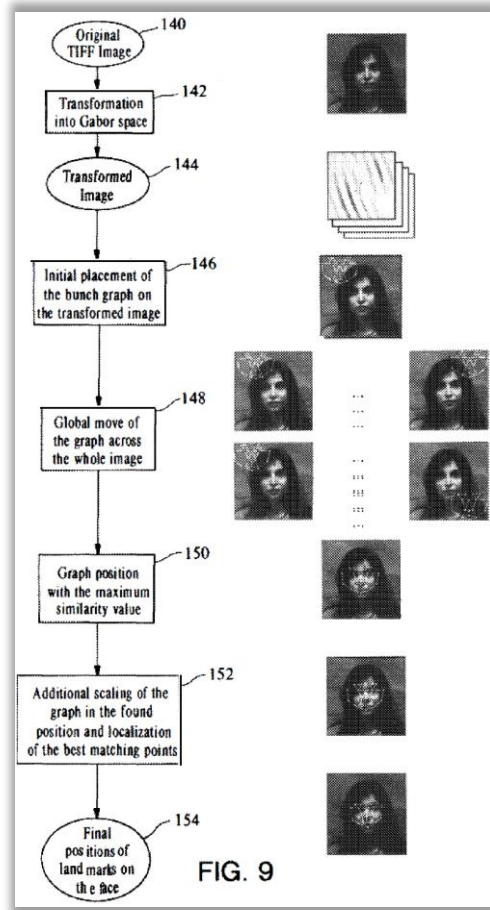
12. The method of claim 9, wherein determining the probability that the object is in each image comprises detecting the object in a current image according to measurements of the object as a collection of components determined from a prior image and a later image relative to the current image.

EX1001 ('980 patent) at claims 4, 12



Steffens + Zhao Does Not Disclose or Render Obvious Claim 4/12

Steffens matches an image graph of a single (current) image to a model graph



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

EX1007 (Steffens) at 7:51-67

EX1007 (Steffens) Fig. 9

Paper 26 (POR) at 47-48

Steffens + Zhao Does Not Disclose or Render Obvious Claim 4/12

Steffens tracks nodes from an earlier image to a later image (forward direction only)

139. Additionally, I disagree with Dr. Bajaj that “between frames” with respect to tracking in Steffens discloses detecting an object in a current image according to measurements of the object from later images or renders it obvious. Steffens confirms to a POSITA that Steffens tracking technique operates in only the forward direction. Steffens states that “facial features may be tracked over consecutive frames,” but makes clear that tracking is performed from earlier nodes to later nodes, *i.e.*, in the forward direction from I_{n-1} to I_n and then from I_n to I_{n+1} . EX1007 at 10:46-61. Steffens does not disclose or suggests tracking backwards in time or detecting an object in a current frame according to graph information for the object from a later frame.

EX2004 (Saber Decl.) ¶139

After the facial features and landmarks are located, the facial features may be tracked over consecutive frames as illustrated in FIGS. 17 and 18. The tracking technique of the invention achieves robust tracking over long frame sequences by using a tracking correction scheme that detects whether tracking of a feature or node has been lost and reinitializes the tracking process for that node.

The position X_n of a single node in an image I_n of an image sequence is known either by landmark finding on image I_n using the landmark finding method (block 180) described above, or by tracking the node from image $I_{(n-1)}$ to I_n using the tracking process. The node is then tracked (block 182) to a corresponding position $X_{(n+1)}$ in the image $I_{(n+1)}$ by one of several techniques. The tracking methods described below advantageously accommodate fast motion.

EX1007 (Steffens) at 10:46-61

Paper 26 (POR) at 48



Steffens + Zhao Does Not Disclose or Render Obvious Claim 4/12

Steffens tracks nodes from an earlier image to a later image (forward direction only)

However, linear motion prediction necessarily infers object detection using multiple frames. Ex1007, 10:62-11:12 (discussing predicting position $X_{(n+1)}$ from position X_n). A POSITA would have known the benefits of extrapolating motion information from prior and later frames. Ex1052, ¶¶36-39. Accordingly, a POSITA reading Steffens would know to extrapolate from a linear motion prediction, which necessarily involves looking at a prior frame and a later frame relative to a current frame.

Paper 32 (Reply) at 17-18

Section VII.B.1 above, do not argue such inherency. I also see no support from Petitioners to support their assertion that linear motion prediction “necessarily infers object detection using multiple frames” is also unsupported. The POSITA would understand that an object need not be detected using measurements from prior and later frames just because the object is tracked across frames. For instance, Steffens is an example of one-direction tracking.

EX2007 (Saber Supp. Decl.) ¶53

Paper 44 (Sur-Reply) at 32



Steffens + Zhao Does Not Disclose or Render Obvious Claim 8/16

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

$$P(X_0^O, X_0^{C0}, X_0^{C1}, \dots, X_0^{CN}, \dots, X_T^O, X_T^{C0}, X_T^{C1}, \dots, X_T^{CN}) = \frac{1}{Z} \prod_{ij} \psi_{ij}(X_i^O, X_j^O) \prod_{ik} \psi_{ik}(X_i^O, X_i^{Ck}) \prod_{ikl} \psi_{kl}(X_i^{Ck}, X_i^{Cl}) \prod_i \phi_i(X_i^O) \prod_{ik} \phi_i(X_i^{Ck}).$$

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

$$P(X_0^O, X_0^{C0}, X_0^{C1}, \dots, X_0^{CN}, \dots, X_T^O, X_T^{C0}, X_T^{C1}, \dots, X_T^{CN}) = \frac{1}{Z} \prod_{ij} \psi_{ij}(X_i^O, X_j^O) \prod_{ik} \psi_{ik}(X_i^O, X_i^{Ck}) \prod_{ikl} \psi_{kl}(X_i^{Ck}, X_i^{Cl}) \prod_i \phi_i(X_i^O) \prod_{ik} \phi_i(X_i^{Ck}).$$

EX1001 ('980 patent) at claims 8, 16

Steffens + Zhao Does Not Disclose or Render Obvious Claim 8/16

Steffens + Zhao fails for the same reasons as Zhao

9. Claims 8/16

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

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

Paper 3 (Petition) at 52

Paper 26 (POR) at 49; Paper 44 (Sur-Reply) at 32



Ground 3: Obviousness over Zhao + Ozer

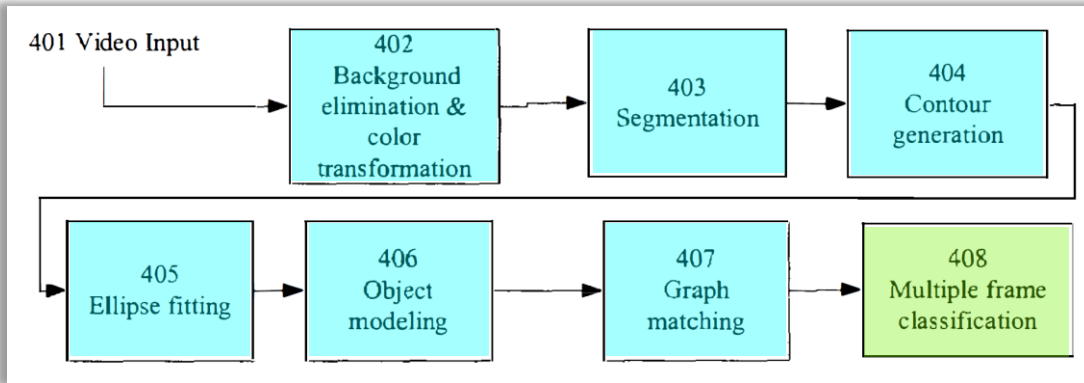
Ground	Assertion	Petitioned Claims
1	Obvious under §103 over Zhao	1-16
2	Obvious under §103 over Steffens in view of Zhao	1-16
3	Obvious under §103 over Ozer in view of Zhao	1-16
4	Obvious under §103 over TLP	1-16

Paper 3 (Petition) at 6



Ozer: Overview

Ozer teaches single-frame object part detection



EX1008 (Ozer) Fig. 4

The system of the present invention can detect non-rigid (e.g. human body) and rigid object parts and recognize their activities in compressed and uncompressed domains. To achieve this, a method with two levels, namely low and high levels, is used. The low-level part performs object detection and extracts parameters for the abstract graph representation of the image being processed. The high level part uses dynamic programming to determine the activities of the object parts, and uses a distance classifier to detect specific activities.

EX1008 (Ozer) at 3:55-64

Early Stage Analysis

Background elimination and color transformation: The first step (402) is the transformation of pixels into another color space regarding to the application. Background elimi-

Segmentation: In 403, the foreground regions are extracted and the object of interest is segmented hierarchically into its smaller unique parts based on the combination of color components and statistical shape features after background elimination. The meaningful adjacent segments

Contour following: Contour points of the segmented regions are extracted and stored (404). FIG. 5 gives an

Ellipse fitting: This step (405) fits ellipses to the contours.

Object modeling by invariant shape attributes: For object detection, it is necessary to select part attributes which are invariant to two-dimensional transformations and are maximally discriminating between objects (406).

Graph matching: In this step (407), we compare the object model with a set of stored models. Each extracted region

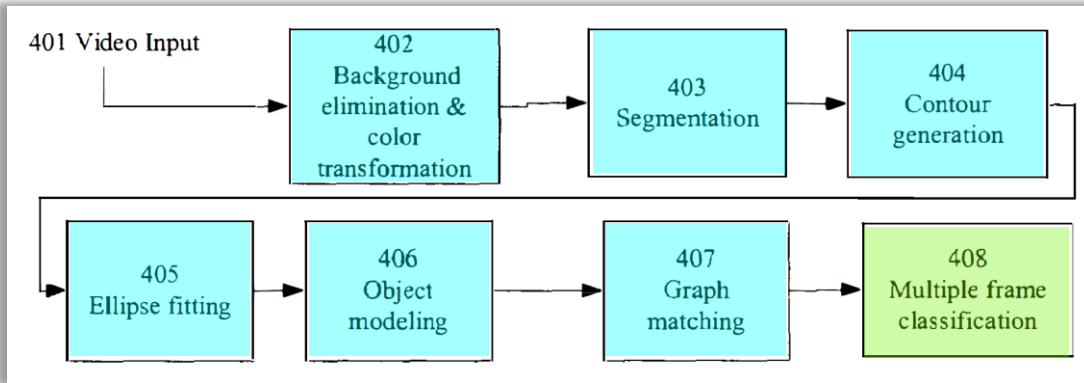
Output of the graph-matching algorithm is the classified object parts. The movements of the object parts are

EX1008 (Ozer) at 6:42-7:35

Paper 26 (POR) at 15-16, 50; Paper 44 (Sur-Reply) at 32

Ozer: Overview

Ozer teaches recognizing the “activity” of object parts



EX1008 (Ozer) Fig. 4

The system of the present invention can detect non-rigid (e.g. human body) and rigid object parts and recognize their activities in compressed and uncompressed domains. To achieve this, a method with two levels, namely low and high levels, is used. The low-level part performs object detection and extracts parameters for the abstract graph representation of the image being processed. The high level part uses dynamic programming to determine the activities of the object parts, and uses a distance classifier to detect specific activities.

EX1008 (Ozer) at 3:55-64

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

EX1008 (Ozer) at 7:36-:42

Ozer + Zhao Does Not Disclose or Render Obvious Element 1/9[a] or 1/9[e]

Term	Patent Owner's Construction
“detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object”	detecting the object in any image <u>based on comparing the probabilities for each image of the plurality of images</u> to a threshold for detecting the object
“spatio-temporal model”	spatial model where an object is detected in an image <u>based on the probabilities that the object is in preceding and/or subsequent images</u>

Ozer + Zhao fails for the same reasons as Zhao

f. Element 1/9[e]

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

Paper 3 (Petition) at 60-61

Paper 26 (POR) at 50; Paper 44 (Sur-Reply) at 32



Ozer + Zhao Does Not Disclose or Render Obvious Claim 4/12

4. The computer implemented method of claim 1, wherein determining the probability that the object is in each image comprises detecting the object in a current image according to measurements of the object as a collection of components determined from a prior image and a later image relative to the current image.

12. The method of claim 9, wherein determining the probability that the object is in each image comprises detecting the object in a current image according to measurements of the object as a collection of components determined from a prior image and a later image relative to the current image.

EX1001 ('980 patent) at claims 4, 12



Ozer + Zhao Does Not Disclose or Render Obvious Claim 4/12

Ozer's object part detection is based on a single image

The video frames are sent from the camera to the video capturer on the PC. The video capturer converts the frames sent from the camera from analog format to digital format and stores the color components of the current frame in three different buffers.

The color components stored in the buffers are further processed by the central processor to classify the object regions. Region parameters are compared with the object parameters in the object database and the region is classified.

EX1008 (Ozer) at 5:10-20

Low-level part performs object detection and extracts parameters for the abstract graph representation of the frame being processed in real time. Local consistency based on low

EX1008 (Ozer) at 3:65-67

tion. However, since motion detection and tracking don't know what is moving, they can easily generate false alarms as well as missing important events. In contrast, this system builds a model of the object of interest on each video frame.

EX1008 (Ozer) at 6:19-24

each video frame.” EX1008 at 6:19-24. These teachings of Ozer confirm to a POSITA that Ozer's detection processes uses data extracted from only a current frame and does not disclose or suggest detecting an object in the current frame based on parameters extracted from prior or later frames. Ozer's approach is different from the invention of '980 patent, as Ozer's graph matching technique performs matching based on data for a single “current” frame—it does not detect an object in a given image based on probabilities of the object being in other images.

EX2004 (Saber Decl.) ¶149

Paper 26 (POR) at 51-52



Ozer + Zhao Does Not Disclose or Render Obvious Claim 4/12

Ozer's activity classification is not object detection

The system of the present invention can detect non-rigid (e.g. human body) and rigid object parts and recognize their activities in compressed and uncompressed domains. To achieve this, a method with two levels, namely low and high levels, is used. The low-level part performs object detection and extracts parameters for the abstract graph representation of the image being processed. The high level part uses dynamic programming to determine the activities of the object parts, and uses a distance classifier to detect specific activities.

EX1008 (Ozer) at 3:55-64

model via graph matching. The high level part uses a pattern classifier, namely Hidden Markov Models, which classifies activity patterns of the body parts in space-time and determines the movements of the object parts. It also uses a distance classifier to detect specific gestures and activities. For articulated objects like human body, the system will find the activities of each body part independently from each other and combine the activities to find the gesture and overall activity of the human body at a later stage by using a quadratic distance classifier.

EX1008 (Ozer) at 6:19-24

model via graph matching. The high level part uses a pattern classifier, namely Hidden Markov Models, which classifies activity patterns of the body parts in space-time and determines the movements of the object parts. It also uses a

EX1008 (Ozer) at 44:44-53

After the detection of the object parts, the system is ready to recognize the activities of each object part and the overall activity of the object.

EX1008 (Ozer) at 4:35-37

EX1008 at 4:35-37. From this a POSITA would understand that Ozer's object classification step is for detecting an object, see EX2006 at 41:8-20, 43:7-44:7, whereas Ozer's subsequent activity tracking step for classifying activities is not part of Ozer's object detection process. Indeed, the "[o]utput" of detection is produced before activity classification even begins, which confirms that the activity classifier does not result in basing object detection in a current frame on data extracted from prior or later frames. EX1008 at 7:34-45.

EX2004 (Saber Decl.) ¶150

Paper 26 (POR) at 52-53



Ozer + Zhao Does Not Disclose or Render Obvious Claim 4/12

Ozer's activity classification is not object detection

Dr. Bajaj admits that activity recognition happens “after” the detection process and that Ozer tracks “only objects that fit the model,” *i.e.*, the previously detected objects. EX1052, ¶63. The fact that motion classification over multiple images uses the results of object detection in those images does not mean that object detection spans multiple images. And in Ozer, it does not. As I previously explained, Ozer's object detection is based on information collected from a single frame, and the detection is made using that information—not information from a previous and a later frame. EX2004, ¶149; EX1008 at 3:65-67, 5:10-20, 6:19-24, 6:52-58.

EX2007 (Saber Supp. Decl.) ¶58

Ground 4: Obviousness over TLP

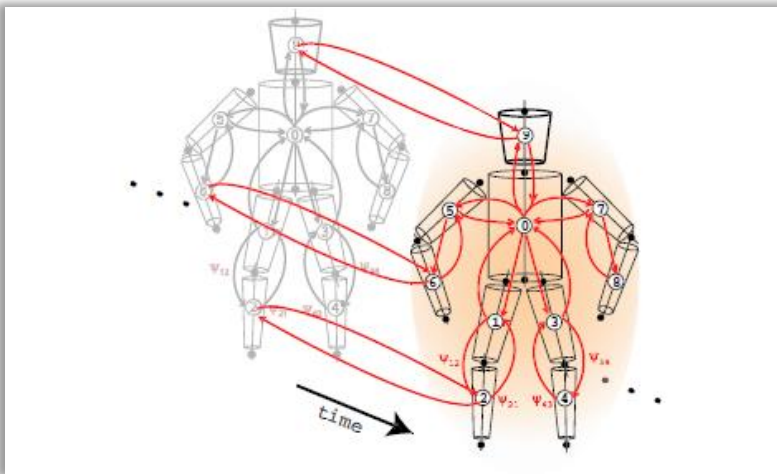
Ground	Assertion	Petitioned Claims
1	Obvious under §103 over Zhao	1-16
2	Obvious under §103 over Steffens in view of Zhao	1-16
3	Obvious under §103 over Ozer in view of Zhao	1-16
4	Obvious under §103 over TLP	1-16

TLP: Overview

TLP models components, *i.e.*, “limbs” or “parts”

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 repre-



TLP Does Not Disclose or Render Obvious Element 1/9[d]

Petitioners' object/component "equivalence" theory is incorrect

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

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

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

Paper 3 (Petition) at 70-71

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 repre-

EX1009 (TLP) at 1

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.

EX1009 (TLP) at 2

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

Paper 26 (POR) at 58-60

TLP Does Not Disclose or Render Obvious Element 1/9[d]

Petitioners' object/component "equivalence" theory is incorrect

The inference algorithm outlined in the next section combines the body model described above with a probabilistic image likelihood model. We define $\phi_i(X_i)$ to be the likelihood of observing the image measurements conditioned on the pose of limb i. Ideally this model would be robust to partial occlusions, the variability of image statistics across different input sequences, and variability among subjects.

EX1009 (TLP) at 4

Separate foreground models are learned for the appearance of each limb. In addition, a pooled background model is learned from non-limb patches sampled from the training images. These are combined into a limb likelihood by taking the likelihood ratio [19]. These likelihood ratios along with background subtraction information are then combined across views, assuming independence of the views conditioned on the limb position.

EX1009 (TLP) at 4

nodes in space and time." EX1009 at 1. Accordingly, when TLP states that "each node has an associated likelihood defined over a rich set of image cues using a learned Gibbs model," a POSITA would have understood TLP as referring to each limb or part (not the human) having an associated likelihood. EX1009 at 1-2. This would have been clear to a POSITA, as TLP states that "[w]e define $\phi_i(X_i)$ to be the likelihood of observing the image measurements conditioned on the pose of **limb i**," or a "**limb** likelihood." EX1009 at 4 (emphasis added).

EX2004 (Saber Decl.) ¶156

Paper 26 (POR) at 58-60



TLP Does Not Disclose or Render Obvious Element 1/9[d]

Petitioners' new "representative" component theory is untimely and improper

A POSITA would have understood to select a component of the object (e.g., the torso) to be representative of the object node x1 (i.e., a 1st layer), and to conditionally link all other components of the body (head, limbs, etc.) forming a 2nd layer of xi nodes. See PAMPAS, EX1019, § 3.3 (finding a jointed object in clutter). Due to this two-layered model, even when the object node x1 is occluded, after one iteration the belief at that node has high weight at the correct location due to the agreement of the nearby arm nodes xi. After a few iterations the belief has converged on the correct object position despite the lack of the central circle. *Id.* Thus, TLP teaches "determining a probability that the object is in each image." Ex1052, ¶¶6-23, 24-34, 65-85.

Paper 32 (Reply) at 25-26

Examples of indications that a new issue has been raised in a reply include new evidence necessary to make out a *prima facie* case for the patentability or unpatentability of an original or proposed substitute claim, such as newly raised rationale to combine the prior art references that was not expressed in the petition. See *Intelligent Bio-Sys., Inc. v. Illumina Cambridge Ltd.*, 821 F.3d 1359, 1369–70 (Fed. Cir. 2016) (holding that the Board did not err in refusing the reply brief as improper under 37 C.F.R. § 42.23(b) because petitioner relied on an entirely new rationale to explain why one of skill in the art would have combined the references at issue). It is also improper for a reply to present new evidence (including new

Consolidated Trial Practice Guide (Nov. 2019) at 74



Paper 44 (Sur-Reply) at 34-35

TLP Does Not Disclose or Render Obvious Element 1/9[d]

Petitioners' new "representative" component theory is incorrect

70. Dr. Bajaj's new argument that the POSITA would have allegedly understood from PAMPAS (EX1019) to select a component in TLP to represent the object (as a second layer of nodes) is incorrect. His argument is based on his incorrect construction of the claims that does not distinguish the "object" in the claims from the "components." As the POSITA would understand, treating a component as arbitrarily "representative" of the object does not make that component's probability a probability of the object. Moreover, PAMPAS does not teach treating a component as the object—it analyzes only components (the "central circle," "arms," and "links" are components) and estimates belief for an occluded component (i.e., the central circle node) based on "agreement of the nearby arm nodes" using belief propagation. EX1019, 6-8. It does not, however, select a component to be representative of an object node, as Dr. Bajaj incorrectly suggests.

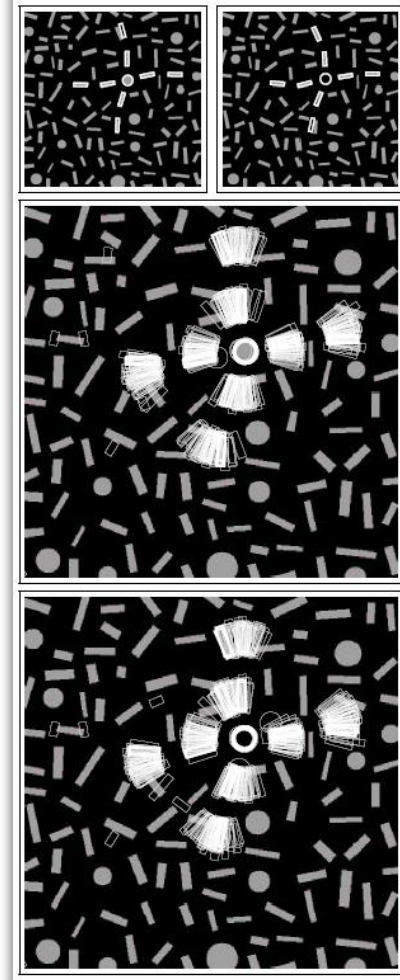
EX2007 (Saber Supp. Decl.) ¶70

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 $x_1 = (x_0, y_0, r_0)$ encodes a position and radius. Each arm node $x_i = (x_i, y_i, \alpha_i, w_i, h_i), 2 \leq i \leq 9$ 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-

EX1019 ("PAMPAS") at 6

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 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.

EX1019 ("PAMPAS") at 7



Paper 44 (Sur-Reply) at 34-35

TLP Does Not Disclose or Render Obvious Element 1/9[a] or 1/9[e]

Term	Patent Owner's Construction
“detecting the object in any image upon comparing the probabilities for each image to a threshold for detecting the object ”	detecting the object in any image <u>based on comparing the probabilities for each image of the plurality of images</u> to a threshold for detecting the object
“spatio-temporal model”	spatial model where an object is detected in an image <u>based on the probabilities that the object is in preceding and/or subsequent images</u>

“object” ≠ “component”

Paper 26 (POR) at 62-64; Paper 44 (Sur-Reply) at 34-35



TLP Does Not Disclose or Render Obvious Claim 3/11

“object” ≠ “component”

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.

11. The method of claim 9, wherein the spatio-temporal model is a graphical model comprising nodes corresponding to each of the collection of components and to the object.

EX1001 ('980 patent) at claims 4, 12

Paper 26 (POR) at 64-65; Paper 44 (Sur-Reply) at 35



TLP Does Not Disclose or Render Obvious Claim 8/16

“object” ≠ “component”

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

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

EX1001 ('980 patent) at claim 8

8). The joint probability distribution for this spatio-temporal graphical object model with N components can be written as:

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

where X_t^O and $X_t^{C_n}$ is the state of the object O, and object's n-th component, C_n , at time t respectively ($n \in (1, N)$ and $t \in (1, T)$); $\psi_{ik}(X_i^O, X_j^O)$ is the temporal compatibility of object state between frames i and j; $\psi_{ik}(X_i^O, X_i^{C_k})$ is the spatial compatibility of the object and it's components at frame i; $\psi_{ik}(X_i^{C_k}, X_i^{C_l})$ is the spatial compatibility between object components at frame i and $\phi_i(X_i^O)$ and $\phi_i(X_i^{C_k})$ denote the local evidence for the object and component states respectively.

EX1001 ('980 Patent) 3:47-4:7; Paper 44 (Sur-Reply) at 6-7

TLP Does Not Disclose or Render Obvious Claim 8/16

Petitioners and Dr. Bajaj concede that claim 8/16 includes “object” and “component” terms

As described for Claim 1/9, TLP presents a probabilistic method for object detection. *See* elements 1/9[d]-[e]. The joint probability distribution formulated in Claims 8/16 would have been known to a POSITA and is the very same pairwise Markov field representation of a spatio-temporal graphical model, which is cited in TLP. *See* EX1009, 4 (“we use a form of non-parametric belief propagation [13 [citing EX1019], 25 [citing EX1012]].”); *see also* EX1004, ¶131 (“The equation represents a typical factorized form of the joint probability distribution $P(\dots)$ for a graphical model G of an object O with state X_i^O at time frame i of the video and $\{k = 1..N\}$ components state X_i^{C-k} ”). This form of the joint probability would have been known to a POSITA as a pairwise Markov field representation of a spatio-temporal graphical model. *Id.* (citing EX1012, §2.1).

Paper 3 (Petition) at 75-76

82. As I explained in my declaration (¶231), the equation recited in Claims 8/16 represent a typical factorized form of the joint probability distribution $P(\dots)$ for a graphical model G of an object O with state X_i^O at time frame i of the video and $\{k = 1..N\}$ components state X_i^{C-k} . This form of the joint probability

EX1052 (Bajaj Supp. Decl.) ¶82

Paper 44 (Sur-Reply) at 36



TLP Does Not Disclose or Render Obvious Claim 8/16

Petitioners' theory is based on their incorrect construction

7 Q. Okay. And so since we're talking
8 about a potential function in the superscript
9 "O," we're just talking about the object, not the
10 ones that are superscript "C," right?

11 A. The object is the sum of its parts.

EX2008 (Bajaj Dep. Tr.) at 90:7-11 (cited in EX2007 ¶77)

19 Q. So the difference between the object
20 state XO and the component state XC, those are
21 different states, right?

22 A. Not at any point in time. The
1 object is the collection of components.

EX2008 (Bajaj Dep. Tr.) at 92:19-93:1 (cited in EX2007 ¶77)

Paper 26 (POR) at 65; Paper 44 (Sur-Reply) at 36 (citing EX2007, ¶77)

