Global Scientific JOURNALS

GSJ: Volume 6, Issue 1, January 2018, Online: ISSN 2320-9186
www.globalscientificjournal.com

ANALYSIS OF MOVING OBJECTS IN VIDEOS

Priyadharsini.N.K[1]
Department of CSE, PACET
priyadharsini.pacet@gmail.com

D. Chitra
Department of CSE,PACET

Abstract— Video processing is a technique of processing individual frames or images. This process involves acquisition, manipulation, transmission, analysis and compression. This paper focuses on video analysis; it includes motion segmentation and motion tracking. Tracking objects in a video containing extremely crowded scenes is a challenging due to the motion and appearance variability produced by large number of people within the scene. The individual pedestrians collectively form a crowd that exhibits a spatially and temporally structured pattern within the scene. The video is divided into sub volumes. The local spatio-temporal motion of the sub volume is extracted. Hidden Markovian model is used to train on the spatio-temporal motion pattern. From the model the spatio-temporal motion pattern that describes how the object moves in a video is obtained. The extracted information is used as the priori for tracking.

Keywords— Bayseian, HMM, spatio-temporal pattern, Tracking, Video object

I. INTRODUCTION

Video content analysis is the process of analysing the video to determine the events. It's used in the domains like human-computer interaction, security, surveillance, video communication and compression, traffic control, medical imaging and video editing. The functionality of video analysis is video tracking, identification and behavior analysis and egomotion estimation. This model focuses on the video tracking. Video tracking is the process of determining the location of the object in the video signal. Video contains large amount of data so the tracking can be a time consuming process. The video tracking is to associate target objects in consecutive video frames. The video tracking algorithm analyses the sequential video frames and outputs the motion of the object between the frames. There are two major components for visual tracking. The components are target representation and localization and filtering and data association. There are numerous researches for video-based object extraction and tracking. One of the simplest methods is to track regions of difference between a pair of consecutive frames [24], and its performance can be improved by using adaptive background generation and subtraction. The difference-based tracking method is efficient in tracking an object under noise-free circumstances; it often fails under noisy and complicated background. The tracking performance degrades if a camera moves either intentionally or unintentionally. Tracking of objects in the presence of shadows, noise and occlusion, a non-linear object feature voting scheme has been proposed in [25]. As an alternative method of the difference-based tracking, a blob tracking method using simple geometric models, e.g., ellipse or rectangle, can track the centroid of an object. Based on the assumption of stationary background, Wren et al. proposed a real-time blob tracking algorithm [26]. For more robust analysis of an object, shape-based object tracking algorithms have been developed, which utilize a priori shape information of an object-of interest, and project a trained shape onto the closest shape in a certain frame. This type of methods includes Active Contour Model (ACM), Active Shape Model (ASM), and the Condensation algorithm [10]. Although the existing shape-based tracking algorithms can commonly deal with partial occlusion, exhibit two major problems in the practical applications, such as: a priori training of the shape of a target object and iterative modelling procedure for convergence. Selected object tracking algorithms of significant interests are summarized in Table 1.

TABLE 1

Properties of various object tracking algorithms

Algorithm	**Tracking entity**	**Occlusion Handling**	**Specific Task**
W4 [25]	Silhouettes of people	Appearance model	Surveillance of people
Amer [26]	Shape, Size, Motion	Non-linear voting	Surveillance, retrieval
Wren [27]	Human body model	Multi-class Statistical model	Tracking Human
Isard [10]	Shape space	State Space Sampling	Tracking objects in the shape space
Yilmaz [28]	Contour	Shape prior	Tracking level sets
Rodriguez [20]	Human body model	Unstructured crowd	Tracking Human
Kratz [14]	Motion	Motion pattern prior	Tracking Human

Object Tracking

The process of object tracking is summarized in the block diagram below:

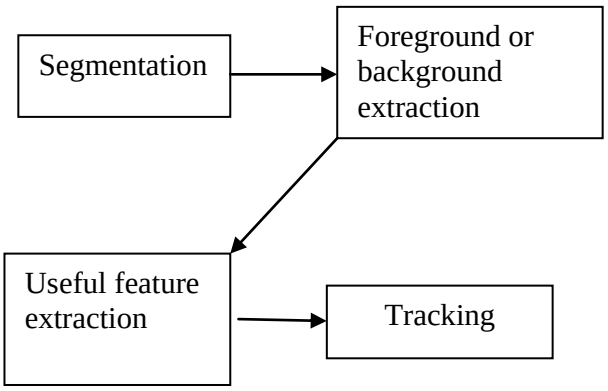


Figure 1 Tracking of Objects

The basic steps in object tracking are:

1. Segmentation
2. Foreground or background extraction
3. Camera modeling
4. Feature extraction and tracking

Segmentation is the process of identifying components of the image. Segmentation involves operations such as boundary detection, connected component labeling and thresholding. Boundary detection finds out edges in the image, any differential operator can be used for boundary detection. Thresholding is the process of reducing the grey levels in the image. Foreground extraction is the process of separating the foreground and background of the image. This process is used for subtraction of images in order to find objects that are moving and those that are not. Another method that can be used in object tracking is Background learning. This approach can be used if fixed cameras are used for video capturing. In this method, an initial training step is carried out before deploying the system. In the training, the system constantly records the background in order to learn it. Once the training is completed the system has complete information about the background. Once the background is known, extracting the foreground is a simple image subtraction. The next step is to extract useful features from the sequence of frames.

The goal of the object tracking is to estimate location and motion parameters of an object in a image sequences. The objective function of tracking depends on distance, similarity or classification measure. The tracking results are often obtained by minimizing or maximizing an objective function.

The size of the cuboids remains same and selected manually due to loss of pixels limit the accuracy of the tracking. The video are divided into spatio-temporal sub-volumes or cuboids defined by a regular grid and compute the local spatio-temporal motion pattern within each. Hidden Markov Model [19] is trained on the local spatio-temporal motion patterns at each spatial location. The spatial location of the training video represents the spatially and temporally varying crowd motion. Using the HMM and previously observed frames of a separate tracking video of the same scene, the local spatio-temporal motion pattern that describes a target moves through the video is predicted. The predicted local spatio-temporal motion pattern is used to hypothesize a set of priors on the motion and appearance variation of individuals that wish to track.

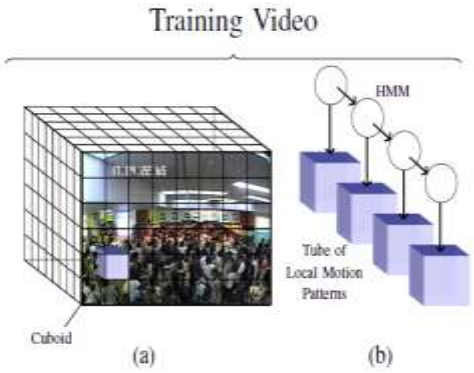


Figure 2 Cuboids

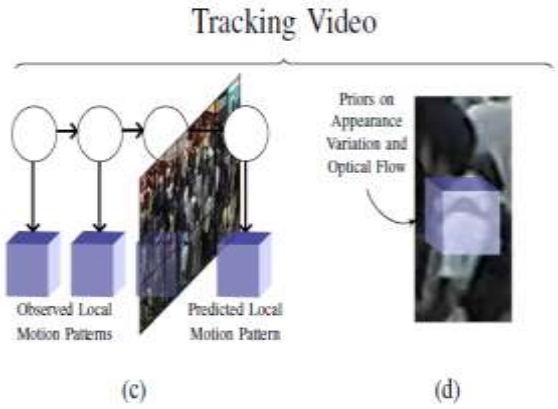


Figure 3 Pattern Prediction

Divide the video into spatio-temporal sub-volumes or cuboids. (a) The local spatio-temporal motion pattern within each cuboid is computed. Train a hidden Markov model (b) on the local spatio-temporal motion patterns at each spatial location. Using the HMM and previously observed frames of a separate tracking video of the same scene, (c) predict the local spatio-temporal motion pattern that describes how a target moves through the video. The predicted local spatio-temporal motion pattern is used to hypothesize a set of priors (d) on the motion and appearance variation of individuals that we wish to track [16].

II. Related Works

Since the literature on tracking is extensive, the work that model motion in cluttered or crowded scenes is reviewed.

Isard et al. [10] developed the condensation based algorithm to track the visual clutter it represents the multi modal distribution. The algorithm uses stochastic framework and track the outline and features of the object. The approach aims at using the probabilistic model of object shape and motion to analyze the video streams. The algorithm is used for both rigid and non-rigid motion.

Black et al. [7] proposed a bayesian framework for representing and recognizing local image motion in terms of two basic models like translational motion and motion boundaries. The method move towards a richer description of image motion using a vocabulary motion primitive. A step in that direction is described with the introduction of an explicit non-linear model of motion boundaries and a Bayesian framework for representing a posterior probability distribution over models and

model parameters. A maximum estimate of image motion is calculated using the probability distribution over the parameter space of discrete samples. The image motion facilitates the correct Bayesian propagation of information over time and its ambiguities make the distribution non-Gaussian. The open issue is that sampling methods has high dimensional space.

Betke et al [5] Proposed statistical data association techniques for visual tracking of enormously large numbers of objects. This approach combines the techniques of multi target track initiation, recursive Bayesian tracking, clutter modeling, event analysis, and multiple hypotheses. Okabe et al proposed a method which tracks features and associate similar trajectories to detect individual moving entities within crowded scenes. Technique assumes that the subjects move in distinct directions and thus disregard possible local motion inconsistencies between different body parts.

Wright et al. [23] proposed a method for analysis of motion patterns. The system uses static and quasi-static backgrounds. This background model produces a crude initial segmentation that is processed by code specific to find and recognizing humans. Ali and Shah and Rodriguez et al. model the motion of individuals across the frame in order to track pedestrians in crowds captured at a distance.

Pless et al. Learn a single motion distribution at each frame location from videos of automobile traffic. These approaches impose a fixed number of possible motions at each spatial location in the frame. In extremely crowded scenes, Pedestrians in the same area of the scene may move in any number of different directions. Motions are encoded in HMM and derive a full distribution of the motion at each spatio-temporal location in the video. In addition, natural body movements appear subtle when captured at a distance but create large appearance changes in near-view scenes.

Nestares and Fleet [18] also use neighboring motion patterns to improve tracking. It increases the continuity of the motion boundary tracking from Black and Fleet by including multiple image neighborhoods. In our work, Model uses a dynamic temporal model of sequential motion patterns rather than assuming continuity across spatial locations.

III. Proposed Scheme

Step 1:

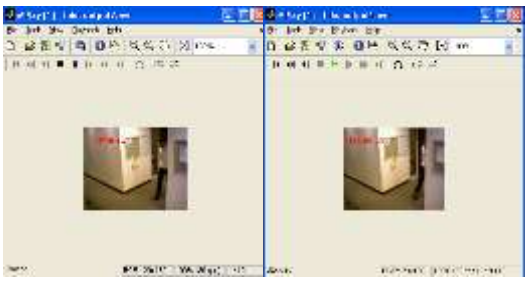
The video is taken as the input.

Figure 4 Videos used for tracking

The videos used are man in the room, foot ball ground and street crossing. The videos are given in the Figure 4

Step 2:

The video is divided into cuboids. Cuboids size should be selected in way that best represents the characteristic movements of a pedestrian. The given video is split into equal size cuboids. The simulation model for cuboid division is given in the Figure 5.



(a) Frame 1 **(b) Frame 19**

Figure 5 Cuboids

Step 3:

For each pixel 'i' in cuboid Spatio-temporal gradient ∇i is calculated using $i = [I_{i,x}, I_{i,y}, I_{i,t}]^T = \left[\frac{\partial I}{\partial x}\ \frac{\partial I}{\partial y}\ \frac{\partial I}{\partial t}\right]^T$ Background estimation is done to track the object and shown in the Figure.

Figure 6 Background Extraction

Step 4:

The enhanced k-means algorithm is used to calculate the centroids.

Algorithm : Finding the initial centroids

Input:

D = {d1, d2,.....,dn} // set of *n* data items

k // Number of desired clusters

Output: A set of *k* initial centroids

Steps:

1. Set m = 1;
2. Compute the distance between each data point and other data points in the set D;
3. Find closest pair of data points from set D and form a data point set Am (1<= m <= k) which contains these two data- points, Delete these two data points from the set D;
4. Find the data point in D that is closest to the datapoint set Am, Add it to Am and delete it from D;
5. Repeat step 4 until number of data points in Am reaches 0.75*(n/k);
6. If m<k, then m = m+1, find another pair of data points from D between which the distance is the shortest, form another data-point set Am and delete them from D, Go to step 4;
7. For each data-point set Am (1<=m<=k) find the arithmetic mean of the vectors of data points in Am, means will be the initial centroids. The clustering of the cuboids is done to

classify the objects. The sample data is generated testing with 4 clusters in 3 dimensions, by generating random data with gaussian density, variance 1, with means (0,0,0), (0,0,6), (0,6,0) and (6,0,0) and Ndata 200, 300, 100 and 500.

Step 5:

KL divergence distance is used to determine the distance between sub volumes and the centroids. The distance used is 0.5, 1, 1.5 and 2.

Step 6:

The hidden states of HMM are represented using the probability of the observed motion pattern O_t^n . The accuracy of predictions depends heavily on the number of hidden states in each HMM. The online clustering algorithm with a distance threshold d_{KL} is used to vary the number of hidden states depending on the local motion patterns within the video.

Step 7:

The hidden states are obtained. The graph given below shows the variation in the hidden states. The hidden states get stabilized after the distance of 3. The collection of HMMs is trained on a video of each scene, and uses it to track pedestrians in videos of the same scene recorded at a different time. The training videos for each scene have different frames. The size of cuboids is used to represent the local spatio temporal motion pattern. The accuracy of our predictions depends heavily on the number of hidden states in each HMM. The clustering algorithm uses a distance threshold dKL to vary the number of hidden states depending on the local motion patterns in the video. Large variations in flow may result in an excessive number of hidden states.

Step 8:

Predict the motion pattern for each space and location of the each cuboid is given using ($\widetilde{\mu_t^n}$ and $\widetilde{\Sigma_t^n}$).

Step 9:

The Bayesian framework is formulated using the

$$p(X_t|Z_{1:t}) \propto p(Z_t|X_t) \int p(X_t|X_{t-1}) p(X_{t-1}|Z_{1:t-1}) dX_{t-1}$$

Step 10:

The state transition distribution is obtained using the optical flow vector and co- variance matrix. The optical flow is the structure tensor's eigenvector with the smallest eigen value.

Step 11:

Co-variance matrix can be estimated using

$$\Delta = [v_1'.v_2'] \begin{bmatrix} \frac{\lambda_3}{\lambda_1} & 0 \\ 0 & \frac{\lambda_3}{\lambda_2} \end{bmatrix} [v_1'.v_2']^{-1}$$

where v_1' and v_2' are the projections of v_1 and v_2 onto the plane t = 1.

Step 12:

The likelihood p ($Z_t|X_t$) is computed using Equation

$$p(Z_t|X_t) = \frac{1}{Z} exp \frac{-d(R.T)}{\sigma}$$

Step 13:

Assuming pedestrians exhibit consistency in their appearance and motion. Model them in a joint likelihood by $p(Z_t|X_t) = p_A(Z_t|X_t) p_M(Z_t|X_t)$ where p_A and p_M are appearance and motion likelihoods.

Step 14:

The motion template is obtained using Equation

$$T_{M,i}^{t} = \alpha \nabla R_{E[x_t|z_{1:t}],i} + (1-\alpha) T_{M,i}^{t-1} \quad (4.3)$$

Step 15:

The pedestrian's motion changes gradually, this error measurement during tracking can be calculated using

$$E_i^t = \alpha \arccos(t_i . t_r) + (1-\alpha) E_i^{t-1}$$

The Figure 7 shows the average magnitude of the error vector using our approach, our approach using all cuboids the target spans and using only an optical-flow. No discernible benefit is achieved by using all of the cuboids. HMMs trained on the optical flow vectors do not retain the rich distribution of flow within the local spatio-temporal motion patterns, and result in high errors.

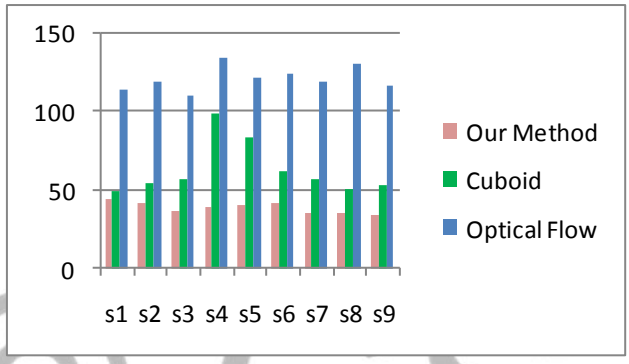


Figure 7 Average Magnitude of Error

The videos used are Football ground, person working in the room and the street. Angular error is the difference between the observed flow vector and predicted flow vector.

IV. RESULT

The video with unstructured crowed are taken. For each cuboid the motion patterns are observed. The video to be tracked is learned and the priors are learned. Hidden Markov models are trained on the video. The KL Divergence distance used is 0.5 and 2. The number of hidden states increases with decrease in the distance. The graph given below shows the variation in the hidden states. The number of states gets stabilized for the distance above 3. So, the distance used here is 0.5, 1, 1.5 and 2. The collection of HMMs is trained on a video of each scene, and uses it to track pedestrians in videos of the same scene recorded at a different time. The training videos for each scene have different frames. The size of cuboids is used to represent the local spatio temporal motion pattern. The accuracy of our predictions depends heavily on the number of hidden states in each HMM. The clustering algorithm uses a distance threshold dKL to vary the number of hidden states depending on the local motion patterns in the video. Large variations in flow may result in an excessive number of hidden states.

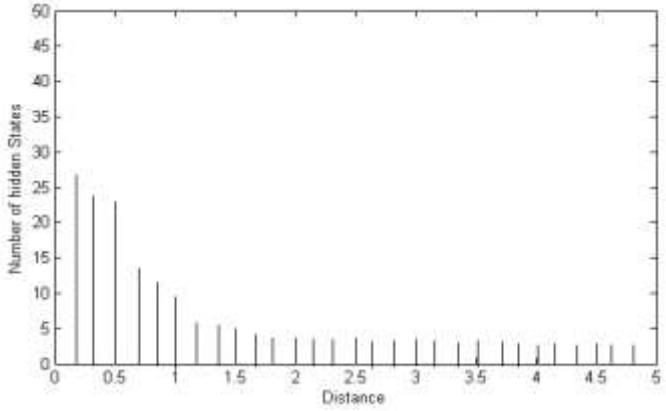


Figure 8 Effects of distance threshold

V. CONCLUSION

A probabilistic method that exploits the inherent spatially and temporally varying structured pattern of a crowd's motion to track individuals in extremely crowded scenes are derived. The input video is divided into equal size cuboids form which Hidden Markovian Model training can be done. Using a collection of Hidden Markovian Model that encode the spatial and temporal variations of local spatio-temporal motion patterns, the proposed method successfully predicts the motion patterns within the video. The predicted video is used to track the object present in the tracking video. The results show that leveraging the steady-state motion of the crowd provides superior tracking results in extremely crowded areas. Further, the algorithm can be extended for the automatic detection of cuboids size. The cuboids sizes may be determined by a semi-supervised approach that approximates the perspective projection of the scene. Varying cuboid size can also be used. Space-time model may be further leveraged to provide robustness severe occlusions.

REFERENCES

[1] S. Ali and M. Shah. A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis. In Proc. of IEEE Int'l Conf on Computer Vision and Pattern Recognition, pages 1–6, 2007.

[2] S. Ali and M. Shah. Floor Fields for Tracking in High Density Crowd Scenes. In Proc. of European Conf on Computer Vision, 2008.

[3] S. M. Arulampalam, S. Maskell, and N. Gordon. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. IEEE Transactions on Signal Processing, 50:174–188, 2002.

[4] J. Barron, D. Fleet, and S. Beauchemin. Performance of Optical Flow Techniques. Int'l Journal on Computer Vision, 12(1):43–77, 1994.

[5] M. Betke, D.E. Hirsh, A. Bagchi, N.I. Hristov, N.C. Makris, and T.H. Kunz. Tracking Large Variable Numbers of Objects in Clutter. In Proc. of IEEE Int'l Conf on Computer Vision and Pattern Recognition, pages 1–8, 2007.

[6] C. M. Bishop. Pattern Recognition and Machine Learning. Springer, October 2007.

[7] M. J. Black and D. J. Fleet. Probabilistic Detection and Tracking of Motion Boundaries. Int'l Journal on Computer Vision, 38(3):231–245, July 2000.

[8] G.J. Brostow and R. Cipolla. Unsupervised Bayesian Detection of Independent Motion in Crowds. In Proc. of IEEE Int'l Conf on Computer Vision and Pattern Recognition, pages 594–601, June 2006.

[9] C. Hue, J.-P. Le Cadre, and P. Perez. Posterior Cramer-Rao Bounds for Multi-Target Tracking. Aerospace and Electronic Systems, IEEE Transactions on, 42(1):37 – 49, Jan. 2006.

[10] M. Isard and A. Blake. CONDENSATION-Conditional Density Propagation for Visual Tracking. Int'l Journal on Computer Vision, 29(1):5–28, August 1998.

[11] Z. Khan, T. Balch, and F. Dellaert. MCMC-Based Particle Filtering for Tracking a Variable Number of Interacting Targets. IEEE Trans. on Pattern Analysis and Machine Intelligence, 27(11):1805–1819, Nov. 2005.

[12] Z. Khan, T. Balch, and F. Dellaert. MCMC Data Association and Sparse Factorization Updating for Real Time Multitarget Tracking with Merged and Multiple Measurements. IEEE Trans. on Pattern Analysis and Machine Intelligence, 28(12):1960–1972, Oct. 2006.

[13] L. Kratz and K. Nishino. Anomaly Detection in Extremely Crowded Scenes Using Spatio-Temporal Motion Pattern Models. In Proc. of IEEE Int'l Conf on Computer Vision and Pattern Recognition, pages 1446–1453, 2009.

[14] L. Kratz and K. Nishino. Tracking With Local Spatio-Temporal Motion Patterns in Extremely Crowded Scenes. In Proc. of IEEE Int'l Conf on Computer Vision and Pattern Recognition, 2010.

[15] S. Kullback and R. A. Leibler. On Information and Sufficiency. The Annals of Mathematical Statistics, 22(1):79–86, 1951.

[16] B. Leibe, E. Seemann, , and B. Schiele. Pedestrian Detection in Crowded Scenes. In Proc. of IEEE Int'l Conf on Computer Vision and Pattern Recognition, June 2005.

[17] Y. Li, C. Huang, and R Nevatia. Learning to Associate: HybridBoosted Multi-Target Tracker for Crowded Scene. In Proc. of IEEE Int'l Conf on Computer Vision and Pattern Recognition, 2009.

[18] O. Nestares and D. J. Fleet. Probabilistic Tracking of Motion Boundaries with Spatiotemporal Predictions. In Proc. Of IEEE Int'l Conf on Computer Vision and Pattern Recognition, pages 358–365, 2001.

[19] L. Rabiner. A Tutorial On Hidden Markov Models and Selected Applications in Speech Recognition. Proc. of the IEEE, 77(2):257–286, Feb. 1989.

[20] M. Rodriguez, S. Ali, and T. Kanade. Tracking in Unstructured Crowded Scenes. In Proc. of IEEE Int'l Conf on Computer Vision, 2009.

[21] E. Shechtman and M. Irani. Space-Time Behavior Based Correlation. In Proc. of IEEE Int'l Conf on Computer Vision and Pattern Recognition, pages 405–412, 2005.

[22] J. Wright and R. Pless. Analysis of Persistent Motion Patterns Using the 3D Structure Tensor. In IEEE Workshop on Motion and Video Computing, pages 14–19, 2005.

[23] B. Wu and R. Nevatia. Tracking of Multiple, Partially Occluded Humans Based On Static Body Part Detection. In Proc. of IEEE Int'l Conf on Computer Vision and Pattern Recognition, pages 951–958, 2006.

[24] T. Zhao, R. Nevatia, and B. Wu. Segmentation and Tracking of Multiple Humans in Crowded Environments. IEEE Trans. on Pattern Analysis and Machine Intelligence, 30(7):1198–1212, 2008.

[25] Haritaog lu I, Harwood D, Davis L. W-4: Real-time surveillance of people and their activities. IEEE Transactions on Pattern Analysis and Machine Intelligence 2000;22(8):809–30.

[26]Amer A. Voting-based simultaneous tracking of multiple video objects. Proceedings of the SPIE Visual Communication Image Processing 2003;5022:500–11.

[27] Wren C, Azerbayejani A, Darrel T, Pentland A. Pfinder: Realtime tracking of the human body. IEEE Transactions on Pattern Analysis and Machine Intelligence 1997;19(7):780–5.

[28] Yilmaz A, Shah M. Contour-based object tracking with occlusion handling in video acquired using mobile cameras. IEEE Transactions on Pattern Analysis and Machine Intelligence 2004;26(11):1531–6.
[29] T. Zhao, R. Nevatia, and B. Wu. Segmentation and Tracking of Multiple Humans in Crowded Environments. IEEE Trans. on Pattern Analysis and Machine Intelligence, 30(7):1198–1212, 2008.