



UvA-DARE (Digital Academic Repository)

Tracking individuals in surveillance video of a high-density crowd

Hu, N.; Bouma, H.; Worring, M.

Published in:

Proceedings of SPIE, the International Society for Optical Engineering

DOI:

[10.1117/12.979489](https://doi.org/10.1117/12.979489)

[Link to publication](#)

Citation for published version (APA):

Hu, N., Bouma, H., & Worring, M. (2012). Tracking individuals in surveillance video of a high-density crowd. Proceedings of SPIE, the International Society for Optical Engineering, 8399, 8399-09. <https://doi.org/10.1117/12.979489>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <http://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Tracking individuals in surveillance video of a high-density crowd

Ninghang Hu^{a,b}, Henri Bouma^{a,*}, Marcel Worring^b

^a TNO, P.O. Box 96864, 2509 JG The Hague, The Netherlands;

^b University of Amsterdam, P.O. Box 94323, 1098 GH Amsterdam, The Netherlands

ABSTRACT

Video cameras are widely used for monitoring public areas, such as train stations, airports and shopping centers. When crowds are dense, automatically tracking individuals becomes a challenging task. We propose a new tracker which employs a particle filter tracking framework, where the state transition model is estimated by an optical-flow algorithm. In this way, the state transition model directly uses the motion dynamics across the scene, which is better than the traditional way of a pre-defined dynamic model. Our result shows that the proposed tracker performs better on different tracking challenges compared with the state-of-the-art trackers, while also improving on the quality of the result.

Keywords: Security, tracking, surveillance, image processing, crowd.

1. INTRODUCTION

Video cameras are widely used in surveillance applications to monitor public areas, such as train stations, airports and shopping centers. When crowds are dense, automatically tracking individuals becomes a challenging task. In this paper we propose a new tracking technique to meet these challenges. Our tracker employs a particle filter tracking framework. Instead of using a fixed pre-defined state transition model, we employ the optical-flow algorithm to estimate state transition. Since optical-flow vectors are observed cues from the scene, they are more accurate than the fixed models. In our method, the optical flow vectors are measured over the scene and quantized within local spatial-temporal regions. Optical flow vectors falling in a region are modeled and the result is fed to the state transition model of particles within the same region. Some recent works require a long training period. But training over a long duration of video may result in a less relevant model for tracking, especially for abnormal motions. In our approach, no training process is required. The local region of flow vectors is directly employed for state transition and it only requires a short part of the video prior to the frame under consideration. To test the robustness of our tracker, in this paper, the performance is analyzed over separate tracking challenges, such as ambiguous appearance, abnormal pedestrian behaviors, partial occlusion, and different density of crowds. Our results show that the proposed tracker performs better on these challenges and the performance is largely improved compared with the state-of-the-art trackers.

2. RELATED WORKS

The definition of crowd varies largely over the literature. We define four types of crowds. (1) *Sparse crowd*: only a few pedestrians are observed in the scene. The pedestrians are definitely not gathered closely together and less than 50% of them are occluded [18][33]. (2) *Moderate crowd*: pedestrians cover over 50% area of the scene, and about 50% - 80% of them are occluded [7][26]. (3) *High density crowd*: pedestrians are observed in the whole area of the scene. Since they are in small proximity, the movements of pedestrians in the scene are physically constrained by the others nearby. 80% - 100% of the people in the scene are occluded [15]. (4) *Extremely dense crowd*: contains an extremely large number of people, and the resolution on each person is extremely low. All people are occluded, and only the head and shoulders of the target can be observed. 100% of the pedestrians in the scene are occluded [24].

To scope our target, in this paper, we focus on tracking in high density crowds, where a large number of people gather closely together, but the upper body of the pedestrians is still visible. Next, we will review different approaches of tracking in the crowded scenes.

* henri.bouma@tno.nl; phone +31 88 866 40 54; <http://www.tno.nl>

Ninghang Hu, Henri Bouma and Marcel Worring, "Tracking individuals in surveillance video of a high-density crowd", Proc. SPIE, Vol. 8399, 839909 (2012); <http://dx.doi.org/10.1117/12.918604>

Copyright 2012 Society of Photo-Optical Instrumentation Engineers (SPIE). One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

2.1 Tracking with frame-by-frame human detection

In a scene with a sparse or moderate crowd, most pedestrians can be fully observed and the pedestrians can be detected with a frame-based human detector. Persons are then tracked by combining detections into tracklets and associating the tracklets into long trajectories [3][13]. Multiple persons are usually tracked at the same time to make the tracker more robust against occlusions. To find the optimal association among multiple tracklets, the Data Association based Tracking (DAT) algorithm [11][17][32] is applied. The DAT performs well in solving the ambiguity problem and reducing the rate of ID switches [19]. The association costs are measured by a weighted sum of different cues, such as appearance (color histogram), motion, and frame gap between two tracklets. The basic approach was extended by using shape and appearance models [33], body-part detectors [29], or a boosting algorithm to train the parameters [19]. However, in a denser crowd where the pedestrians are heavily occluded by the others, frame-based detection is highly unstable, and associating among a large number of trajectories is very expensive.

2.2 Tracking with local feature points

Since the local feature points are large in quantity and in general they are unlikely to all be affected by occlusions, the target can be tracked by associating local feature points. Brostow e.a. [7] were among the first to do this. They assume that the feature points belonging to the same person are close in space and their motion exhibit high correlation over time. A similar approach is proposed by Li et al. [18]. Sugimura et al. [26] make an extension by employing gait features to separate pedestrians that are close in space. In a dense crowd, the gait features are not noticeable. In high density crowds, the local feature points are frequently occluded which results in a huge number of short tracking fragments. Associating these fragments with DAT can be very challenging and time consuming. The other limitation of such an approach is that targets moving together with the same speed cannot be identified as separate persons, whereas a local body movement is usually wrongly estimated as a different target.

2.3 Tracking with optical flow

Optical flow algorithms have been widely used for tracking pedestrians. Initial approaches assume that optical flow on the target is uniform, and the target is tracked by computing the mean flow around the target location [27][30]. Denman et al. [8] extended this approach by using foreground-background segmentation to get a precise target region. Such an approach is hardly possible in crowded scenes. To deal with the high ambiguity of tracking in the frames, some approaches keep multiple hypotheses for the location of the target, and they integrate optical flow algorithms in a particle filter tracking framework [12]. This framework consists of two components, measuring likelihood and particle propagation. Some papers [16][20][21] proposed to improve the likelihood measure by building a motion template of the target with flow vectors, while they choose a fixed dynamic model for particle propagation. The fixed model makes the simplifying assumption that targets move with a constant speed in consecutive frames. However, the simple pre-defined model does not meet with the requirements of tracking in crowded scenes. In crowded scenes, targets are frequently occluded and their region changes over time. As a result, the target location does not change linearly. This requires a propagation model that adaptively changes over time and space with non-linear behavior. Rather than building a dynamic model, Rodriguez [24] and Kratz [15] incorporate optical flow algorithms in the particle propagation.

2.4 Tracking with the motion of local areas

In the high density and the extremely dense crowd situations, the most promising tracking algorithms use motion information in local areas. Ali and Shah [1] assume that pedestrians in the crowd behave in a similar way as particles in the flow. Their application is limited to tracking pedestrians that move in a similar direction as the crowd. Rodriguez et al. [24] solve the tracking problem as in topic retrieval. They first divide the video into short clips and for each local area in the clip, flow vectors are quantized into four categories based on the direction where vectors are heading. The quantized clips are trained within the Correlated Topic Model (CTM), generating a set of topics. For each new frame, a probability distribution over the topics is measured, and the probability of the motion is then derived. The new target position is estimated as a combination of the observation and the tracker prediction. Similar to Rodriguez, Kratz e.a. [15] also divide the video into spatial-temporal areas and model the motions in each local space. But instead of coarsely quantizing the motion vectors into four directions, they model the motion vectors of a local area with a Gaussian density distribution. The temporal variation of the motions is then learned by training a Hidden Markov Model (HMM) at each spatial area. With the HMMs, motions in the frame under consideration can be predicted. These approaches incorporate a training process that extracts motion statistics from the scene. For robust training, a long duration of the training video is acquired. However, when the duration of training video increases the frames are less relevant for tracking.

In this paper, we propose a novel approach that tracks pedestrians with the previous observation while disregarding the other previous frames. We model the local motions in the similar way as Kratz and Nishino [15], but instead of learning the model with a long sequence of training data, we apply the local motions directly in particle propagation. To the best of our knowledge, such a method has not been proposed before. Before we describe our own approach, we study the particle filter in more detail.

3. COLOR-BASED PARTICLE FILTER TRACKING

In a cluttered scene with a high density crowd, the heavy occlusions and the high ambiguity call for maintaining multiple hypotheses while tracking. The particle filter [12] provides a robust way of solving such problems. In this section, we will first introduce the framework for the color-based particle filter tracker. Then we will focus on measuring the two major components of the framework, the likelihood and the state transition model.

3.1 Particle filter tracking framework

Our goal is to track individuals in a high density crowd, i.e. finding the most probable location $S^t(x^t, y^t)$ of the person given a sequence of observed frames $[O^1, O^2, \dots, O^t]$. Typically, the particle filter makes use of a recursive Bayesian framework [12][15]:

$$P(S^t | O^{1:t}) \propto P(O^t | S^t) \int P(S^t | S^{t-1}) P(S^{t-1} | O^{1:t-1}) d(S^{t-1}) \quad (1)$$

where $P(S^t/O^{1:t})$ is the posterior probability, $P(O^t/S^t)$ is the likelihood of the target and $P(S^t/S^{t-1})$ is the transition model from time $t-1$ to time t . The Bayesian framework follows a first-order Markov process, which means the current hypothetical location of the pedestrian only depends on the previous state and all the other past states or observations are considered irrelevant. In the particle filter, the likelihood and the state transition model are re-measured every iteration. In the following two subsections, we introduce the measurement of the two components separately.

3.2 Likelihood

The likelihood $P(O^t/S^t)$ refers to the likelihood of the observation O^t being the same as the target that we are tracking. The likelihood is typically measured inversely proportional to the (e.g., Bhattacharyya) distance between two color histograms. RGB or HSV are two commonly used color spaces for the histograms, the first of which is sensitive to changes in illumination and the latter may become unstable for colors with low value or saturation.

3.3 State transition model

The state transition model $P(S^t/S^{t-1})$ is the other component in the particle filter, and it determines how the particles are propagated in each iteration. A traditional state transition model assumes that the pedestrian moves with a constant speed in consecutive frames $(S_n^{t-1})' = (S_n^t)'$. With such assumption, we can update the state S_n^t of the particle n by:

$$S_n^t = S_n^{t-1} + \delta(S_n^{t-1})' + \sigma_n^{t-1} \quad (2)$$

where δ is a constant for weighting, $(S_n^{t-1})'$ is the speed of the target at time $t-1$, and σ_n^{t-1} is the Gaussian random variable.

With the constant-velocity assumption, the appearance likelihood of the target is the only observed cue, thus the tracking is highly dependent on the performance of the likelihood measurement. Due to the high ambiguity of the scene, measuring likelihood from the appearance is very hard. To make a more robust tracker for crowded scenes, we introduce a novel approach, which uses the previous observation as the state transition model in the particle filter.

4. TRACKING WITH PREVIOUS OBSERVATION

In this section, we present a novel approach that employs the previous observation as the transition model in the particle filter.

4.1 Capturing motion with optical flow

The first step of our approach is to extract the motion by computing optical flow vectors. The optical flow algorithm measures the shift of pixels between two consecutive frames.

4.2 Modeling observation in local areas

To extract the motion knowledge locally, the volume of motion vectors is subdivided into local spatial-temporal areas, forming a set of cuboids with motion statistics. Kratz [15] proposed to compute the 3D gradients of the intensity in the local area and model them with 3D Gaussians. In our approach, however, we use the optical flow vectors directly, since we consider the flow vectors to be more stable than the 3D gradients. As the cuboid contains flow vectors within a short time span and a similarly small spatial area, we assume that the density distribution of the vectors follows a 2D Gaussian distribution.

4.3 Embedding previous observation for tracking

In the context of particle-filter tracking, the propagation is driven by a state transition model, which accounts for the particles' displacement in the frames. In our paper, the state transition model is approximated by the previous observation captured in the same spatial location as the particle.

Assuming a particle is located in a cuboid defined by spatial region R and temporal duration T , and the particle is to be propagated from frame $t-1$ to frame t . The state transition model is approximated by the previous observation O_{R}^{T-1} . Formally, the particle state in the next frame is propagated by:

$$S_n^t = S_n^{t-1} + O_R^{T-1} = S_n^{t-1} + N(\mu, \Sigma) \quad (3)$$

By employing the previous observation in particle propagation, the two observed cues of tracking, i.e. the motion and the appearance of the target, are fused properly in the particle filter. In such a way, targets can be tracked with both observed cues at the same time and the ambiguity of the appearance can be reduced by adding the motions.

5. EXPERIMENTS AND RESULTS

5.1 Dataset

The approach was evaluated on two crowded scenes (Figure 1). The first scene contains a high density crowd in the music festival Love Parade (loveparade.de). The second scene is recorded at the train station of Amsterdam. For the first scene, we evaluated our system with two separate sequences of 300 frames (approximately 10 seconds). The dimension of a single frame is 1280x720 pixels and in our experiment we only used a sub-region of 350x1030 pixels (average size of the pedestrians is 19x29) which was relevant for tracking. To analyze the tracking performance, we manually selected 10 normal pedestrians (moving with the crowd) and 3 abnormal pedestrians (not moving with the crowd) for evaluation. For the second scene, a sequence of 300 frames was used with the dimension of 480x640 pixels (average pedestrian is 25x56). In this scene 10 pedestrians were selected as ground-truth targets.

In the following, we first introduce the evaluation criteria used in the paper. Then we discuss the implementation of the different tracking approaches separately. Finally, we evaluate the tracking algorithms and compare the tracking results in both scenes.



Figure 1. Two frames from the sequences with the music festival and the train station.

5.2 Evaluation criteria

The related works [1][15][24] use the average distance error as the evaluation criteria. The drawback of such measurement is that it does not reflect how well the tracker can focus on the target. In this paper, we decompose the evaluation into two parts. First, we determine whether targets are tracked correctly based on overlap, which is our primary criterion. Only when the target is tracked correctly, we measure the average distance error of the track as a measure of accuracy.

5.3 Tracking with color-based particle filter

In this paper, we employ as a baseline the traditional color-based particle filter [23], which assumes the target moves with a constant speed in consecutive frames. Based on experiments [10], we chose the parameter to weight the target's speed in the previous frame to be $\delta=1$ and the variance of the random component.

5.4 Tracking with motion patterns

We also implemented the state-of-the-art approach by Kratz and Nishino [15] for comparison. We follow their approach to model the local motions with a Gaussian density distribution in each spatial-temporal area. The temporal motion dynamics are learned by training a Hidden Markov Model (HMM). In the HMM, the states are generated by applying an on-line clustering algorithm over the Gaussian density distributions, where the Kullback-Leibler (KL) divergence is applied as the distance measurement between the cluster and the Gaussians. Here the states are considered as motion patterns [14]. Since they also employ the particle filter as their tracking framework, in the following, we denote their approach as PF+MP+HMM for the convenience of illustration. With a small number of frames, the HMMs cannot be trained sufficiently, while with a large number of frames, it becomes less relevant for tracking. Based on experiments [10] a duration of 300 frames was chosen.

5.5 Tracking with the previous observation.

Instead of modeling the HMMs, our proposed tracker, PF+PO, uses the previous observation directly as the state transition model in the particle filter. The PO is a 2-dimensional Gaussian density distribution, compactly representing the optical flow vectors in a local temporal-spatial area. The optical-flow vectors are computed with the Gunnar Farneback's algorithm [9] in OpenCV [6]. A set of trajectories generated by our proposed tracking algorithm is shown in Figure 13. The proposed tracker is evaluated with varying durations of the previous observation. The PF+PO tracker appears to perform optimally when the PO contains between 1 to 10 frames [10] and we use 10 as the optimal duration for comparison.

5.6 Evaluation and comparison

To compare the baseline particle filter (PF), state-of-the-art motion patterns (PF+MP+HMM) and the novel previous-observation (PF+PO), we applied each approach to the two scenes. To reduce the randomness, each pedestrian is tracked for 10 iterations. A few examples are shown in Figure 2.



Figure 2. By exploiting the previous observation in the local area, most of the pedestrians are tracked as they moved through a high density crowd.

Figure 3 compares the tracking results over time. The result shows that PF+PO is able to track the pedestrians for longer time than the others. We observed that the tracking result of PF is highly dependent on the appearances of the pedestrians. Figure 4 shows that the bad performing targets of PF, which exhibit either low saturation or low value in the HSV space, fail due to an unstable hue component.

Table 1 shows the quantitative results in both scenes. Benefiting from the constant speed assumption, the baseline tracker PF achieves the best performance when tracking with full occlusions (especially in the first scene). By modeling the motions with HMM, the PF+MP+HMM tracker scores higher for partial occlusions and normal pedestrians than PF in the first scene, but still lower than PF+PO because the motion statistics in the HMM are more relevant for the training video than the frame of tracking. When tracking the abnormal pedestrians, the performance of PF+MP+HMM is even worse than our baseline tracker. This is because the abnormal motion is not modeled in the HMM, thus the HMM cannot assign good predictions to the abnormal pedestrians. In the second scene, the PF+MP+HMM tracker completely fails, because the tracking algorithm cannot handle the areas with hardly any motion. Our proposed tracker PF+PO scores the highest on partial occlusions, abnormal behaviors, normal behaviors, and on average. Particularly, when tracking the abnormal pedestrians in the first scene, the performance is improved by 30% compared with the other trackers.

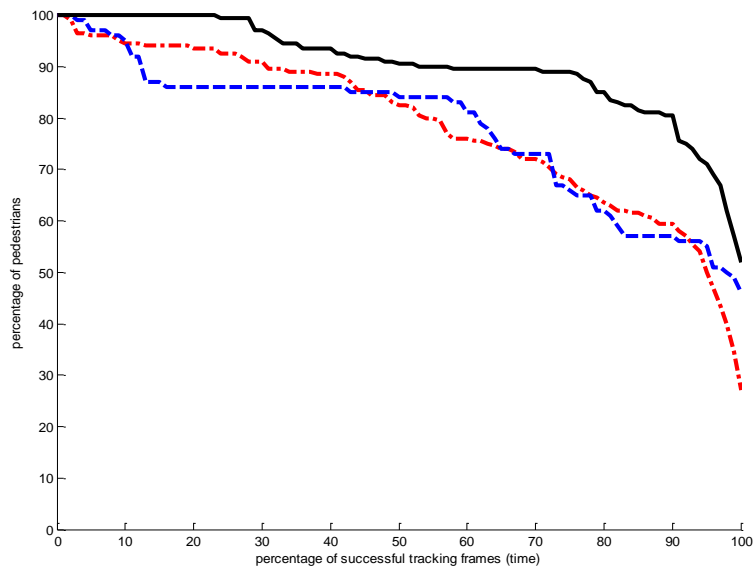


Figure 3. The percentage pedestrians that is successfully tracked for a number of frames for PF (red; dash-dot), PF+MP+HMM (blue; dashed) and PF+PO (black; solid).

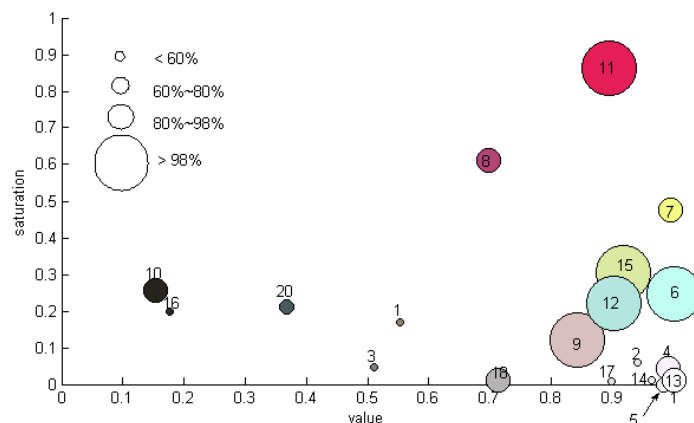


Figure 4. The graph shows the tracking performance of PF over pedestrians, where the color represents the appearance of the target and the size its performance. The PF shows bad performance for low saturation and low value.

Table 3. Tracking results of different approaches in the two scenes. The tracking performance is the average ratio of persons tracked over the complete duration.

Dataset	Method	Partial Occl.	Full Occl.	Abnormal	Normal	Average
<i>Musical Festival</i>	<i>Number of Targets</i>	18	2	3	10	20
	PF	0.73	0.87	0.68	0.82	0.75
	PF+MP+HMM	0.79	0.38	0.66	0.87	0.75
	PF+PO	0.93	0.42	0.91	0.92	0.88
<i>Train Station</i>	<i>Number of Targets</i>	8	2	-	-	10
	PF	0.88	0.67	-	-	0.84
	PF+MP+HMM	0.48	0.10	-	-	0.40
	PF+PO	0.94	0.65	-	-	0.88

6. CONCLUSIONS

In this paper, we proposed a method that uses the previous observation (PO) to track individuals in high-density crowds. The proposed tracker employs a particle-filter tracking framework, where the particles are propagated according to the previous observations. These observations are defined as a two-dimensional Gaussian, which models the dense distribution of the optical flow vectors in a local spatial-temporal area.

According to our experiments, the traditional color-based particle filter (PF) fails mainly on the targets with black, gray and white appearances, due to an unstable hue component. The result shows that modeling the temporal variation of motions with HMMs may not be a suitable choice for tracking in our video data. To train the HMMs properly, a large size of the training data is preferred. However, the training data becomes less relevant for tracking as the size increases. The results also show that choosing an optimal length in between helps to improve the tracking performance on normal pedestrians. For the abnormal persons, however, the performance is even worse than the PF tracker. Besides, the HMM tracker is very sensitive to regions with hardly any motion. By evaluating the proposed (PF+PO) tracker on varying duration of the previous observation, we find that better tracking results can be achieved by using a small number of frames in the previous observation.

REFERENCES

- [1] Ali, S. and Shah, M., "Floor fields for tracking in high density crowd scenes", ECCV II, 1–14 (2008).
- [2] Barron, J.L, Fleet, D.J, and Beauchemin, S.S., "Performance of optical flow techniques", IJCV 12(1), 43–77 (1994).
- [3] Berclaz, J., Fleuret, F. and Fua, P., "Robust people tracking with global trajectory optimization", IEEE CVPR 1, 744–750 (2006).
- [4] Bouma, H., Borsboom, S., Hollander, R., Landsmeer, S., Worring, M., "Re-identification of persons in multicamera surveillance under varying viewpoints and illumination", Proc. SPIE 8359, (2012).
- [5] Bouma, H., Hanckmann, P., Marck, J.-W., Penning, L., Hollander, R., Hove, J.M., Broek, S.P. van den, Schutte, K., Burghouts, G.J., "Automatic human action recognition in a scene from visual inputs", Proc. SPIE 8388, (2012).
- [6] Bradski, G. and Kaehler, A., [Learning OpenCV: Computer vision with the OpenCV library], O'Reilly Media, USA, (2008).
- [7] Brostow, G.J. and Cipolla, R., "Unsupervised Bayesian detection of independent motion in crowds", IEEE CVPR 1, 594–601 (2006).
- [8] Denman, S. Chandran, V. and Sridharan S., "Adaptive optical flow for person tracking", IEEE Digital Image Computing: Techniques and Applications, 44–50 (2005).

- [9] Farneback, G., “Two-frame motion estimation based on polynomial expansion”, *Image Analysis*, 363–370 (2003).
- [10] Hu, N., [Using previous observations to track people in high-density crowds], MSc Thesis University of Amsterdam, The Netherlands, (2011).
- [11] Huang, C., Wu, B., and Nevatia, R., “Robust object tracking by hierarchical association of detection responses”, *ECCV*, 788–801 (2008).
- [12] Isard, M. and Blake, A., “Condensation-conditional density propagation for visual tracking”, *IJCV* 29(1), (1998).
- [13] Jiang, H., Fels, S. and Little, J.J., “A linear programming approach for multiple object tracking”, *IEEE CVPR*, (2007).
- [14] Kratz, L. and Nishino, K., “Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models”, *IEEE CVPR*, 1446–1453 (2009).
- [15] Kratz, L. and Nishino, K., “Tracking with local spatio-temporal motion patterns in extremely crowded scenes”, *IEEE CVPR*, 693–700 (2010).
- [16] Kristan, M., Pers, J., Leonardis, A., and Kovacic, S., “Probabilistic tracking using optical flow to resolve color ambiguities”, *Computer Vision Winter Workshop*, 3–10 (2007).
- [17] Leibe, B., Schindler, K. and Van Gool, L., “Coupled detection and trajectory estimation for multi-object tracking”, *IEEE ICCV*, (2007).
- [18] Li, Y. and Ai, H., “Fast detection of independent motion in crowds guided by supervised learning”, *IEEE ICIP* 3, (2007).
- [19] Li, Y., Huang, C., and Nevatia, R., “Learning to associate: Hybrid-Boosted multi-target tracker for crowded scene”, *IEEE CVPR*, 2953–2960 (2009).
- [20] Lucena, M., Fuertes, J.M. and De la Blanca, N.P., “Evaluation of three optical flow-based observation models for tracking”, *IEEE CVPR* 4, 236–239 (2004).
- [21] Lucena, M.J., Fuertes, J.M., Gomez, J.I., De la Blanca, N.P., and Garrido, A., “Optical flow-based probabilistic tracking”, *IEEE Int. Symp. Signal Processing and Its Applications* 2, 219–222 (2003).
- [22] Metternich, M.J., Worring, M., Smeulders, A.W.M., “Color-based tracking in real-life surveillance data”, *Trans. Data hiding and multimedia security LNCS 6010*, 18–33 (2010).
- [23] Nummiaro, K., Koller-Meier, E., and Van Gool, L., “An adaptive color-based particle filter”, *Image and Vision Computing* 21(1), 99–110 (2003).
- [24] Rodriguez, M., Ali, S., Kanade, T., “Tracking in unstructured crowded scenes”, *ICCV*, 1389–1396 (2009).
- [25] Sebastian, P., Voon, Y.V. and Comley, R., “The effect of colour space on tracking robustness”, *IEEE Industrial Electronics and Applications*, 2512–2516 (2008).
- [26] Sugimura, D. and others, “Using individuality to track individuals: Clustering individual trajectories in crowds using local appearance and frequency trait”, *IEEE ICCV*, 1467–1474 (2009).
- [27] Tsutsui, H., Miura, J., and Shirai, Y., “Optical flow-based person tracking by multiple cameras”, *Int. Conf. Multisensor Fusion and Integration for Intelligent Systems*, 91–96 (2001).
- [28] Withagen, P.J., Schutte, K., Groen, F.C.A., “Likelihood-based object detection and object tracking using a color histograms and EM”, *Proc. IEEE Int. Conf. Image Processing* (1), 589–592 (2002).
- [29] Wu, B. and Nevatia, R., “Detection and segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses”, *IJCV* 82(2), 185–204 (2009).
- [30] Yamane, T., Shirai, Y., and Miura, J., “Person tracking by integrating optical flow and uniform brightness regions”, *IEEE Int. Conf. Robotics and Automation* 4, 3267–3272 (1998).
- [31] Yilmaz, A., Javed, O., and Shah, M., “Object tracking: A survey”, *ACM Computing Surveys* 38(4), (2006).
- [32] Zhang, L., Li, Y. and Nevatia, R., “Global data association for multi-object tracking using network flows”, *IEEE CVPR*, (2008).
- [33] Zhao, T. and Nevatia, R., “Tracking multiple humans in crowded environment”, *IEEE CVPR* 2, (2004).