# Data Fusion by Belief Propagation for Multi-Camera Tracking

Wei DU

Institut Montefiore, B28, University of Liège B-4000 Liège, Belgium weidu@montefiore.ulg.ac.be

Abstract - Multi-camera tracking poses a data fusion problem that integrates image measurements from different cameras. A novel approach to tracking using multiple cameras is proposed that combines particle filters and belief propagation in a unified framework. In each view, a target is tracked by a dedicated particle-filter-based local tracker. The trackers in different views collaborate via belief propagation so that a local tracker operating in one view is able to take advantage of additional information from other views. The message passing mechanism in belief propagation guarantees that wrong information is not propagated across views, thus avoiding a common problem in multi-camera tracking. Target states in each view and in 3D are inferred based on the multi-view image measurements by a set of particle filters, and a sequential belief propagation algorithm implements collaboration between the viewspecific particle filters. We demonstrate the effectiveness of our approach on sequences of video surveillance and soccer games.

**Keywords:** multi-camera tracking, data fusion, particle filters, belief propagation, sequential belief propagation.

## 1 Introduction

The use of multiple cameras is often preferable in many surveillance and sports analysis applications not only for resolving occlusions but also for extracting useful information such as precise target trajectories and events of interest. In general, multi-camera tracking is a data fusion problem that integrates image measurements from different cameras to detect and recursively localize targets of interest in a scene.

Commonly, a multi-camera tracking system contains a set of local trackers in each view, a communication procedure for exchanging information across views, and a global tracker for fusing data from each view. One important issue is the communication procedure which allows the collaboration among cameras. With a proper communication scheme, good measurements in some views can compensate for poor measurements in other views. A less sophisticated communicaJustus Piater Institut Montefiore, B28, University of Liège B-4000 Liège, Belgium justus.piater@ulg.ac.be

tion scheme may cause wrong information to be propagated across views so that the failure of one tracker breaks the whole system. In this paper, we address the communication issue in multi-camera tracking.

Recently, particle filters, also known as Sequential Monte Carlo Methods, have become popular tools to solve tracking problems [1]. The popularity stems from their simplicity, flexibility, and systematic treatment of non-linearity and non-Gaussianity. However, in the context of multi-sensor fusion, standard particle filters have problems with conflicting measurements in different sensors due to the unknown dependence relationships between them. Although the conditional dependence can be represented by graphical models [2], in many situations, direct inference based on graphical models is intractable due to the lack of analytical representations and the high dimensionality of the state space. Belief propagation (BP) is an efficient algorithm for solving the inference problem by passing messages between nodes in graphical models [3], which provides a rigorous way for the communication in multi-camera tracking.

A novel stochastic approach of tracking and fusing image measurements in multiple cameras is proposed that integrates the strength of particle filters and BP. Intuitively, a target is tracked in a view by a dedicated particle-filter-based local tracker. The trackers in different views interact with each other by passing messages via BP so that a local tracker operating in one view is able to take advantage of additional information from other views. In particular, target states in different views and in 3D are represented by different but dependent random variables, and a two-level, tree-structured graphical model is built to describe the dependence relationships between them. Thus, multicamera tracking amounts to inferring the multi-view target states based on the multi-view image measurements. By combining particle filters and BP in a unified framework, a sequential belief propagation algorithm is adopted to solve the inference problem, forming a collaborative tracking and fusion scheme. In doing so, we largely avoid propagating incorrect information across views, a common problem in multi-camera tracking, due to the asymmetric property of the message passing in BP [4]. We demonstrate the effectiveness of our approach on sequences of video surveillance

This work was sponsored by the Région Wallonne under DGTRE/WIST contract 031/5439.

and a soccer game.

Most related work on multi-camera tracking used similar hierarchical fusion frameworks in which targets are tracked independently in each view, and a fusion module integrates estimates from different views to generate precise trajectories [5]. Some frameworks even include feedback from the fusion module back to each view [5]. To prevent bias due to tracking errors in some views, many approaches assess the tracking performance to obtain a confidence weight for each view, and fuse only views providing reliable data [6]. However, these so-called best views are selected according to the quality of the tracking results. Therefore, the risk of failure is high if the targets being tracked are approached by other, similar objects. An example is the tracking of a soccer player surrounded by teammates. In such situations, view selection based on single-view confidence values is inherently unreliable.

In contrast to previous work, our approach combines the strengths of several tools. First, graphical models are used to describe the dependencies between target states in different views and in 3D. Our model has a tree structure that contains a set of leaf nodes and a central node. Then, Sequential Belief Propagation (SBP) [7] is adopted to have a set of particle-filterbased local trackers collaborate via a message passing procedure. Both graphical models and particle filters have been used in multi-sensor fusion such as multicamera tracking [8] or audio-video sensor fusion [9]. By combining them with BP, we propose a new formulation that solves the problem of conflicting measurements in a systematic way. Although the approach is designed for multi-camera tracking, the general framework is also suitable for other multi-sensor or multi-cue fusion tasks.

## 2 Model Description

Suppose L cameras are used and each camera collects one measurement for a target at a time instant. Due to occlusions, image noise or other sources of uncertainty, inaccurate measurements may be collected. Therefore, the problem of multi-camera tracking is to infer the target state in 3D, denoted by  $x_t$ , by fusing measurements in different views.

Let  $z_{t,j}$  be the image measurement in view j at time t,  $\mathbf{z}_t = \{z_{t,1}, \ldots, z_{t,L}\}$  the multi-view image measurement at time t, and  $\mathbf{z}^t = \{\mathbf{z}_1, \ldots, \mathbf{z}_t, \}$  the multi-view measurements up to time t. An optimal estimate of  $x_t$  is given by

$$\hat{x}_t = \int x_t p(x_t | \mathbf{z}^t) \, \mathrm{d}x_t,$$

where  $p(x_t | \mathbf{z}^t)$  is the posterior of  $x_t$  conditioned on  $\mathbf{z}^t$ .

Based on the Bayes' rule and the Markov property, the marginalization of  $p(x_t | \mathbf{z}^t)$  is formulated as

$$p(x_t | \mathbf{z}^t) \propto$$
(1)  
$$p(\mathbf{z}_t | x_t) \int p(x_t | x_{t-1}) p(x_{t-1} | \mathbf{z}^{t-1}) \, \mathrm{d}x_{t-1},$$

where  $p(\mathbf{z}_t|x_t)$  is the multi-view image likelihood function and  $p(x_t|x_{t-1})$  is a state prediction process that is associated with a prior motion model.

We assume that the measurements from different views  $z_{t,j}$ ,  $j = 1, \ldots, z_{t,L}$  are conditionally independent given the target state  $x_t$ , i.e.

$$p(\mathbf{z}_t|x_t) = \prod_{j=1}^{L} p(z_{t,j}|x_t), \qquad (2)$$

where  $p(z_{t,j}|x_t)$  is the image likelihood function in view j. Substituting Eq. 2 into Eq. 1, we obtain the update equation for  $p(x_t|\mathbf{z}^t)$ ,

$$p(x_t | \mathbf{z}^t) \propto (3)$$

$$\prod_{j=1}^{L} p(z_{t,j} | x_t) \int p(x_t | x_{t-1}) p(x_{t-1} | \mathbf{z}^{t-1}) \, \mathrm{d}x_{t-1}.$$

There are two reasons why direct inference using Eq. 3 is intractable. First, the analytical representation of Eq. 3 is not available in most situations such as non-linear motion and non-Gaussian distributions. Second, the measurements are not produced by the target alone, but jointly by the target and background clutter. Thus,  $p(\mathbf{z}_t|x_t)$  is easily affected by an inaccurate measurement in one particular view. For instance, if the target is occluded in one view, the measurement in this view is either incomplete if the occlusion is introduced by clutter in the background, or spurious if produced by similar objects. Consequently,  $p(\mathbf{z}_t|x_t)$  is biased by this inaccurate measurement, even if accurate measurements are available in other views.

Therefore, we reformulate the problem in the framework of BP. Random variables are defined to represent target states in different views, denoted by  $x_{t,j}$ ,  $j = 1, \ldots, L$ . To avoid confusion, the notation of the target state in 3D  $x_t$  is replaced by  $x_{t,0}$ . Putting all states together results in a multi-view target representation, denoted by  $\mathbf{x}_t = \{x_{t,0}, \ldots, x_{t,L}\}$ . A benefit of this representation is that it facilitates the integration of multi-view information, which helps overcome the occlusion problem.

Given the above definitions, a graphical model is built, as shown in Fig. 1(a), to model the dependence relationships between target states in different views and in 3D. The model consists of three types of nodes: the central node associated with  $x_{t,0}$ , the leaf nodes associated with  $x_{t,j}$ , and the measurement nodes associated with  $x_{t,j}$ ,  $j = 1, \ldots, L$ . Note that  $z_{t,j}$  is only associated with  $x_{t,0}$ . We assume that target states in different views are independent given  $x_{t,0}$ , which yields a tree-structured model. One advantage of this model is that it is acyclic so that BP performs exact inference [3]. Connecting the graphical models at different times results in a dynamic Markov model as shown in Fig. 1(b), which describes the evolution of the system.

In both models in Fig. 1, the undirected link between  $x_{t,j}$ ,  $j = 1, \ldots, L$ , and  $x_{t,0}$  describes the mutual influence between each leaf node and the central node, and is associated with a potential function  $\psi_{0,j}^t(x_{t,0}, x_{t,j})$ . The directed link from  $x_{t,j}$  to  $z_{t,j}$ ,  $j = 1, \ldots, L$ , represents the image measurement



Figure 1: The graphical representations of the multiview target representation: (a) tree-structured graphical model at time t, and (b) dynamic Markov model for the motion of the target.

process and is associated with an image likelihood function  $p_j(z_{t,j}|x_{t,j})$ . In Fig. 1(b), the directed link from  $x_{t-1,i}$  to  $x_{t,i}$ , i = 0, ..., L, represents the state prediction process and is associated with a motion model  $p(x_{t,i}|x_{t-1,i})$ .

Thus, the multi-camera tracking problem is reformulated as the recursive inference of  $p(\mathbf{x}_t | \mathbf{z}^t)$  based on the dynamic Markov model in Fig. 1(b),

$$p(\mathbf{x}_t | \mathbf{z}^t) \propto p(\mathbf{z}_t | \mathbf{x}_t) \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | \mathbf{z}^{t-1}) \, \mathrm{d}\mathbf{x}_{t-1},$$
(4)

where  $\mathbf{x}_t$  is the multi-view target state at time t, and  $\mathbf{z}^t$  denotes the multi-view image measurements up to time t. Direct inference using Eq. 4 is still intractable. In practice, we infer  $p(x_{t,i}|\mathbf{z}^t)$ ,  $i = 0, \ldots, L$ , collaboratively using SBP as described below.

## 3 Sequential Belief Propagation for Multi-Camera Tracking

#### 3.1 Sequential Belief Propagation

Conceptually, SBP computes each  $p(x_{t,i}|\mathbf{z}^t)$ ,  $i = 0, \ldots, L$ , based on local measurements and incoming messages from other nodes. We first consider the message passing procedure in the graphical model of Fig. 1(a). The message passed from leaf node j to the central node is

$$m_{0j}(x_{t,0}) = \int p_j(z_{t,j}|x_{t,j}) \,\psi_{0,j}^t(x_{t,0}, x_{t,j}) \,\mathrm{d}x_{t,j}, \quad (5)$$

where  $p_j(z_{t,j}|x_{t,j})$  is the image likelihood function in view j, and  $\psi_{0,j}^t(x_{t,0}, x_{t,j})$  is a symmetric potential

function that will be discussed in Section 3.2. Likewise, the message passed from the central node to leaf node j is

$$m_{j0}(x_{t,j}) = \int \left[\prod_{l \neq j} m_{0l}(x_{t,0})\right] \psi_{0,j}^t(x_{t,0}, x_{t,j}) \,\mathrm{d}x_{t,0}.$$
(6)

As the central node is not associated with any image measurement, it simply passes along the messages received from leaf nodes.

Our goal is to infer  $p(x_{t,j}|\mathbf{z}^t)$ ,  $j = 1, \ldots, L$ , and  $p(x_{t,0}|\mathbf{z}^t)$  based on the dynamic Markov model in Fig. 1(b). We assume independent motion models for target states in different views and in 3D, thus

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}) = \prod_{i=0}^{L} p(x_{t,i} | x_{t-1,i}).$$
(7)

Given  $p(x_{t-1,i}|\mathbf{z}^{t-1}), i = 0, ..., L$ , the above message equations are updated as

$$m_{0j}(x_{t,0}) = \int p_j(z_{t,j}|x_{t,j}) \psi_{0,j}^t(x_{t,0}, x_{t,j})$$
(8)

$$\left\{ \int p(x_{t,j}|x_{t-1,j}) \, p(x_{t-1,j}|\mathbf{z}^{t-1}) \, \mathrm{d}x_{t-1,j} \right\} \, \mathrm{d}x_{t,j},$$
$$m_{j0}(x_{t,j}) = \int \prod_{l \neq j} m_{0l}(x_{t,0}) \, \psi_{0,j}^t(x_{t,0}, x_{t,j}) \qquad (9)$$

$$\left\{ \int p(x_{t,0}|x_{t-1,0}) \, p(x_{t-1,0}|\mathbf{z}^{t-1}) \, \mathrm{d}x_{t-1,0} \right\} \, \mathrm{d}x_{t,0}.$$

Then, the marginal posteriors of  $x_{t,j}$ ,  $j = 1, \ldots, L$ , and  $x_{t,0}$  conditioned on  $\mathbf{z}^t$  are given respectively by

$$p(x_{t,j}|\mathbf{z}^t) \propto p_j(z_{t,j}|x_{t,j}) m_{j0}(x_{t,j})$$
(10)

$$p(x_{t,0}|\mathbf{z}^{t}) \propto \prod_{l=1,...,L} m_{0l}(x_{t,0}) p(x_{t-1,j}|\mathbf{z}^{t-1}) \,\mathrm{d}x_{t-1,j},$$
(11)
$$\int p(x_{t,0}|x_{t-1,0}) \, p(x_{t-1,0}|\mathbf{z}^{t-1}) \,\mathrm{d}x_{t-1,0}.$$

The above formulation shows that SBP involves both particle filters for propagating marginal posteriors over time, and BP for passing messages between nodes. In this work, we update messages and marginal posteriors using Eq. 8, Eq. 9, Eq. 10, and Eq. 11 iteratively. As the graphical model in Fig. 1(a) is a two-level tree, the SBP algorithm theoretically converges after two iterations [3]. However, due to the Monte-Carlo implementation, several more iterations are required to produce robust results.

In the Monte-Carlo implementation of the SBP algorithm, both the marginal posteriors and the messages passed between the central node and the leaf nodes are represented by weighted particles, and the same particle sets are used to represent the messages and the marginal posteriors. Intuitively, the algorithm employs particle filters to track targets in each view, where each particle is evaluated not only by local measurements in a given view but also using incoming messages from other views. Consult [10] for the implementation details.

#### 3.2 Potential Function

One issue in the SBP-based multi-camera tracking algorithm is the proper definition of the potential function that describes the mutual influence between the  $x_{t,j}$ ,  $j = 1, \ldots, L$ , and  $x_{t,0}$ . The potential function generally depends intimately on the application and on the definition of the state space, which often includes position, size, velocity, etc. In most video-surveillance and team-sports applications, the position of a target is usually the most important parameter. We assume that the targets of interest move on the ground plane and that the calibrations of the cameras are available either beforehand or by online updating using domain knowledge. Thus, the target states in different views  $x_{t,j}$ ,  $j = 1, \ldots, L$ , and in 3D  $x_{t,0}$  can be aligned through an image-to-ground homography.

Therefore, the potential functions  $\psi^t_{0,j}$  are defined as

$$\psi_{0,j}^{t}(x_{t,0}, x_{t,j}) \propto \lambda N(x_{t,0}; \mu_{x_{t,0}}, \Lambda_0) + (12) (1-\lambda) N(x_{t,0}; H_j^t(x_{t,j}), \Gamma_j^t(x_{t,j})),$$

where the first term is a standard Gaussian outlier process, the second term describes the spatial correlation between  $x_{t,0}$  and  $x_{t,j}$  and can be thought of as the distance between  $x_{t,0}$  and  $x_{t,j}$  after mapping  $x_{t,j}$ from the image plane to the ground-plane reference frame using the image-to-ground homography  $H_j^t$ .  $\Gamma_j^t$ propagates the uncertainties of  $x_{t,j}$  from view j to the ground plane using perturbation theory [11].

#### 3.3 Measurement Model

A measurement model is required to compute the likelihood  $p_j(z_{t,j}|x_{t,j}), j = 1, \ldots, L$ . Following Pérez et al. [12], a classical measurement model based on Hue-Saturation-Value color histograms is adopted due to their advantages of being insensitive to illumination effects and ease of implementation. Thus, in each view at a given time instant, the color model associated with a sampled particle is obtained by histogramming techniques. This model will be compared to a previously-learned reference color model, where the Bhattacharyya coefficient is computed to measure the similarity. The effectiveness of this model has been demonstrated earlier [12] and is confirmed by this work.

#### 4 Discussions

The strength of our approach lies in the fact that each particle-filter-based local tracker integrates information from multiple cameras using BP, which provides a systematic solution to conflicting measurements. It can be seen that good measurements can compensate for poor measurements due e.g. to occlusions. The approach is superior to other multi-sensor fusion frameworks [5] in that the full information from all views is taken into consideration during tracking. Even a view in which the target is completely occluded "contributes" to the tracking results by propagating uniformly distributed beliefs. As this view is not informative, it will not affect the tracking in other views. As was pointed out by Sun et al. [4], the message passing in BP is asymmetric: The entropy of the messages from high-confidence nodes to low-confidence nodes is smaller than the entropy of the messages from lowconfidence nodes to high-confidence nodes. Consequently, the propagation of incorrect information is avoided.

We find that the potential function is critical to the success of the approach. An important issue is that the mapping of the target positions from the image plane to the ground plane has large uncertainty if the camera viewing direction is highly oblique. In this case, a relatively large number of particles is needed to model the target distribution. This motivates the use of more views to reduce the uncertainty.

## 5 Results

We demonstrate the effectiveness of our approach on sequences of both video surveillance and a soccer game. Here, we assume the standard constant-velocity motion model in all experiments. To simplify, we manually initialize the regions of the targets in the first frame of each camera and learn the reference color models.

#### 5.1 Video Surveillance

Dataset Two from PETS2001<sup>1</sup> contains sequences taken from two calibrated cameras and is used to evaluate the algorithm. Fig. 2 shows the result of tracking a pedestrian in subsequences of Camera 1 and Camera 2 from Frame 600 to Frame 850. The pedestrian is completely occluded by a tree in Camera 1 but is visible all the time in Camera 2. Thus, the algorithm successfully tracks the target during the occlusion in Camera 1 by receiving messages from Camera 2 through the central node. Although the estimates during the occlusion are biased due to the absence of image measurements, they are corrected when the target reappears after the occlusion. The trajectories of two targets from Frame 600 to Frame 850 are shown in Fig. 3.

Note that similar results on the same sequences have been obtained using other methods [13, 8]. However, both cited contributions require explicit assessment of tracking performance in each view either for occlusion reasoning or for view selection. This assessment is achievable only if the occlusions are introduced by background clutter, as in this experiment. In soccer scenarios, the proximity of similar objects renders such assessments unreliable, as shown in the next section.

#### 5.2 Soccer Game

The algorithm is also tested on sequences of a soccer game taken from two uncalibrated cameras. We first compute the image-to-ground homographies either by using known features such as border lines on the field where enough of them are visible, e.g. near the penalty

<sup>&</sup>lt;sup>1</sup>http://visualsurveillance.org/pets2001

area, or by cumulating small estimates of motion between consecutive frames where no or few features are visible [14]. To compare with classical work in this field, we implemented the Condensation algorithm [15]. As expected, the results of Condensation, shown in Fig. 4, are hampered by occlusion and nearby clutter of similar appearance. In the end, the track of the target is lost and a wrong target is followed. In contrast, our algorithm integrates information from both cameras and is thus able to keep track of the correct target in both views, as shown in Fig. 5.

This comparison illustrates the strength of our algorithm. The proximity of similar objects poses the fundamental problem of selecting the correct mode that represents the true target from a multi-modal distribution produced jointly by multiple, similar objects. Neither the centralized fusion nor the best-view selection strategy are able to deal with this problem: The former will integrate the wrong results from camera M, and the latter method may select camera M as the best view, in which the wrong estimate happens to have a high match score. In contrast, our SBP-based multi-camera tracking algorithm enables the cameras to "talk" to each other so that the correct mode in a multi-modal distribution can be enhanced by receiving messages from other views.

## 6 Conclusion and Future Work

We presented a general multi-camera tracking and fusion framework that allows the cameras to "talk" to each other by passing messages so that the information from all views is taken into account during the tracking in each individual view. This approach is inherently insensitive to occlusions in only some of the views. Moreover, due to the use of Sequential Belief Propagation, the approach itself has some capacity of dealing with total occlusions in all views, as target hypotheses are generated according to a motion model. In practice, chances are that these hypotheses are not far from the truth if the duration of the occlusions is not too long. In our experience, the targets of interest are very unlikely to be occluded in all views for long periods of time under usual camera configurations for video surveillance and sports broadcasting.

We are currently developing algorithms for multitarget, multi-camera tracking, which involves the correspondences of the targets both in time and across views. Another interesting extension is to adapt the algorithm to cameras with non-overlapping fields of view, which will broaden the scope of applications of the tracker.

## References

- A. Doucet, N. de Freitas, and N. Gordon. Sequential Monte Carlo Methods in Practice. Springer-Verlag, New York, 2001.
- [2] M. Jordan. Learning in Graphical Models. MIT Press, Cambridge, MA, 1999.

- [3] J.S. Yedidia, W.T. Freeman, and Y. Weiss. Generalized belief propagation. Advances in Neural Information Processing Systems (NIPS), 13:689– 695, 2000.
- [4] J. Sun, N. Zheng, and S. Harry. Stereo matching using belief propagation. *IEEE Transactions* on Pattern Analysis and Machine Intelligence, 25(7):787–800, 2003.
- [5] A. Gad, M. Farooq, J. Serdula, and D. Peters. Multitarget tracking in a multisensor multiplatform environment. In the Seventh International Conference on Information Fusion, pages 206– 213, Stockholm, Sweden, 2004.
- [6] C. Beugnon, T. Singh, J. Llinas, and R.K. Saha. Adaptive track fusion in a multisensor environment. In the Third International Conference on Information Fusion, pages 24–31, Paris, France, 2000.
- [7] G. Hua and Y. Wu. Multi-scale visual tracking by sequential belief propagation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 826–833, Washington, DC, 2004.
- [8] Y.D. Wang, J.K. Wu, and A.A. Kassim. Multiple cameras tracking using particle filtering. In *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, pages 33–39, Breckenridge, CO, 2005.
- [9] P. Pérez, J. Vermaak, and A. Blake. Data fusion for visual tracking with particles. *Proceeding of* the IEEE, 92(3):495–513, 2004.
- [10] W. Du, J.B. Hayet, J. Piater, and J. Verly. Collaborative multi-camera tracking of athletes in team sports. In Workshop on Computer Vision Based Analysis in Sport Environments, in conjunction with ECCV2006, Graz, Austria, 2006.
- [11] A. Criminisi, I. Reid, and A. Zisserman. A plane measuring device. *Image Vision and Computing*, 17(8):625–634, 1999.
- [12] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet. Color-based probabilistic tracking. In *European Conference on Computer Vision*, pages 661–675, Copenhagen, Denmark, 2002.
- [13] J. Black and T. Ellis. Multi camera image tracking. In the Second IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Kauai, HI, 2001.
- [14] J.B. Hayet, T. Mathes, J. Czyk, J. Piater, J. Verly, and B. Macq. A modular multi-camera framework for team sports tracking. In *International Conference on Advanced Video and Signal based Surveillance*, pages 493–498, Como, Italy, 2005.
- [15] M. Isard and A. Blake. Condensation conditional density propagation for visual tracking. *In*ternational Journal of Computer Vision, 29(2):5– 28, 1998.





Camera 1 (600)

Ground

Camera 2 (600)



Camera 1 (680)



Ground



Camera 2 (680)



Camera 1 (740)



Ground



Camera 2 (740)



Camera 1 (800)

Ground



Camera 2 (800)

Figure 2: Results of tracking a pedestrian in PETS2001 Dataset 2. In the left and right columns, the colored rectangles have the following meaning: 1) blue for the particles sampled in the current view; 2) green for the particles mapped from the other view using the homography between the two views; 3) white for the estimate of the target state in a view; 4) red for the estimate of the target state in the other view mapped to this view using the same homography. In the center column, the colored dots have the following meaning: 1) blue dots for the particles of the target state in 3D; 2) red dots for the particles mapped from the two views using the image-to-ground homography; 3) white for the estimate of the target state in 3D.



Camera 1 (600-850)

Ground (600-850)

Camera 2 (600-850)

Figure 3: Results of tracking several targets in PETS sequences.



(a) Camera M (419)

(c) Camera M (432)



Figure 4: Results of Condensation. Due to the occlusion and the teammate approaching the target of interest, Condensation loses the track at Frame 519.



Camera M (432)

Camera B (432)



Camera M (508)

Camera B (508)



Camera M (519)

Camera B (519)

Figure 5: Results of our SBP-based multi-camera tracker. Although the player is occluded in Camera M by an opponent and is approached by a teammate, he is visible mostly in Camera B so that a strong belief propagated from Camera B attracts the target to the true state at Camera M.