Tracking by Cluster Analysis of Feature Points and Multiple Particle Filters[1]

Wei Du, Justus Piater

University of Liège, Department of Electrical Engineering and Computer Science,
Institut Montefiore, B28, Sart Tilman Campus, B-4000 Liège, Belgium
weidu@montefiore.ulg.ac.be, justus.piater@ulg.ac.be

Abstract. A moving target produces a coherent cluster of feature points in the image plane. This motivates our novel method of tracking multiple targets by cluster analysis of feature points and multiple particle filters. First, feature points are detected by a Harris corner detector and tracked by a Lucas-Kanade tracker. Clusters of moving targets are then initialized by grouping spatially co-located points with similar motion using the EM algorithm. Due to the non-Gaussian distribution of the points in a cluster and the multi-modality resulting from multiple targets, multiple particle filters are applied to track all the clusters simultaneously: one particle filter is started for one cluster. The proposed method is well suited for the typical video surveillance configuration where the cameras are still and targets of interest appear relatively small in the image. We demonstrate the effectiveness of our method on different PETS datasets.

1 Introduction

Tracking of moving targets is an elementary task in many computer vision applications such as video surveillance, sports analysis, human computer interaction, etc. Many different types of features have been used for tracking including points, edges, color, and templates. In this paper, we explore point features as they are ubiquitous and can be easily detected by e.g. the popular Harris corner detector [1].

Most previous work on point tracking focused on reconstructing individual point trajectories as long as possible. For instance, the Kanade-Lucas-Tomasi (KLT) algorithm [2] matches points by minimizing the sum of squared intensity differences. As minimization is sensitive to local extrema, KLT fails easily in case of occlusions and target deformation. In Arnaud *et al* [3], a stochastic filtering framework that blends a dynamic prior model and measurements provided by a matching technique was introduced and proved capable of dealing with abrupt motion changes and partial occlusions. In Shafique *et al* [4], optimal matching was adopted to exploit similarity information of feature points in multiple frames so that tracking is done by means of k-frame point correspondence using graph theory. However, the key problem remains: when a target is occluded or deforms, feature points

---

become less stable - corners disappear during occlusion or turn to edges during deformation - making tracking or matching individual points difficult.

In this paper, a novel method that attacks the instability problem with a different methodology is presented. The ultimate goal for most trackers is to detect and track moving targets, and the tracking of points is the means to achieve this goal. By observing that a moving target produces a coherent cluster of feature points in the image plane, tracking is converted to cluster analysis of feature points. First, feature points are detected by a Harris corner detector and tracked by a KLT tracker. Clusters of moving targets are then initialized by grouping spatially co-located points with similar motion using the EM algorithm [5]. Due to the non-Gaussian distribution of the points in a cluster and the multi-modality resulting from multiple targets, multiple particle filters [6] are applied to track the clusters in the following sequences. Therefore, instead of tracking individual points, we capture the stochastic properties of the clusters of feature points during tracking so that missing or unstable feature points don't affect the tracking results very much. Our method is well suited for the typical video surveillance configuration where the cameras are still and targets of interest appear relatively small in the image, thus feature points on them show strong coherence in space and motion. We demonstrate the effectiveness of our method on different PETS datasets [7].

The idea of tracking by cluster analysis was introduced by Pece [8] and borrowed into this work. Our contributions are, first, to apply it to points instead of regions, thus avoiding background modeling which is sensitive to illumination changes; second, to take motion coherence into account when computing measurements of clusters, which improves the robustness of cluster analysis; third, to integrate cluster analysis in the framework of particle filtering, which stabilizes the estimation of the cluster parameters significantly.

Section 2 describes the overview of our method and states the problem. Automatic initialization by EM based cluster analysis is given in Section 3. Section 4 introduces multiple target tracking using multiple particle filters. Results on sequences from PETS 2001 are illustrated in Section 5.

2 Overview

The motivation of this work is to develop a multitarget tracker for video surveillance applications. By detecting Harris corners and applying KLT in each frame, all the feature points with their associated velocities in the sequence are obtained, as shown in Figure 1. Points on moving targets exhibit large displacements, whereas points on the static background are characterized by very little motion.

An intuitive solution of tracking targets via feature points is to cluster coherent points using the EM algorithm [9]. However, the problems of using EM directly are, first, the number of points in a cluster varies from target to target and over time, depending on the size and appearance of the target. Sometimes few points in a cluster are detected due to the lack of texture information. Then, the spatial distribution of points in a cluster is not well represented by a Gaussian model; a finite uniform distribution is more appropriate. In contrast, the motion distribution of a cluster is well approximated by a Gaussian.
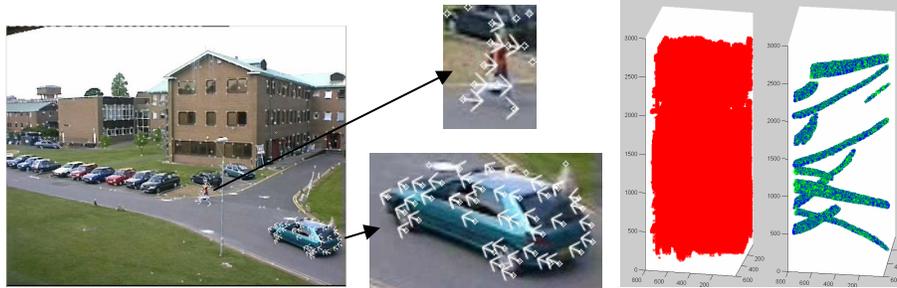
Fig. 1. Result of the Harris corner detection and the KLT tracking. In the left panel, point distributions of clusters are shown in the image plane. All the corners in the sequence are displayed in the spatio-temporal space in the right panel. After removing background points, the structure of the trajectories of moving targets can be clearly seen.

We apply multiple particle filters to solve these problems, as particle filters are well known for their ability to handle clutter and non-Gaussianity [10]. The main idea behind it is simple: Since feature points in a cluster are too sparse to model its distribution, a set of particles are sampled in a cluster. Each particle is evaluated according to some distance function so that it receives a weight reflecting the likelihood that the particle originates from the cluster. The cluster parameters are then updated from the weighted particles. Based on a prior motion model, the cluster distribution is propagated in the sequence so that the target is tracked. Multiple particle filters are applied to track multiple targets simultaneously. New filters are started when a large number of feature points exist that are not associated with any existing filters. Their parameters are initialized by clustering points using the EM algorithm. Existing filters are terminated when the total weights of their particles drop below a threshold. This happens in case of occlusions and targets leaving the scene.

2.1 Problem Statement

A feature point x_i is represented by its image coordinates u_i and its velocity s_i. A cluster of a target O_i is represented by a set of coherent feature points $\{x_j = (u_j, s_j), j = 1...n_i\}$, and is parameterized by a Gaussian $(o_i, \Sigma_i^o, v_i, \Sigma_i^v)$, where o_i is the spatial center, Σ_i^o is the spatial covariance, v_i is the average velocity, and Σ_i^v is the velocity covariance. The spatial and motion distributions of the points in a cluster are assumed independent.

Therefore, the problem of tracking is stated as: given the parameters of clusters in the previous frame, detect how many clusters are present in the current frame and assign each feature point to a cluster. In the following sections, we show how it is solved by initializing with EM based cluster analysis and tracking with multiple particle filters.

3 EM based Cluster Analysis

Automatic initialization is crucial to the success of a video surveillance system. Targets should be located when they first appear. An EM based cluster analysis algorithm is applied when a large number of feature points exist that are not associated with any existing clusters. Note that new targets may not only occur at the borders but anywhere within the image.

Deciding the number of clusters in the data is usually the hardest problem in cluster analysis. A voting technique was devised to solve this problem. Intuitively, each point spreads a weight to its neighbors based on the distance between them. After voting, each point computes its weight by collecting all the votes received. Points near the center of a cluster tend to have a larger weight. This method is incidentally the first phase ("sparse voting") of tensor voting [11]. By looking for local maxima, the number of new clusters and their centers are detected.

Using these results for initialization, an EM algorithm is applied to estimate the cluster parameters. The probability that a feature point i originates from a cluster j can be estimated from its location and the velocity, defined as $f_j(i) \propto \exp(-\mathrm{dist}(x_i, O_j)))$, where the distance between a point and a cluster is

$$\mathrm{dist}(x_i, O_j) = \left(\begin{bmatrix} u_i \\ s_i \end{bmatrix} - \begin{bmatrix} o_j \\ v_j \end{bmatrix} \right)^T \begin{bmatrix} \Sigma_j^o & 0 \\ 0 & \Sigma_j^v \end{bmatrix}^{-1} \left(\begin{bmatrix} u_i \\ s_i \end{bmatrix} - \begin{bmatrix} o_j \\ v_j \end{bmatrix} \right). \tag{1}$$

According to Bayes' theorem, the posterior probability that point i is generated by one of the clusters j is $p_j(i) = \dfrac{w_j f_j(i)}{\sum w_j f_j(i)}$, where w_j is the prior probability of cluster j defined as the fraction of image pixels generated from cluster j. Points are associated with the cluster that maximizes the posterior probability. Once all the points are assigned, the parameters of each cluster are re-estimated by summing the evidence over all its points. This is iterated until EM converges to a local maximum of the likelihood of the observed data. A phase of K-Means clustering is inserted to obtain a better initialization so that the EM algorithm converges with fewer iterations. In fact, in cases where the targets are well separated, EM does not change the output of K-Means at all. Results are shown in Figure 2.
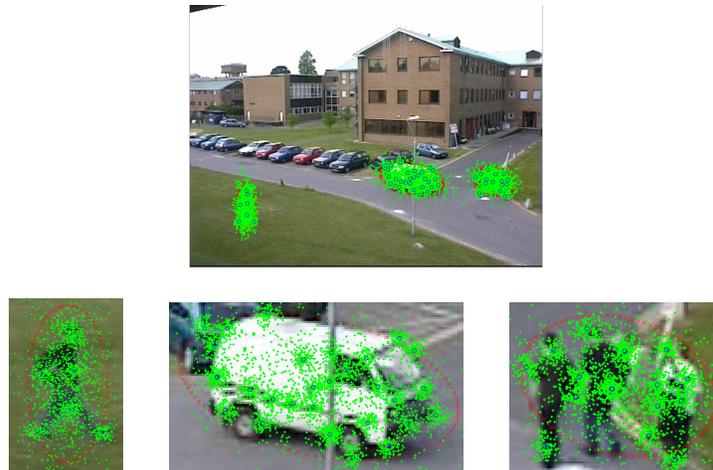
4 Multiple Particle Filters

Multiple particle filters are a simplified implementation of the mixture particle filter which is capable of maintaining the multi-modality of the posterior distribution and of tracking multiple targets simultaneously [6, 12, 13]. With a similar idea, we model each cluster with an individual particle filter, start a filter when a cluster is detected and terminate it when the cluster disappears.

4.1 Initialization of a particle filter

Given the initial parameters of a cluster obtained from the cluster analysis step, a particle filter is started. Two sets of particles are sampled in each filter: one from the initial distribution of the cluster and the other around each feature point in the cluster, shown in Figure 3.
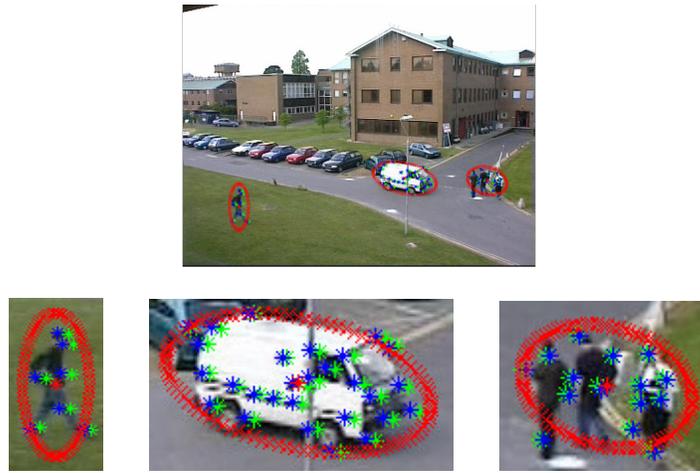
Fig. 2. Results of initialization of clusters in the first frame. Feature points detected by the Harris corner detector and tracked by KLT are grouped into clusters representing targets.

Fig. 3. Results of initialization of multiple particle filters. The green dots are sampled particles.

Let $x_{m,t}^k = (u_{m,t}^k, s_{m,t}^k)$ be the m-th particle in the k-th filter (corresponding to the k-th cluster) at time t, so

$$x_{m,0}^k = \text{sample}(O_k) \text{ or } x_{m,0}^k = x_i + \begin{bmatrix} \varepsilon_u \\ \varepsilon_s \end{bmatrix}, \tag{2}$$

where $x_i \in O_k$, ε_u and ε_s are random variables modeling respectively the changes in space and motion.

The reason of sampling 2 sets of particles is because of the non-Gaussianity of the feature points in a cluster. In this way, the particles are scattered in the cluster and the distribution is fully and well approximated. In all experiments, 100 particles are sampled around a feature point, and the number of particles sampled from the cluster distribution is proportional to the size of the cluster.

4.2 Tracking by Multiple Particle Filters

A particle in filter k is propagated in the sequence based on the constant velocity assumption,

$$x_{m,t+1}^k = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_{m,t}^k + \begin{bmatrix} \varepsilon_u \\ \varepsilon_s \end{bmatrix}, \tag{3}$$

and weighted by a function of the distances between the particle and the feature points around it, defined as

$$w_{m,t+1}^k = \sum_i \exp(-\text{dist}(x_{m,t+1}^k, x_{i,t+1})), \tag{4}$$

where the distance is

$$\text{dist}(x_{m,t+1}^k, x_{i,t+1}) = \left(x_{m,t+1}^k - x_{i,t+1}\right)^T \begin{bmatrix} \Sigma_u & 0 \\ 0 & \Sigma_s \end{bmatrix}^{-1} \left(x_{m,t+1}^k - x_{i,t+1}\right). \tag{5}$$

Σ_u and Σ_s are set to balance the influence of the distance in space and in velocity. The parameters of cluster k are then estimated from the weighted particles.

The computed parameters are not good enough because of the possible target deformation and the unstable feature point detection. For instance, the way that the weight is computed in Equation 4 tends to attract particles to the closest feature point. As a result, when new feature points appear in a frame, there may be few particles of large weights near them (especially when these new feature points are near the border of the cluster) so that their contributions to the estimation of the cluster parameters are unfortunately ignored.

To solve this problem, a one-step clustering is inserted to assign all the feature points to one of the clusters using their current parameters based on the distance defined by Equation 1. New particles are sampled around each feature point. The new sampled particles plus all the existing particles in a cluster are then reweighted by a function that averages the previously computed weight and the distances between particles and their clusters, defined as

$$w_{m,t+1}^k = \alpha \sum_i \exp(-\mathrm{dist}(x_{m,t+1}^k, x_{i,t+1})) + (1-\alpha)\exp(-\mathrm{dist}(x_{m,t+1}^k, O_{k,t+1})). \qquad (6)$$

The first term of the above equation describes the similarity measurement of the particle with its neighboring feature points, while the second term penalizes how coherent the particle is with the cluster. Finally, the parameters of the cluster are refined from its re-weighted particle set.

The final step of a particle filter is to resample particles based on their weights so that particles with small weights are likely to be discarded and those with large weights are duplicated. Note that a fixed number of particles in a filter are resampled during tracking.

In summary, the tracker consists of the following steps: (1) Prediction: particles are propagated using Equation 3. (2) Weighting: their importance weights are computed using Equation 4. (3) Clustering: assign feature points in the current frame to a cluster; new particles are sampled around each feature point. (4) Reweighting: particles are reweighted using Equation 6, and the parameters of the clusters are refined. (5) Resampling: resample particles using the Monte Carlo Sampling technique. These steps are iterated to propagate the distributions of the clusters in the sequence.

At the Clustering step, if a large number of feature points exist that are not associated with any existing filters, a new particle filter will be started and initialized by the EM based cluster analysis, as is stated in Section 3. At the Weighting step, if the total weight of all the particles in a filter drops below a threshold, the filter will be terminated. This happens when the target is occluded or leaves the scene.

5   Results

The proposed method is evaluated on different sequences from PETS2001. Figure 4 shows the result of tracking a subsequence of 300 frames in the sequence of Camera 1 of Dataset 1. Note that two crossing targets in the sequence are tracked separately during the occlusion, shown in the right panel of Figure 4, since they exhibit different motion.

Four challenging subsequences from the noisy sequence of Camera 1 of Dataset 3 are selected to evaluate the robustness of the method, as is demonstrated in Figure 5. They contain substantial and rapid illumination changes, shadows, severe occlusions and groups of people entering and leaving. The algorithm proves robust to substantial changes in illumination since the Harris corner detector is relatively insensitive to lighting changes. As shadows move along with the targets that cast them, they are tracked as a part of the targets and introduce only small jitter in the trajectories. The algorithm has problems maintaining a stable number of clusters in case of severe occlusions, because shadows connect distinct clusters and people move from one cluster to another. We are currently

studying complementary methods for tracking individual targets using model-based approaches.

Figure 6 and 7 illustrate the results of the comparison of our method with direct KLT tracking and our previous background-subtraction method [14]. The first comparison shows that KLT tracker fails during target deformation and occlusions, because when corners turn to edges, the tracks of points slide along edges, and when occluded, points drift from one target to another; meanwhile, our method is able to capture the stochastic properties of targets and is not affected by unstable feature points. The second comparison shows that background models are difficult to maintain in the presence of rapid lighting changes and fail in such situations (consult Piater *et al* [14] for more details), whereas our method is less sensitive and continues to track. The only problem is that shadows show up or disappear when illumination changes rapidly, which affects the parameters of clusters.

Nevertheless, a practical drawback of our method is that tracks of targets tend to be lost if they move slowly or possess little texture. Another drawback is that the method is only capable of dealing with partial occlusions. In case of complete occlusion, new targets are detected and are not linked to their correspondences before occlusion due to the lack of other information such as the appearance of the targets. However, an advantage of our method is that the errors will not be propagated in the sequence so that interactive reinitialization is unnecessary.

## 6   Conclusions and Future Work

This paper presents a novel method of tracking moving targets via feature points. The method is suitable for the video surveillance configuration where the cameras are still and targets are relatively small in the image so that feature points on a target form coherent spatio-temporal clusters. The EM algorithm and multiple particle filters are applied to cluster feature points and to track all the targets simultaneously. As demonstrated, the method is robust and capable of dealing with partial occlusions, shadows and illumination changes. We are currently focusing on tracking in difficult situations such as severe occlusions. Complementary methods for tracking individual targets over long sequences are being developed using model-based approaches and probabilistic data association. An extension of the current work to moving cameras is also ongoing and will broaden its application to e.g. sports analysis.

References

1. Harris, C., Stephens, M.: A Combined Corner and Edge Detector, Fourth Alvey Vision Conference, pp. 147-151, 1988.
2. Lucas, D.B., Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision, International Joint Conference on Artificial Intelligence, pp. 674-679, 1981.
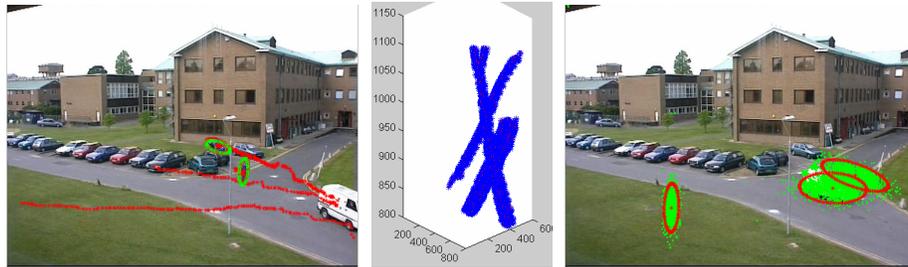
Fig. 4. Results of tracking. All the particles in the sequence are displayed in the spatio-temporal space in the middle panel.
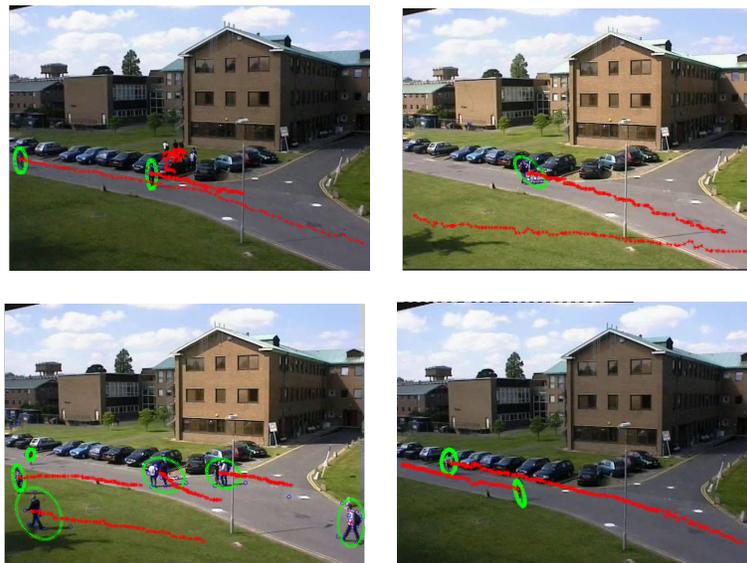

Fig. 5. Results of tracking four noisy sequences to evaluate the robustness of our method. Note that only long trajectories with large certainties are displayed.

3. Arnaud, E., Memin, E., Cernuschi-Frias, B.: A Robust Stochastic Filter for Point Tracking in Image Sequences, Asian Conference on Computer Vision, Korea, 2004.
4. Shafique, K., Shah, M.: A Noniterative Greedy Algorithm for Multiframe Point Correspondence, IEEE Transactions on PAMI, Vol. 27, No. 1, pp. 51-65, January 2005.
5. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical Society, Vol. 39, No. 1, pp. 1-38, 1977.
6. Lookingbill, A., Lieb, D., Stavens, D., Thrun, S.: Learning Activity-based Ground Models from a Moving Helicopter Platform, ICRA, Spain, 2005.
7. Ferryman, J.: PETS 2001 Datasets, http://visualsurveillance.org/PETS2001.
8. Pece, A.E.C.: Generative-Model-Based Tracking by Cluster Analysis of Image Differences, Robotics and Autonomous Systems, Vol. 49, No. 3, pp. 181-194, 2002.
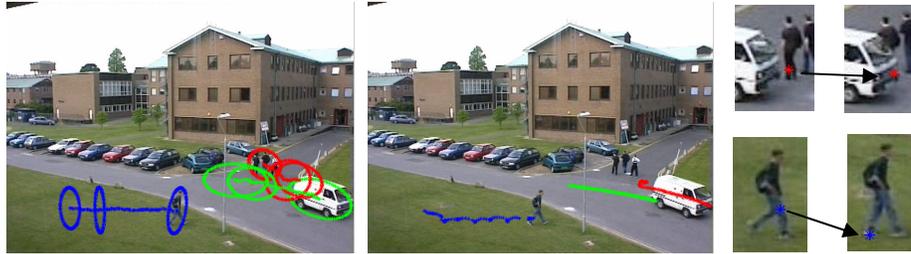
Fig. 6. Comparison of the method with direct KLT tracking. 3 KLT tracks are displayed in the middle panel. Note that the red one drifts from the group of people to the vehicle during occlusion (the first row of the right panel), and the blue one jumps from one leg of a pedestrian to another during deformation (the second row of the right panel). As our method captures the stochastic properties of the clusters of feature points, these unstable feature points don't affect the tracking results very much, shown in the left panel.
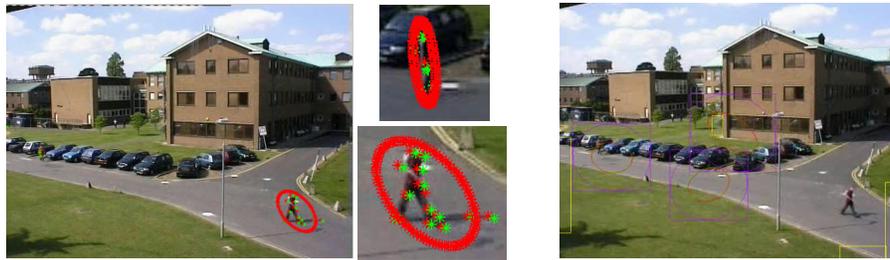
Fig. 7. Comparison of the method with our previous work which integrates background subtraction and motion history detection in the framework of a Kalman filter [14]. As expected, the track of the pedestrian is lost during the rapid illumination changes, shown in the right panel, whereas our new method succeeds in tracking in such situations, shown in the left panel. In the middle panel, two subimages of the same pedestrian at different time are displayed to show how the parameters of the cluster are affected by the shadow.

9. Du, W., Piater, J., Verly, J.: Tracking by Perceptually Grouping Feature Points into Clusters, submitted.
10. Doucet, A., Freitas, N., Godor, N., eds: Sequential Monte Carlo Methods in Practice, Springer Verlag, 2000.
11. Medioni, G., Tang, C.K.: Inference of Integrated Surface, Curve and Junction Descriptions from Sparse 3-D Data, IEEE Transactions on PAMI, Vol. 20, No. 11, pp. 1206-1223, 1998.
12. Vermaak, J., Doucet, A., Perez, P.: Maintaining Multi-Modality through Mixture Tracking, International Conference on Computer Vision, Nice, France, 2003.
13. Okuma, K., Taleghani, A., Freitas, N.D., Little, J.J., Lowe, D.G.: A Boosted Particle Filter: Multitarget Detection and Tracking, ECCV, Vol. 1, pp. 28-39, 2004.
14. Piater, J., Crowley, J.: Multi-Modal Tracking of Interacting Targets Using Gaussian Approximations. Proceedings of the Second IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Hawaii, USA, 2001.