Multiple People Tracking using Contextual Trajectory Forecasting

Pranav Mantini
Dept. of Comp. Sci., Univ. of Houston
Houston, Texas 77204-3010
Email: pmantini@uh.edu

Shishir K. Shah
Dept. of Comp. Sci., Univ. of Houston
Houston, Texas 77204-3010
Email: sshah@central.uh.edu

Abstract—**People tracking is the ability to identify the position of a specified person in the camera view with the progression of time. Trajectory forecasting is the task of predicting the likely path that a person might take to reach a destination. Contextual trajectory forecasting (CTF) leverages the 3D geometric information and static objects in the environment along with observed behavioral norms for human path prediction. In this paper, we enhance CTF to also account for dynamic objects in the environment (like other humans) for prediction. The proposed tracking algorithm makes use of traditional HSV histogram appearance features for detection and combines it with the enhanced CTF for tracking. A maximum likelihood minimum mean square error data association filter is used to probabilistically associate the appearance detections and the CTF predictions for tracking. Two real world scenarios with 49 ID's were used to evaluate the proposed algorithm. The result show a significant improvement over a baseline tracking algorithm (HSV histogram) and a state-of-the-art online multi person tracking algorithm.**

I. Introduction

Tracking has applications in multiple disciplines like surveillance, robot motion planning, etc. For example, in surveillance, it can be used to monitor a scene and detect abnormal activities. In robot motion planning, it can be used to identify people and plan a path to avoid collisions [1]. Recent methods in people tracking follow a two stage cycle of detection and prediction as shown in figure 1. In the detection phase, an appearance model is used to describe the object of interest and the location of the object is initialized for the tracking process. In the prediction stage, a motion model is used to predict the future location of the detected objects and based on this prediction, a localized area is defined where the object might exist.

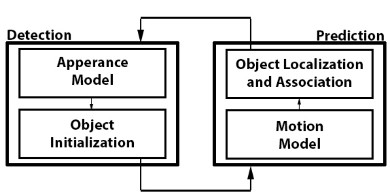

Fig. 1. Steps involved in the tracking process.

Given a good appearance model and the motion model, object initialization, localization and association can be trivial tasks. The core of the detection stage is the appearance model and that of the prediction stage is the motion model. Hence significant research has been focused on improving appearance models for detection and motion models for prediction. This work proposes the use of a human motion model for prediction, which can be used in conjunction with any human appearance based detection model. Motion models assume an underlying law for predicting the future state of the object being tracked. For example, consider tracking a free falling ball. The laws of gravity can be used to generate a motion model. However, in this case the objects being tracked are humans and the design of the underlying law/system is non-trivial. Human motion can be complex, attributing to the multitude of factors that influence it, like destination, geometry, other humans, etc. [2]

Destination is the motivating factor for human motion. Most people traverse with the objective of reaching a destination, usually within the shortest time possible while adhering to social norms imposed by the geometry and other humans [2].

Geometry of the environment (like the walls and doorways) and static objects have an effect on the human motion. For example in a narrow hallway with walls on either side, majority of the humans prefer to walk at the center of the hallway as opposed to the edge. Furthermore, in a classroom, consider how one would navigate around tables and chairs or any other object to get to a seat [2].

Humans in the environment are dynamic and the human motion of the subject is effected by other humans and vice versa. For example, consider how humans plan their motion to navigate around other humans while maintaining some socially acceptable distance.

Human motion can also be effected by other social and cultural factors. In this work, we propose a motion model for prediction that accounts for geometry, objects, and other humans in the environment. The proposed method takes as input the entire 3D geometry of the environment and performs motion prediction in 3D. The contributions of this paper are,

- We propose a model for motion estimation that accounts for static elements (geometry of the environment and the objects in it), dynamic elements (moving humans in the environment), destination and employs observed human behavioral norms for prediction.
- We propose a method for tracking humans by leveraging the aforementioned motion estimation model which can

handle occlusions and even allow for tracking across non-overlapping cameras.

II. Previous Work

Tracking is an important low level algorithm for numerous applications. Accounting to this, an immense value of research has been conducted in this area. Yilmaz *et al.* [1] and Watada *et al.* [3] performed a broad survey in object and human tracking respectively. As mentioned earlier, appearance models and motion models are the core of the detection and prediction stages. This work is focused on the prediction stage for tracking, and hence appearance methods are first briefly discussed followed by a detailed survey on existing prediction methods for tracking.

A. Detection methods

The objective of the detection phase is to identify the location of the objects of interest in the image. Yilmaz **et al.** [1] categorized the detection methods as follows:

1) Points based methods: The objects to be tracked are represented by a set of one or more interest points. Image features like contours of lines and end points [4], [5], [6] or color and contrast of object intensities [7], [8] were used to identify points of interest.

2) Segmentation based methods: The objective of segmentation is to partition the object to be tracked from the image. Commonly used segmentation techniques for tracking were the mean shift algorithm [9], [10] and histograms [11], [12], [13].

3) Supervised learning based methods: These methods use a dataset representing the object to train a classifier for identifying the object of interest. This classifier was used to detect the regions with object of interest in the images for tracking [14], [15].

4) Background subtraction methods: The objective is to isolate the foreground pixels or the objects of interest by identify and removing all the background pixels. This was the popular and conventional technique for tracking [16], [17], [18].

B. Prediction methods

Majority of the early methods used for tracking were based on detection alone, however, methods proposed latter supplemented the detection methods with the prediction phase for estimating the future location of the object. This allowed for faster tracking methods for two reasons, first, the prediction provided with an approximate location of the object for the detection algorithm and second, detection algorithm had to be run only on every few frames as opposed to every frame since the location of the object could be predicted. Motion models can be designed at a 2D level on the image plane or at a 3D level on the ground plane. In general, tracking in 3D can have an advantage over 2D when handling occlusions. The proposed model performs tracking in 3D on the ground plane and hence these method are discussed more in detail than model performing tracking in 2D.

1) Prediction on image plane: In these methods, the first few frames were used to learn the motion of the object and then a statistical algorithm was initialized based on the learned motion to predict the future states of the object. The most commonly applied techniques were Kalman filter [19], [20], [21] and particle filter [22], [23], [24]. Although these methods can handle occlusion to a certain extent better than detection methods, they perform poorly when tracking an object with complex motion like humans. For further reading on this methods or detections method, the readers can refer to [3], [1].

2) Prediction on the ground plane: A fair component of work in this area has been conducted in the robotics community, as laser range sensors allowed for a natural way to work in 3D on the ground plane in contrast to a video sensor in computer vision which required calibration and homography mapping. These methods can further be sub-categorized as follows.

Non-behavioral models: These models do not account for the complexity of human behavior and assume a linear interpolation or constant velocity for prediction. Fod *et al.* [25] proposed the use of a constant velocity model for prediction in a laser range sensor environment. The previous scans were used to estimate the velocity of the object and a Kalman filter was used to estimate the future position for tracking. Schulz *et al.* [26] introduced sample-based Joint Probabilistic Data Association Filters (SJPDAFs) for people tracking using a laser sensor which uses particle filter to track the state of the object and apply (JPDAFs) for association. Similar to Kalman filters, the previous measurements were used for prediction using a particle filter. Cui *et al.* [27] demonstrated tracking using rows of laser scanners and a video camera by employing a common Kalman filter for prediction. Arras *et al.* [28] also assumed a constant velocity model to track people's legs with laser scanner using Kalman filter and also explicitly handled occlusion for tracking. These models can cope with occlusion to a certain extent and can be used for tracking objects with linear motion. However they might not be sufficient to track humans as they exhibit complex motion.

Human behavioral models: These models either learn human behavior from observation or explicitly model it for prediction. In the former, human trajectories are observed in the scenario and a motion pattern is learned for prediction in contrast to the latter where the influence of the various factors (geometry, objects and humans) on human motion are explicitly modeled.

Liao *et al.* [29] used the floor map of the environment to generate a Voronoi graph and assumed that people travel along the edges of the map. Observed motion patterns with a laser scanner on a robot was used to calculate the transition probabilities along the edges of the Voronoi graph. These probabilities along the graph were used for prediction in tracking. Bruce and Gordon [30] observed people trajectories using laser sensor on a robot to learn destinations in the environment. Later the human motion is predicted to an estimated goal location along the path predicted by a planner. Bennewitz *et al.* [31] clustered observed human trajectories from a laser range sensor and employed Expectation-Maximization algorithm to form motion patterns. Hidden Markov model was used to predict the future states of people for tracking. Weser *et al.* [32] proposed the use of self organizing maps to learn motion patterns from trajectories obtained from a laser range sensor. A particle filter was used to predict the future position of humans using the learned motion patterns for tracking. These models can handle occlusions in static environments but fail when deployed in a dynamic environment with moving humans because the learned motion pattern is not accurate anymore. Furthermore introduction of a new static object in the environment would require new observations for training and generating the motion patterns.

These drawbacks were overcome by explicitly modeling human behavior and supplementing it with observed motion patterns for prediction. Antonini *et al.* [33] proposed the use of discrete choice models with varying velocity options to build a probability distribution and sample the future state by accounting for other humans and environment. The predictions were used for detection and tracking in a video sequences. Pellegrini *et al.* [34] proposed Linear Trajectory Avoidance (LTA) model taking into consideration other humans and static objects such that the pedestrians steer clear to avoid collision. The model was incorporated for tracking in video data. Yamaguchi *et al.* [35] defined an energy function that evaluates the future states based on destination, other humans, static objects and group behavior. An energy minimization framework was used for predicting the future states. The results were demonstrated using a tracking algorithm in video sequences. Luber *et al.* [36] proposed the use of Social Force Model [37] for tracking in data collected by laser scanner and video data. Social Force model was one of the earliest work in human motion dynamics which modeled the interaction between humans, objects and geometry as repulsion and attraction forces. Gong *et al.* [38] implemented Multi-Hypothesis motion planning for video tracking. This model takes into account the geometry and hypothesizes multiple routes around objects, but fails to model the social interaction between objects and humans. Luber *et al.* [39] generated a spatial affordance map which represented the global human activity of the environment assuming events occur as a Poisson's process. This map was incorporated into a multi-hypothesis tracker for enhancing motion prediction for tracking using a laser range sensor.

Mantini and Shah [2] proposed a human motion prediction model that generates an occupancy map for any geometry based on observed human behavior. This model takes into account the static objects and geometry into account and predicts trajectory to a given destination. This model was also shown to enhance re-identification by incorporating human behavioral context [40], [41]. This model is closely related to the proposed method, however [2] does not account for interaction with dynamic objects (humans). In the proposed method the model is enhanced to handle social interaction with humans and incorporated into a tracking algorithm.

III. Methodology

Let G be the geometry of the environment and $P = \{p_1, p_2, ...\}$ be accessible points on the floor. Let there exists a function $F : P \rightarrow \Re$ that quantifies the accessibility $F(p)$ of a point p on the floor with respect to the geometry and other humans in the environment. Let $\tau = \{(p^\tau_1, t^\tau_1), (p^\tau_2, t^\tau_2), ..., (p^\tau_n, t^\tau_n)\}$ be human motion trajectory from p^τ_1 to p^τ_n such that $(p^\tau_1, ..., p^\tau_n) \in P$ and t^τ_i is the time stamp when the human is located at p^τ_i. Given the function F, the trajectory can be modeled as a Markov chain model.

$$P(p_{i+1}|p_i, p_{i-1}, ..., p_1, F) = P(p_{i+1}|p_i, F) \qquad (1)$$

Given this probability distribution, consecutive points can be sampled from the floor to form a trajectory. The problem then simplifies to generating a function F that accounts for destination, geometry and other humans in the environment and assigns higher values to the points on the floor that adhere to social norms. The original method is recapped to describe how the third component that models the effect of dynamic objects influences the occupancy map.

A. Contextual Trajectory Forecasting

Contextual Trajectory Forecasting (CTF) model predicts the future location of humans based on their destination, geometry and other humans in the environment. It is motivated by the idea, humans in general try to reach a destination in the shortest time/distance possible while adhering to social norms imposed by the geometry and other humans in the environment. This is an extension to the model described in [2], the whole model is described for the completeness of the paper. Consider the geometry shown in figure 2. The effect of the factors (distance, geometry and humans) on this floor plan are demonstrated below.

Fig. 2. Geometry of a Floor Plan.

1) Destination: Consider the destination a human is trying to reach be as shown in figure 3. A distance map $D(p_i)$ is created which indicates the distance of the point p_i from the destination. Geodesic distances are used instead of Euclidean distances as they are more accurate on account of the complex polygonal geometry of the floor plans. In the absence of the effect of the geometry or other human in the environment, one could create a probability distribution that is inversely proportional to the distance map and sample points consecutively to generate a trajectory to the destination. This would result in a trajectory that represents the shortest path to the destination.

$$F(p_i) \propto -D(p_i) \qquad (2)$$

2) Geometry: It is well understood that 3D geometry of the environment and the objects in it impose specific constrains on human motion. For example, when walking in the hallways, humans tend to maintain a certain distance from the surrounding walls. When they encounter an object in the path, humans tend to go around them while still maintaining some distance from it. Given the geometry of the environment and static objects in it, hypothetically if a large number of trajectories followed by humans are observed in the environment, it can be assumed that certain points would be accessed more often than the other. For example, points on the floor that are next to the wall or an object might be accessed less often than those that are farther away from any static geometry. Hence, there could be a certain distribution to the floor that is static and dependent on the geometry. Mantini and Shah [2] proposed an accessibility map $A(p_i)$ which defines the accessibility of any point on the floor based on its surrounding geometry. They proposed a linear relationship between the geometry and accessibility, and used linear regression to obtain the accessibility $A(p_i)$ for a point p_i in any geometry with out the need for training. The obtained accessibility map is shown in figure 4.

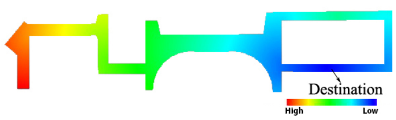

Fig. 3. Distance Map to Destination.

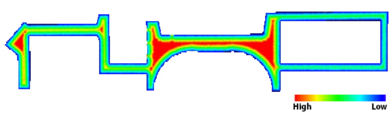

Fig. 4. Accessibility Map based on the geometry.

The accessibility map was combined with the distance map (Figure 5) to obtain a distribution that allows sampling points for the trajectory that represent the shortest distance while following social norm concerning geometry and objects. The function F for any point p_i in the geometry was defined as

$$F(p_i) = -D(p_i)/A(p_i) \tag{3}$$

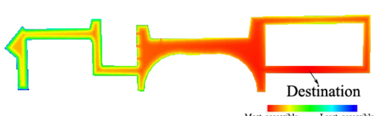

Fig. 5. Accessibility Map Combined with Distance Map to Destination.

3) Humans: Objects and geometry are static and so is their effect on the accessibility map and human motion. However, humans in the environment are dynamic and hence their effect on the accessibility of a point is also dynamic. A human's motion is effected by the other humans in the environment and vice versa. The effect of other humans on the accessibility map is modeled using the Theory of Proxemics [42]. Theory of Proxemics in an observational study, that define how humans utilize the physical space around them. This theory classifies the space close to a human into four discrete regions: Intimate, Personal, Social and Public distance. The proposed method adapts a continuous effect on the accessibility map. Let a human be present at the point p_i on the floor. The effect of this human on the accessibility map is defined as an exponentially increasing function with distance from the location of the human.

$$H(p_j) = 1 - \exp^{-d(p_i,p_j)/k} \tag{4}$$

Where, $d(p_i, p_j)$ is the euclidean distance between the points p_i and p_j and k is a constant. This would make the accessibility at the location of the human ($p_i = p_j$) to be zero and increase exponentially as the distance increase. This is combined with the effect of the geometry and the destination to obtain F.

$$F(p_i) = \frac{-D(p_i)}{A(p_i)} \prod_j (1 - exp^{d(p_i,p_j)/k}) \tag{5}$$

where p_j is the position of the humans in the environment in view of the human whose motion is predicted. The obtained function F (shown in Figure 6) illustrates the effect of two human on the accessibility map. This would be the accessibility map for a third human trying to reach the destination.

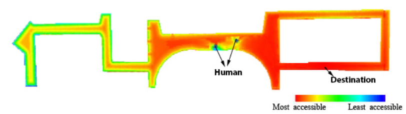

Fig. 6. Effect of other humans on the Accessibility Map.

4) Trajectory Sampling: The function F is used to build a transition matrix for sampling points in the Markov chain. If the current location is p_t, we assume that the only possible points of transition are the neighbors of $\{p_{t1}, ..., p_{tm}\}$. The probability of transitioning to these neighbors is defined as

$$\propto \left\{ \begin{array}{ccc} F(p_{tm}) - F(p_t) & if & D(s_{tm}) - D(s_t) \leq 0 \\ 0 & otherwise & \end{array} \right\} \tag{6}$$

In order to reach the destination in the shortest time possible while conforming to the social norms, only the points closer to the destination are chosen ($D(p_{tm}) - D(p_t) <= 0$). The neighbors are sampled to form consecutive points in the trajectory.

B. Tracking using Contextual Trajectory Forecasting

The proposed tracking framework involves five steps as shown in figure 7.

1) Initialize the 3D model, occupancy map, location and appearance of the humans to be tracked.
2) Predict the future location of the human using CTF.
3) Localize a search region around the predicted locations and perform human detection.
4) Associate the observed data with the existing data using maximum likelihood - minimum meas square error filter based on the location and appearance.
5) Update the location and the histogram input to step 2 to continue prediction.

Let $Y = \{y_1, y_2, ...\} = \{(p_{y_1}, h_{y_1}), (p_{y_2}, h_{y_2}), ...\}$ be the description of human $i = \{1, 2, ..., n\}$ being tracked, where p_i is the physical location in 3D geometry and h_i the HSV histogram of the human at time t_i. Given the geometry, occupancy map and the corresponding destination of the human, the future location of the human is predicted using CTF. Let p_i be the predicted location of the human at time t'_i. The

Method	Misses	False Positives	True Positives
Baseline	0.295	0.348	0.723
Zhang *et al.* [18]	0.509	0.0611	0.793
Proposed	0	0.325	1.076

TABLE I
MISSES, FALSE POSITIVES AND TRUE POSITIVES SHOWN AS ID'S PER FRAME FOR GEOMETRY A.

point p_i' is projected on tho the image plane and a search region s_i' is defined. The search region is subjected to a human detection algorithm to obtain observations $Z = \{z_1, z_2, ...\} = \{(p_{z_1}, h_{z_1}), (p_{z_2}, h_{z_2}), ...\}$ where $j = \{1, 2, ..., m\}$ be the set of all observations.

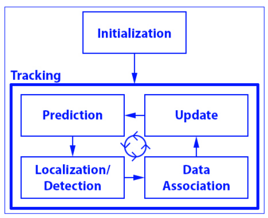

Fig. 7. Tracking Framework.

A maximum likelihood minimum mean square error data association filter is used to assign the observed data (Z) to the current state data (Y). Let $A_i = \{(y_{i_1}, z_{i_1}), (y_{i_2}, z_{i_2})...\}$ be an association such that $y_{i_j} \in Y, z_{i_j} \in Z$ and $A_i \in A$, where A is the set of all mutually exclusive and exhaustive events between the sets Y and Z.

$$\begin{aligned} i &= (y, z) \\ &= \arg \max_{y \in Y, z \in Z} P(y, z|A_i) \end{aligned} \tag{7}$$

$$\begin{aligned} P(z, y|A_i) &= P(z, y|A) \\ &= P(z = z_i|y = y_i) \\ &= P((p_{z_i}, h_{z_i})|(p_{y_i}, h_{y_i})) \\ &= P(p_{z_i}|p_{y_i}) * P(h_{z_i}|h_{y_i}) \end{aligned} \tag{8}$$

$$\begin{aligned} P(p_{z_i}|p_{y_i}) &\propto (1 - d(p_{z_i}, p_{y_i})) \\ P(h_{z_i}|h_{p_i}) &\propto d_h(h_{z_i}, h_{y_i})) \end{aligned} \tag{9}$$

Where $d(p_{z_i}, p_{y_i})$ is the Euclidean distance between the point p_{z_i} and p_{y_i} and $d_h(h_{z_i}, h_{y_i})$ is the histogram intersection distance between h_{z_i} and h_{y_i}. Finally, the corresponding state of the human are updated according to the association model.

IV. EXPERIMENTS

Two real world scenarios were considered to evaluate the performance of the tracker. The geometry of the environment and their corresponding views are as shown in figure 8. A total of 49 ID's were used to evaluate the tracking algorithm of which 30 were from geometry B and the rest from geometry A. The dataset consisted of 15,000 frames containing 3 scenarios with 4 people, 5 scenarios with 3 people, 5 scenarios with 2 people walking simultaneously, the rest consisted of tracking 1 person. We compare the results of the proposed tracker against

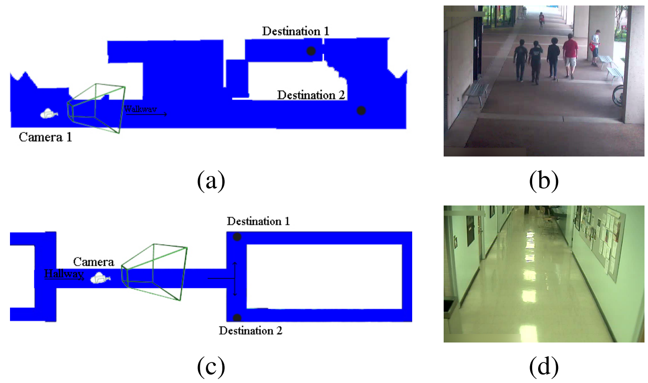

Fig. 8. (a) Geometry A; (b) View of the camera located in Geometry A; (c) Geometry B; (d) View of the camera located in Geometry B

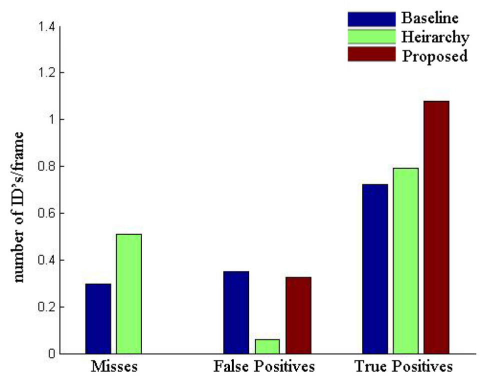

Fig. 9. Misses, false positives and true positives for Geometry A

two other tracking algorithms. As a baseline, the tracker is first compared against the results from using histogram data alone from data association, i.e. with out the prediction. This will quantify the effect of the prediction algorithm on tracker's performance. Further more, the results are compared against [18], which is a state of the art online multi-person tracker proposed by Zhang *et al.* The results for geometry A are shown in figure 9, table I and for geometry B are shown in 10, table II and are quantified as misses, false positives and true positives. The proposed algorithm has no misses, this is because, in the absence of a detection, the location of the object can be estimated from the trajectory prediction. The hierarchy tracker has the lowest number of false positives, this is because, if the detector fails to identify an object continuously, the algorithm stops tracking. Hence it has higher number of misses than the base line. The proposed algorithm has the highest number of true positives, out performing the baseline and the hierarchy tracker.

V. CONCLUSION

We have implemented a methodology to predict the future position of human subjects using contextual trajectory forecasting. Finally we have successfully coupled CTF with a traditional appearance based tracking algorithm. Preliminary results show that using the 3D geometry and contextual

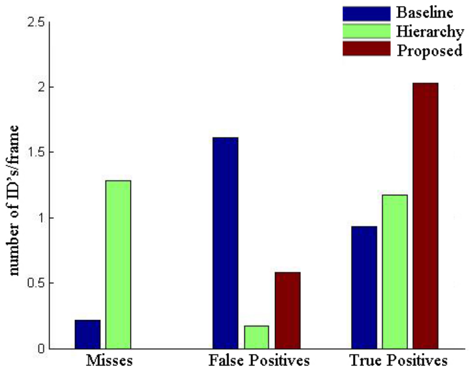

Fig. 10. Misses, false positives and true positives for Geometry B

Method	Misses	False Positives	True Positives
Baseline	0.215	1.614	0.933
Zhang *et al.* [18]	1.284	0.172	1.174
Proposed	0	0.584	2.0252

TABLE II
MISSES, FALSE POSITIVES AND TRUE POSITIVES SHOWN AS ID'S PER FRAME FOR GEOMETRY B.

trajectory forecasting can enhance tracking performance significantly and the results were compared with the state of the art detection based tracking methods. A Large scale study will be taken into consideration in the future.

ACKNOWLEDGMENT

This work was supported in part by the US Department of Justice 2009-MU-MU-K004. Any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of our sponsors.

REFERENCES

[1] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv. '06*.
[2] P. Mantini and S. K. Shah, "Human trajectory forecasting in indoor environments using geometric context," in *Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing*, ser. ICVGIP '14.
[3] J. Watada, Z. Musa, L. Jain, and J. Fulcher, "Human tracking: A state-of-art survey," in *Knowledge-Based and Intell. Inf. and Engg. Sys. '10*.
[4] M. Yokoyama and T. Poggio, "A contour-based moving object detection and tracking," in *Vis. Surv. and Perf. Eval. of Tracking and Surv., '05.*
[5] Y. Wang, E. Teoh, and D. Shen, "Lane detection using b-snake," in *Inf. Intell. and Sys., '99.*
[6] Y.-y. Cao, G.-y. Xu, and T. Riegel, "Moving object contour detection based on s-t characteristics in surveillance," in *Human Interface and the Management of Information. '07*.
[7] K. Yokoi, "Probabilistic bprrc: Robust change detection against illumination changes and background movements," in *MVA '09*.
[8] S. Ghiasi, H. Moon, A. Nahapetian, and M. Sarrafzadeh, "Collaborative and reconfigurable object tracking," *The Journal of Supercomputing, '04*.
[9] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, '02*.
[10] G. Bradski, "Real time face and object tracking as a component of a perceptual user interface," in *WACV '98*.
[11] Z. Zivkovic and B. Krose, "An em-like algorithm for color-histogram-based object tracking," in *CVPR '04*, 2004.
[12] J. Villalba Espinosa, J. Gonzlez Linares, J. Ramos Czar, and N. Guil Mata, "Kernel-based object tracking using a simple fuzzy color histogram," in *Advances in Computational Intelligence '11*.
[13] S. Shahed Nejhum, J. Ho, and M.-H. Yang, "Visual tracking with histograms and articulating blocks," in *CVPR '08*.
[14] C.-J. Pai, H.-R. Tyan, Y.-M. Liang, H.-Y. M. Liao, and S.-W. Chen, "Pedestrian detection and tracking at crossroads," in *ICIP '03*.
[15] Z. Qui, D. Yao, Y. Zhang, D. Ma, and X. Liu, "The study of the detection of pedestrian and bicycle using image processing," in *Intell. Trans. Sys., '03*.
[16] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in *CVPR '99*.
[17] S. Saravanakumar, A. Vadivel, and C. Saneem Ahmed, "Multiple human object tracking using background subtraction and shadow removal techniques," in *ICSIP '10*.
[18] R. Zhang and J. Ding, "Object tracking and detecting based on adaptive background subtraction," 2012 Int. Workshop on Info. and Elec. Engg.
[19] S.-K. Weng, C.-M. Kuo, and S.-K. Tu, "Video object tracking using adaptive kalman filter," *J. Vis. Comun. Image Represent. '06*.
[20] S. Huang and J. Hong, "Moving object tracking system based on camshift and kalman filter," in *CECNet '11*.
[21] M. Mirabi and S. Javadi, "People tracking in outdoor environment using kalman filter," in *ISMS '12*.
[22] M. Isard and A. Blake, "Contour tracking by stochastic propagation of conditional density," in *Computer Vision ECCV '96*.
[23] X. Xu and B. Li, "Rao-blackwellised particle filter for tracking with application in visual surveillance," in *Vis. Surv. and Perf. Eval. of Tracking and Surv., '05.*
[24] Z. Ye and Z.-Q. Liu, "Tracking human hand motion using genetic particle filter," in *SMC '06*.
[25] A. Fod, A. Howard, and M. J. Mataric, "Laser-based people tracking," in *ICRA '02*.
[26] D. Schulz, W. Burgard, D. Fox, and A. B. Cremers, "People tracking with a mobile robot using sample-based joint probabilistic data association filters," 2003.
[27] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, "Tracking multiple people using laser and vision," in *IROS '05*.
[28] K. Arras, S. Grzonka, M. Luber, and W. Burgard, "Efficient people tracking in laser range data using a multi-hypothesis leg-tracker with adaptive occlusion probabilities," in *ICRA '08*.
[29] L. Liao, D. Fox, J. Hightower, H. Kautz, and D. Schulz, "Voronoi tracking: location estimation using sparse and noisy sensor data," in *IROS '03*.
[30] A. Bruce and G. Gordon, "Better motion prediction for people-tracking," *ICRA '04*.
[31] M. Bennewitz, W. Burgard, G. Cielniak, and S. Thrun, "Learning motion patterns of people for compliant robot motion," *Internationl Journal of Robotics Research '05*.
[32] M. Weser, D. Westhoff, M. Huser, and J. Zhang, "Multimodal people tracking and trajectory prediction based on learned generalized motion patterns," in *Multisensor Fusion and Integration for Intel. Sys., '06*.
[33] G. Antonini, S. Martinez, M. Bierlaire, and J. Thiran, "Behavioral priors for detection and tracking of pedestrians in video sequences," *International Journal of Computer Vision, '06*.
[34] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *ICCV '09*.
[35] K. Yamaguchi, A. Berg, L. Ortiz, and T. Berg, "Who are you with and where are you going?" in *CVPR '11*.
[36] M. Luber, J. Stork, G. Tipaldi, and K. Arras, "People tracking with human motion predictions from social forces," in *ICRA '10*.
[37] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Phys. Rev. E '95*.
[38] H. Gong, J. Sim, M. Likhachev, and J. Shi, "Multi-hypothesis motion planning for visual object tracking," ser. ICCV '11.
[39] M. Luber, G. Diego Tipaldi, and K. O. Arras, "Place-dependent people tracking," *Int. J. Rob. Res. '11*.
[40] P. Mantini and S. K. Shah, "Enhancing re-identification through contextual trajectory forecasting," in *Technologies for Homeland Security (HST), 2015 IEEE International Conference on*, 2015.
[41] P. Mantini and S. Shah, "Camera placement optimization conditioned on human behavior and 3d geometry," 2016.
[42] E. T. Hall, *The Hidden Dimension*. Anchor Books. ISBN 0-385-08476-5, 1966.