A Survey of Image Processing Techniques for Identification and tracking of objects from a video sequence

Farhana Wani[#1], Adeel ahmed khan[#2], Dr. S Basavaraj Patil[#3]

[#]Dept. of Computer Science and Engineering
BTL Institute of Technology
Bangalore, India

Abstract— Visual surveillance is an emerging research topic in image processing. This paper discusses about various image processing techniques and tools which are available for identification and tracking of moving objects in a crowd. In general, the processing framework of visual surveillance task includes the following stages: detection of moving objects, their classification, tracking, and identification of the behavior. In this paper, we provide a brief overview of the techniques for object detection and tracking as well as some insights to the behavior of the crowd. Object detection and tracking algorithms can be proactively used to respond to accidents, crime, suspicious activities, terrorism, and may provide insights to improve evacuation planning and real-time situation awareness during public disturbances. For the visual surveillance task, the virtual analyst has to work on the information provided by the detection and tracking segment of the system and based on that, it has to discover some interesting patterns within a crowd.

Keywords— Visual surveillance, motion detection, classification, tracking, behavior recognition, Hidden Markov Model

I. INTRODUCTION

In the visual surveillance task there has been increasing interest in finding methods for detecting moving objects in a crowd along with extracting their trajectories. Using this data we can derive useful information about a crowd and analyze the activities of various entities that are present in a crowd. In order to fight crime or monitor for abnormal events, there has been an increase in the installation of video cameras in public areas to keep track of various activities that are happening within a crowd. Visual surveillance in dynamic scenes has a wide range of potential applications, such as a security guard for communities and important buildings, traffic surveillance in cities and detection of military targets, etc. Detection of various activities in a crowd has been always a challenging task mainly because of the limitation imposed by large amounts of video data that is available, hence leaving the task of analysis of behavior of the crowd to human operators. Manual analysis of video happens to be costly, labor intensive, and prone to errors. Moreover, there are several limitations in the manual analysis mainly due to scarce human resources and inability of humans to monitor simultaneous signals [1]. Thus, in order to overcome the shortcomings of the human surveillance capabilities we require a virtual analyst for the recognition of various activities in a crowd. Less well studied is the collective behavior of small groups of people in a crowd. In this paper, we discuss various detection and tracking techniques to discover and extract the trajectories of moving objects which can be cars or people. Hence the low-level processing techniques used are common in almost all the visual surveillance tasks which are as follows: detection of moving objects, classification, tracking, and identification of behavior. Evaluating the group structure of crowds has significant real-world applications. Present models of evacuation consider all people as separate agents making independent decisions. These “particle flow” models tend to underestimate the time it takes for people to leave an area because groups of individuals who are together try to leave together, limiting the speed of the group to that of its slowest member. There are various group behavior models which improve the strategies for police intervention during public disturbances. Rather than considering an irrational homogeneous crowd, police should be looking at small groups, only a few of which might merit to disturbing elements.

Much of the work has been done in the computer vision system for the automatic detection and tracking of moving objects, less well studied is the analysis of behavior of objects which can be cars or pedestrians that comprise the crowd. However the task becomes much more complex particularly when the crowd involves interactions between people. In the visual surveillance task Hidden Markov Models (HMMs) have been extensively used in recent years for modelling and identifying activities of the crowd that are captured in the video. For showing less structured group or interactive activities involving multiple temporal processes, Dynamic Probabilistic Networks (DPNs) can be used[2,3]. One way to construct a DPN is to extend a standard HMM to a set of interconnected multiple HMMs. A Multi-Observation

Mixture Counter Hidden Markov Model (MOMC-HMM) was introduced by Brand and Kettnaker[4] to represent multiple observations of different objects at each state. Vogler and Metaxas [5] proposed Parallel Hidden Markov Models (PaHMMs) that factorize state space into multiple independent temporal processes without causal connections in between. Any interconnection among temporal processes is implicitly assumed to be by strict zero-order synchronization, i.e. simultaneousness. This is generally untrue. Brand and Oliver *et al.* [6,7] exploited Coupled Hidden Markov Models (CHMMs) to take into account the causal connections among multiple temporal processes. They are essentially fully coupled pairs of HMMs such that each state is conditionally dependent on all past states of all processes at the previous time instance. However, it can be shown that such a fully connected statespace cannot be factorized effectively therefore leading to poor network topology [8].

A dynamically Multi-Linked Hidden Markov Model (DML-HMM) introduced by Shaogang Gong and Tao Xiang[65] for the recognition of group activities involving multiple different object events in a noisy outdoor scene, with its topology being intrinsically determined by the underlying causality and temporal order discovered automatically using Schwarz’s Bayesian Information Criterion based factorization. A relational clustering approach has also been formulated by Anthony Hoogs Steve Bush Glen Brooksby[9] to address the problem of recognizing group activities, or activities with an arbitrary, variable number of participants.

II. Object Recognition

A visual surveillance task starts with detection of motion. Motion detection segments regions corresponding to moving objects from the rest of an image. All other subsequent processes like tracking and behavior recognition in the crowd are greatly dependent on it. There are two different detection strategies. The videos that are captured from high elevation where people appear small, individual pedestrians are detected by using Reversible Jump Markov Chain Monte Carlo(RJMCMC) to find a set of overlapping rectangles that best explain or “cover” the foreground pixels in a binary segmentation generated by adaptive background subtraction. The method being similar to that of [10], [11], [12] and is capable of extracting overlapping individuals in crowds up to moderate density. For higher resolution videos, pedestrian detection is attained in each frame using a combination of motion and contour (edge gradient) information, using a set of templates learned offline from training examples extracted from the same camera viewpoint. An HoG detector, implemented from the description in Dalal and Triggs [13] can be used for the same. Based on [14], motion information can be used to determine regions that are more likely to contain moving pedestrians, in the form of a background subtraction mask. Background subtraction, also known as Foreground Detection, is a method in the fields of image processing and computer vision where an image's foreground is extracted for further processing (object recognition etc.). Generally an image's regions of interest are objects (humans, cars, text etc.) in its foreground. Visual surveillance systems for stationary cameras usualy include some sort of motion detection. Motion detection is used to segment moving objects from the rest of the image i.e. the background. Motion detection process can be split into environment modeling, motion segmentation, and object classification. There are several approaches for motion segmentation, some of which are explained as follows:

A. Background subtraction

A very common object segmentation approach is background subtraction. Background subtraction compares an image with an estimate of the image as if it contained no foreign objects. It extracts foreground objects from sections where there is a substantial change between the observed and the estimated image. Common algorithms contain methods by Heikkila and Silven[15], Stauffer and Grimson (adaptive Gaussian mixture modelor GMM) [16], Halevy and Weinshall [17], Cutler and Davis[18], and Toyama *et al.* (Wallflower) [19]. A thorough general survey of image change algorithms can be seen in [20].

B. Temporal differencing

Temporal differencing performs the pixel-wise differences between two or three successive frames in an image sequence to extract moving objects. Temporal differencing is typically computationally economical and is very adaptive to dynamic environments, but it regularly fails at properly extracting the shape of the object in motion and can cause small holes to appear inside moving entities. For getting better results, hybrid approaches can be used which often combine both background subtraction and temporal differencing methods to obtain more robust segmentation approaches.

C. Optical flow

Optical flow is a vector-based method that estimates motion in video by matching points on objects over multiple frames. A reasonably high frame rate is essential for accurate measurements.

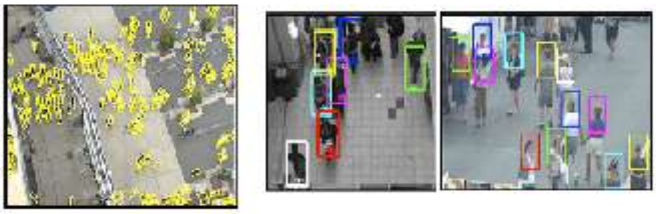

Fig. 1. Left: Sample detections in low-resolution

video using RJMCMC. Right: Sample detections in higher resolution video using an HoG detector for body (left) and head-and-shoulders (right).

After extracting moving regions or objects in an image, the next step in the behavior-recognition process is object classification. In general, for object classification in surveillance video, there are shape-based, motion-based, and feature-based classification methods:

A. *Shape-based classification*

Based on the geometry of the extracted regions (boxes, silhouettes, blobs) containing motion, we can classify objects in video surveillance. VASM [21]takes image blob dispersedness, image blob area, apparent aspect ratio of the blob bounding box, etc, as key features, and classifies moving-object blobs into four classes: single human, vehicles, human groups, and clutter, using a viewpoint-specific three-layer neural network classifier. Lipton *et al.* [22] use the dispersedness and area of image blobs as classification metrics to classify all moving-object blobs into humans, vehicles and clutter. Temporal consistency constraints are considered so as to make classification outcomes more accurate. Kuno*et al.*[23] use simple shape parameters of human silhouette patterns to separate humans from other moving objects.

B. *Motion-based classification*

This classification method is based on the notion that object motion features and patterns are unique enough to distinguish between objects. Cutler *et al.*[24]describes a similarity-based procedure to identify and examine periodic motion. Humans can have distinct types of motion which in turn can be used to identify "types" of human movements such as walking, running, or skipping, as well as for human identification. By tracing a moving object, its self-similarity is calculated as it progresses over time. For periodic motion, its self-similarity measure is also periodic. Therefore time-frequency analysis is applied to detect and characterize the periodic motion, and tracking and classification of moving objects are realized using periodicity. In Lipton's work [25], residual flow is used to examine rigidity and periodicity of moving objects. It is probable that rigid objects present little residual flow, whereas a non-rigid moving object such as a human being has a higher average residual flow and even exhibit a periodic component. Based on this concept, human motion is distinguished from motion of other objects, such as vehicles.

III. OBJECT TRACKING

After the objects have been detected, the next task for the visual surveillance system is to track the objects from one frame to another consecutively following the sequence. The tracking algorithms can be used in parallel with the object detection during the initial processing.

Tracking over time typically includes matching objects in successive frames using features such as points, lines or blobs. There are numerous mathematical tools available for tracking which include the Kalman filter, the Condensation algorithm, the dynamic Bayesian network, the geodesic method, etc. Tracking procedures are divided into four main types: region-based tracking, active-contour-based tracking, feature-based tracking, and model-based tracking. These categories can also be combined together and used as a hybrid algorithm.

A. *Region-Based Tracking*

Region-based tracking algorithms track objects according to deviations of the image sections corresponding to the moving objects. For such algorithms, the background image is retained dynamically [26], [27], and motion regions are typically identified by subtracting the background from the current image. Wren *et al.* [28] explore the use of small blob features to track a single human in an indoor location. In their work, a human body is considered as a combination of some blobs respectively representing various body parts such as head, torso and the four limbs. Gaussian distributions of pixel values can be used to model both human body and background scene. Lastly, the pixels belonging to the human body are allocated to the different body part's blobs using the log-likelihood measure. Therefore, by tracking each small blob corresponding to the body parts of human, the moving human is successfully tracked. Recently, McKenna *et al.* [29] propose an adaptive background subtraction method in which color and gradient information are combined to deal with shadows and unpredictable color cues in motion segmentation. Tracking is then performed at three levels of abstraction: regions, people, and groups. Each region is represented by a bounding box and regions can merge and split. A human is composed of one or more regions assembled together under the condition of geometric structure constraints on the human body, and a human group consists of one or more people grouped together. Hence, using the region tracker and the individual color appearance model, seamless tracking of multiple people is attained, even during occlusion. As far as region-based vehicle tracking is concerned, there are some typical systems such as the CMS mobilizer system supported by the Federal Highway Administration (FHWA), at the Jet Propulsion Laboratory (JPL) [30], and the PATH system developed by the Berkeley group [31].

B. *Active Contour-Based Tracking*

Active contour-based tracking algorithms track moving objects by representing their outlines as bounding contours and updating these contours dynamically in consecutive frames [32], [33], [34],[35]. These algorithms aim at directly mining shapes of subjects and deliver more effective descriptions of objects than region-based algorithms. Paragios *et al.* [36] detect and track multiple moving objects in image sequences using a geodesic active contour objective function and a level set formulation scheme. Peterfreund [37] explores a new active contour model based on a Kalman filter for tracking non-rigid moving targets such as people in spatio-velocity space. Isard *et al.* [38] adopt stochastic differential equations to describe complex motion models, and combine this approach with deformable templates to cope with people tracking. Malik *et al.* [39], [40] have effectively applied active contour-based methods to vehicle tracking.

C. Feature-Based Tracking

Feature-based tracking algorithms accomplish detection and tracking of moving objects by extracting elements, clustering them into higher level features and then matching the features between images. Feature-based tracking algorithms can further be classified into three subcategories according to the nature of selected features: global feature-based algorithms, local feature-based algorithms, and dependence-graph-based algorithms.

1) The features used in global feature-based algorithms include centroids, perimeters, areas, some orders of quadratures and colors [41], [42], etc. Polana *et al.* [43] provide a good example of global feature-based tracking. A person is circumscribed with a rectangular box whose centroid is nominated as the feature for tracking. Even when occlusion occurs between two persons during tracking, as long as the velocity of the centroids can be distinguished efficiently, tracking is still successful.
2) The features used in local feature-based algorithms include line segments, curve segments, and corner vertices [44], [45], etc.
3) The features used in dependence-graph-based algorithms include a variety of distances and geometric relations between features [46].

IV. Behavior Analysis

After successfully tracking the moving objects from one frame to another in an image sequence, the next task for the visual surveillance system is understanding object behaviors from image sequences. Behavior understanding includes examining and recognition of motion patterns, and the creation of high-level description of actions and interactions. Models of the collective behavior tend to be bimodal. At one extreme are models that consider the entire crowd as one entity. Scholars have assumed that crowds transform individuals so that the resulting collective begins to exhibit a homogeneous “group mind” that is highly emotional and irrational [47]. At the other extreme are models considering everyone as independent members acting to maximize their own utility. For example, crowd behavior has been simulated by considering people as particles making local choices based on the principle of least effort [48]. One theory is that crowds consist primarily of small groups, defined as a “collection of individuals who have relations to one another that make them interdependent to some significant degree” [49]. Despite being instinctively reasonable, there has been unfortunately little work to validate this hypothesis. Johnson [50] claims that most crowds consist of small groups rather than isolated individuals (see also [51]). An unpublished study by McPhail found that 89 percent of people attending an event came with at least one other person, 52 percent with at least 2 others, 32 percent with at least 3 others, and that 94 percent of those coming with someone left with the people they came with [52].Behavior recognition including interpreting sequences of actions of one person or interactions of two or three are commonly based upon Hidden Markov Models [53] or Dynamic Bayes Networks [54]. A HMM is a kind of stochastic state machines [55]. It allows a more refined analysis of data with spatio-temporal variability. The use of HMMs consists of two stages: training and classification. In the training stage, the number of states of a HMM must be defined, and the corresponding state transition and output probabilities are adjusted in order that the generated symbols can correspond to the observed image features of the examples within a specific movement class. In the classification stage, the probability with which a particular HMM produces the test symbol sequence corresponding to the observed image features is computed. HMMs generally outperform DTW for undivided time series data, and are therefore widely applied to behavior understanding. For instance, Starner *et al.* [56] propose HMMs for the recognition of sign language. Oliver *et al.* [57] suggest and compare two different state-based learning architectures, namely, HMMs and coupled hidden Markov models (CHMMs) for modeling people behaviors and interactions such as following and meeting. The CHMMs are proven to work much more efficiently and accurately than HMMs. Brand *et al.* [58] illustrate that, by the usage of the entropy of the joint distribution to learn the HMM, a HMM’s internal state machine can be designed to establish observed behaviors into meaningful states. This technique has found applications in video monitoring and explanation, in low bit-rate coding of scene behaviors, and in abnormality detection. The limitation of all these approaches is that they can be typically applied to a small, known number of individuals. There is recent indication that more efficient recognition of group activities is possible by using a model of the group activity process to enable better analysis of the actions of individual members [59], [60].

Collective locomotion behavior is also considered in the traffic analysis and crowd simulation community. Models describing traffic flow can be categorized at the macroscopic or the microscopic level [61].Because macroscopic studies concentrate more on the space allocation for pedestrians in a facility than on the direct interaction between pedestrians, they are not as appropriate for predicting pedestrian groups as for evacuation planning. Microscopic models consider pedestrians as individual entities, with the collective crowd dynamics evolving from the interaction between agents. For example, in the social forces model [62], the behavior of an individual is subject to long-range forces caused by other pedestrians and environmental components such as obstacles and preferred areas. Similar approaches are used for multi-target tracking [63], [64] and abnormal behavior detection [66] in crowds. Another example is the Cellular Automaton (CA) model, where individuals move according to a preference matrix that specifies the probabilities for a particular walking direction and speed. Time and state are discretized in CA models, making them responsive to high-performance crowd simulation. Floor field models were presented to substitute individual agents' intelligence with a floor field that is altered by the pedestrians and in turn adjusts their preference matrices. The benefit of using the floor field is that it can turn long-range interactions into local forces. In visual surveillance, floor fields are estimated from visual data and used to support target tracking in dense crowds.

V. CONCLUSION

In this paper, we have provided an overview of all the object detection and tracking methods and how based on such information we can derive interesting information about the behavior of the objects that form the dynamic crowd. Currently in many visual surveillance tasks, we need to analyze the behavior of the people in a crowd e.g. whether a person is travelling alone or in a group, and if he is travelling in a group what kind of interactions are going on between them. All this information can be helpful for detecting any kind of suspicious activities that are happening within a crowd, thus identifying any kind of vulnerabilities that the crowd is susceptible to.

Three techniques for motion segmentation are addressed: background subtraction, temporal differencing, and optical flow. We have discussed four approaches to tracking: region based, active-contour based, feature based and model based. A brief overview of the HMMs has been provided for understanding the behavior of the objects that are present in scene. In addition, we examine the state-of-the-art of behavior description.

Our future extensions include further investigation on the different kinds of interactions among the entities that are possible within a crowd. Also we will be using some clustering techniques to analyze the groups that are forming within a crowd that can be used for realistic crowd simulation.

REFERENCES

[1] N. Sulman, T. Sanocki, D. Goldgof, and R. Kasturi, "*How effective is human video surveillance performance*?" in Proc. Int. Conf. Pattern Recog., 2008, pp. 1–3.

[2] Z. Ghahramani. Learning dynamic bayesian networks. In Adaptive Processing of Sequences and Data Structures. Lecture Notes in AI, pages 168–197, 1998.

[3] D. Heckerman. A tutorial on learning with Bayesian networks. Technical Report MSR-TR-95-06, Microsoft Research, 1995.

[4] M. Brand and V. Kettnaker. Discovery and segmentation of activities in video. *PAMI*, 22(8):844–851, August 2000.

[5] C. Vogler and D. Metaxas. A framework for recognizing the simultaneous aspects of American Sign Language. *CVIU*, 81:358–384, 2001.

[6] M. Brand, N. Oliver, and A. Pentland. Coupled hidden markov models for complex action recognition. In *CVPR*, pages 994–999, Puerto Rico,

[7] N. Oliver, B. Rosario, and A. Pentland. A bayesian computer vision system for modeling human interactions.*PAMI*,22(8):831–843, August 2000.

[8] Z. Ghahramani. Learning dynamic bayesian networks.In *Adaptive Processing of Sequences and Data Structures. Lecture Notes in AI*, pages 168–197, 1998.

[9] Anthony Hoogs Steve Bush Glen Brooksby

[10] T. Zhao and R. Nevatia, "Bayesian Human Segmentation in Crowded Situations," Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, pp. 459-466, 2003.

[11] G.M.Q. Yu and I. Cohen, "Multiple Target Tracking Using Spatio-Temporal Monte Carlo Markov Chain Data Association," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2007.

[12] W. Ge and R.T. Collins, "Marked Point Processes for Crowd Counting," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2009.

[13] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, pp. 886-893, 2005.

[14] P. Viola, M. Jones, and D. Snow, "Detecting Pedestrians Using Patterns of Motion and Appearance," Proc. IEEE Int'l Conf. Computer Vision, pp. 734-741, 2003.

[15] J. Heikkila and O. Silven, "A real-time system for monitoring of cyclists and pedestrians," in *Proc. IEEE Workshop Visual Surveillance*, 1999,pp. 74–81.

[16] C. Stauffer andW. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Int. Conf. Comput. Vis. PatternRecog.*, 1999, vol. 2, pp. 246–252.

[17] G. Halevy and D. Weinshall, "Motion of disturbances: Detection and tracking of multi-body non-rigid motion," in *Proc. IEEE Int. Conf. Comput.Vis. Pattern Recog.*, 1997, pp. 897–902.

[18] R. Cutler and L. Davis, "View-based detection and analysis of periodic motion," in *Proc. Int. Conf. Pattern Recog.*, 1998, pp. 495–500.

[19] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," in *Proc. IEEE Int. Conf.Comput. Vis.*, 1999, pp. 255–261.

[20] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Trans. Image Process.*,vol. 14, no. 3, pp. 294–307, Mar. 2005.

[21] N. Sulman, T. Sanocki, D. Goldgof, and R. Kasturi, "How effective is human video surveillance performance?" in *Proc. Int. Conf. PatternRecog*., 2008, pp. 1–3.

[22] F. Yin, D. Makris, and S. A. Velastin, "Performance evaluation of object tracking algorithms," in *Proc. IEEE Int. Workshop Perform.Eval.Tracking Surveillance*, 2007, pp. 733–736.

[23] S. Muller-Schneiders, T. Jager, H. S. Loos, and W. Niem, "Performance evaluation of a real time video surveillance system," in *Proc. IEEE Int. Workshop Visual Surveillance Perform. Eval.Tracking Surveillance*, 2005, pp. 137–143.

[24] T. List, J. Bins, J. Vazquez, and R. B. Fisher, "Performance evaluating the evaluator," in *Proc. IEEE Int. Workshop Visual Surveillance Perform.Eval.Tracking Surveillance*, 2005, pp. 129–136.

[25] F. Ziliani, S. A. Velastin, F. Porikli, L. Marcenaro, T. Kelliher, A. Cavallaro, and P. Bruneaut, "Performance evaluation of event detection solutions: The CREDS experience," in *Proc. IEEE Conf. Adv. Video Signal Based Surveillance*, 2005, pp. 201–206.

[26] K. Karmann and A. Brandt, "Moving object recognition using an adaptive background memory," in *Time-Varying Image Processing and Moving Object Recognition*, V. Cappellini, Ed. Amsterdam, The Netherlands: Elsevier, 1990, vol. 2.

[27] M. Kilger, "A shadow handler in a video-based real-time traffic monitoring system," in *Proc. IEEE Workshop Applications of Computer Vision*, Palm Springs, CA, 1992, pp. 11–18.

[28] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: real-time tracking of the human body," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 780–785, July 1997.

[29] S. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking groups of people," *Comput. Vis. Image Understanding*, vol.80, no. 1, pp. 42–56, 2000.

[30] JPL, "Traffic surveillance and detection technology development" Sensor Development Final Rep., Jet Propulsion Laboratory Publication no. 97-10, 1997.

[31] J. Malik, S. Russell, J. Weber, T. Huang, and D. Koller, "A machine vision based surveillance system for Californaia roads," Univ. of California, PATH project MOU-83 Final Rep., Nov. 1994.

[32] A. Baumberg and D. C. Hogg, "Learning deformable models for tracking the human body," in *Motion-Based Recognition*, M. Shah andR. Jain, Eds. Norwell, MA: Kluwer, 1996, pp. 39–60.

[33] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-based object detection in images by components," *IEEE Trans. Pattern Recognit. Machine Intell.*, vol. 23, pp. 349–361, Apr. 2001.

[34] A. Galata, N. Johnson, and D. Hogg, "Learning variable-length Markov models of behavior," *Comput. Vis. Image Understanding*, vol. 81, no. 3, pp. 398–413, 2001.

[35] Y. Wu and T. S. Huang, "A co-inference approach to robust visual tracking," in *Proc. Int. Conf. Computer Vision*, vol. II, 2001, pp. 26–33.

[36] N. Peterfreund, "Robust tracking of position and velocity with Kalmansnakes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp.564–569, June 2000.

[37] M. Isard and A. Blake, "Contour tracking by stochastic propagation of conditional density," in *Proc. European Conf. Computer Vision*, 1996, pp. 343–356.

[38] D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, and S.Russel, "Toward robust automatic traffic scene analysis in real-time," in *Proc. Int. Conf. Pattern Recognition*, Israel, 1994, pp. 126–131.

[39] J. Malik and S. Russell, "Traffic Surveillance and Detection Technology Development: New Traffic Sensor Technology," Univ. of California, Berkeley, California PATH Research Final Rep., UCB-ITS-PRR-97-6, 1997.

[40] C. A. Pau and A. Barber, "Traffic sensor using a color vision method," in *Proc. SPIE—Transportation Sensors and Controls: Collision Avoidance, Traffic Management, and ITS*, vol. 2902, 1996, pp. 156–165.

[41] B. Schiele, "Vodel-free tracking of cars and people based on color regions," in *Proc. IEEE Int. Workshop Performance Evaluation of Tracking and Surveillance*, Grenoble, France, 2000, pp. 61–71.

[42] R. Polana and R. Nelson, "Low level recognition of human motion," in *Proc. IEEE Workshop Motion of Non-Rigid and Articulated Objects*, Austin, TX, 1994, pp. 77–82.

[43] B. Coifman, D. Beymer, P. McLauchlan, and J. Malik, "A real-time computer vision system for vehicle tracking and traffic surveillance," *Transportation Res.: Part C*, vol. 6, no. 4, pp. 271–288, 1998.

[44] J. Malik and S. Russell, "Traffic surveillance and detection technology development (new traffic sensor technology)," Univ. of California, Berkeley, 1996.

[45] T. J. Fan, G. Medioni, and G. Nevatia, "Recognizing 3-D objects using surface descriptions," *IEEE Trans. Pattern Recognit. Machine Intell.*, vol. 11, pp. 1140–1157, Nov. 1989.

[46] R.W. Brown, "Mass Phenomena," Handbook of SocialPsychology, G. Lindzey, ed., vol. 2, pp. 833-876, Addison Wesley, 1954.

[47] G. Still, "Crowd Dynamics," PhD thesis, Univ. of Warwick, 2000.

[48] D. Cartwright and A. Zander, Group Dynamics: Research and Theory, third ed. Harper, 1968.

[49] N.R. Johnson, "Panic at the Who Concert Stampede: An Empirical Assessment," Social Problems, vol. 34, pp. 362-373, 1987.

[50] A. Aveni, "The Not-So-Lonely Crowd: Friendship Groups in Collective Behavior," Sociometry, vol. 49, pp. 96-99, 1977.

[51] C. McPhail, "Withs across the Life Course of Temporary Sport Gatherings," unpublished manuscript, Univ. of Illinois, 2003.

[52] M. Brand, N. Oliver, and A. Pentland, "Coupled hidden Markov models for complex action recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1997, pp. 994–999.

[53] T. Starner, J. Weaver, and A. Pentland, "Real-time American sign language recognition using desk and wearable computer-based video,"*IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 1371–1375,Dec. 1998.

[54] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 831–843, Aug. 2000.

[55] M. Brand and V. Kettnaker, "Discovery and segmentation of activities in video," IEEE Trans. Pattern Anal. Machine Intell., vol. 22, pp. 844–851,Aug. 2000.

[56] M. Ryoo and J. Aggarwal, "Recognition of High-Level Group Activities Based on Activities of Individual Members," Proc. IEEE Workshop Motion and Video Computing, pp. 1-8, Jan. 2008.

[57] W. Zhang, F. Chen, W. Xu, and Y. Du, "Hierarchical Group Process Representation in Multi-Agent Activity Recognition," Image Comm., vol. 23, pp. 739-739, Jan. 2008.

[58] A. May, Traffic Flow Fundamental. Prentice Hall, 1990.

[59] D. Helbing and P. Molnar, "Social Force Model for Pedestrian Dynamics," Physical Rev. E, vol. 51, no. 5, pp. 4282-4286, 1995.

[60] P. Scovanner and M. Tappen, "Learning Pedestrian Dynamics from the Real World," Proc. IEEE Int'l Conf. Computer Vision, 2009.

[61] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool, "You'll Never Walk Alone: Modeling Social Behavior for Multi-Target Tracking,"

[62] R. Mehran, A. Oyama, and M. Shah, "Abnormal Crowd Behavior Detection Using Social Force Model," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2009.

[63] V. Blue and J. Adler, "Cellular Automata Microsimulation for Modeling Bi-Directional Pedestrian Walkways," Transportation Research B, vol. 35, no. 3, pp. 293-312, 2001.

[64] A. Schadschneider, "Cellular Automaton Approach to Pedestrian Dynamics—Theory," Proc. Int'l Conf. Pedestrian and Evacuation Dynamics, pp. 75-86, 2002.

[65] Shaogang Gong and Tao Xiang,"Recognition of Group Activities using Dynamic Probabilistic Networks," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2009.

[66] S. Ali and M. Shah, "Floor Fields for Tracking in High Density Crowd Scenes," Proc. European Conf. Computer Vision, pp. 1-14, Oct. 2008.