Maximum-Likelihood Object Tracking from Multi-View Video by Combining Homography and Epipolar Constraints

Maximum-Likelihood Object Tracking from Multi-View Video by Combining Homography and Epipolar Constraints

Yixiao Yun Irene Yu-Hua Gu
Department of Signals and Systems
Chalmers University of Technology, Göteborg, 41296, Sweden
{yixiao, irenegu}@chalmers.se

Hamid Aghajan
Department of Electrical Engineering
Stanford University, CA 94305, USA
aghajan@stanford.edu

***Abstract*—This paper addresses problem of object tracking in occlusion scenarios, where multiple uncalibrated cameras with overlapping fields of view are used. We propose a novel method where tracking is first done independently for each view and then tracking results are mapped between each pair of views to improve the tracking in individual views, under the assumptions that objects are not occluded in all views and move uprightly on a planar ground which may induce a homography relation between each pair of views. The tracking results are mapped by jointly exploiting the geometric constraints of homography, epipolar and vertical vanishing point. Main contributions of this paper include: (a) formulate a reference model of multi-view object appearance using region covariance for each view; (b) define a likelihood measure based on geodesics on a Riemannian manifold that is consistent with the destination view by mapping both the estimated positions and appearances of tracked object from other views; (c) locate object in each individual view based on maximum likelihood criterion from multi-view estimations of object position. Experiments have been conducted on videos from multiple uncalibrated cameras, where targets experience long-term partial or full occlusions. Comparison with two existing methods and performance evaluations are also made. Test results have shown effectiveness of the proposed method in terms of robustness against tracking drifts caused by occlusions.**

***Index Terms*—multiple cameras, multiple view geometry, planar homography, epipolar geometry, visual object tracking**

I. Introduction

In visual object tracking, occlusion is a commonly encountered problem, as appearances of occluded targets could be significantly different from their reference models, thus leading to tracking drifts. Much efforts were made to tackle this issue. Many existing approaches dealt with occlusions in a single camera view [1] [2] [3] [4]. These methods can handle occlusion to some extent, but become less effective when objects undergo long-term full occlusions.

On the other hand, object tracking using multiple camera views has drawn growing research interest in recent years [5] [6] [7] [8] [9] [10], largely driven by its wide spatial coverage which is advantageous for handling complex scenarios, including occlusions. It is observed that occlusion of target usually does not occur in all views, information from un-occluded views can hence be used by exploiting some underlying multi-view geometric constraints to infer the target state in an occluded view. [9] [10] propose mechanisms to detect occlusion and maintain tracking in the occluded view by mapping the transformation matrix of object bounding box from un-occluded views using homography. The disadvantage lies in the assumption that occlusions produce large image differences. It might not hold since the occluding object might have similar appearance to the object of interest.

Motivated by the above issue, we propose a novel method for multi-view object tracking without requiring occlusion detection, as tracking results are mapped between views at each time instant. Potential tracking drift in occluded scene can be avoided by tracking results from un-occluded views. It is computationally efficient since the multi-view object appearance model is based on region covariance, which utilizes integral images for fast calculation. Further, we propose to use mean shift guided particle filter for tracking in each individual view where the number of particles is reduced.

The remainder of the paper is organized as follows: Section II and III present the big picture and the details of proposed scheme, respectively; Section IV describes our tracker in each individual view; Section V shows experimental results on multi-view videos containing severe occlusion scenarios; finally Section VI concludes the paper.

II. Problem Formulation: The Big Picture

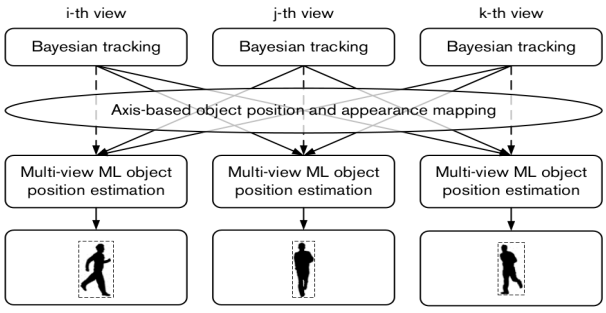

Fig. 1. A block diagram of the proposed multi-view tracking scheme. Without loss of generality, three camera views are depicted.

The proposed scheme consists of three layers for each individual view, as illustrated in Fig.1.

In the first layer, the purpose is to obtain independent maximum a posteriori (MAP) estimation of object positions in each view. This is done by performing individual-view object tracking, each formulated in a Bayesian framework. A particle filter is used to approximate the recursive Bayesian estimation, where mean shift using anisotropic kernel weighted color histogram is embedded to guide the particle filter. The appearance of candidate object described by each particle is represented by region covariance. Observation likelihood is measured based on geodesic between region covariances of the reference and candidate objects on a Riemannian manifold. In this way, each individual tracker results in a MAP estimation of object position in its view, where the vertical axis of tracked object is obtained.

In the second layer, the position and appearance of tracked object in each view are mapped separately to other camera views. The aim is to make the position and appearance of mapped object consistent with the destination view in terms of pixel coordinate, scale and 2D orientation. For illustration, we arbitrarily take an i-th view in Fig.1 as a destination view. First, the vertical axes of tracked object in j-th and k-th views are warped to i-th view by combining the constraints of planar homography, epipolar geometry and vertical vanishing point. Then, based on warped axes, positions and appearances of tracked object in j-th and k-th views are mapped via 2D similarity transformation to the i-th view. In a similar way, the reference model containing multi-view object appearances for i-th view is obtained by mapping the reference objects from j-th and k-th views, assembling covariances from these mapped regions and the reference object region in i-th view.

The third layer is the decision-making level. The basic idea is to pick up the best estimation of object position based on the maximum likelihood (ML) from individual views. The appearance of tracked object is mapped together with the position from other views to construct the likelihood measure that is consistent to the destination view. Thus, tracking results from all views are used to collaboratively improve the estimation of object position in i-th view. The distance function computes the dissimilarity between the reference model for i-th view and the appearance of tracked object either in i-th view or mapped from other views. The dissimilarity measure is the geodesic between region covariances on a Riemannian manifold.

The essence for occlusion handling in our proposed scheme is that multiple individual trackers in different views interact with each other at each time instant, hence the tracking drift in occluded view is mitigated by using un-occluded views. This differs from occlusion detection strategies where the tracker could suffer from false negative detection of occlusion occurrences, therefore the proposed scheme may lead to more robust tracking under full occlusion scenarios.

III. Multi-View ML Object Position Estimation

Assuming the total number of cameras used is M ($M \geq 2$). Given the positions and appearances of tracked object from M individual trackers in each view, our aim is to map these tracking results from combination of $(M-1)$ views to the remaining view, where the mapped position and appearance are consistent with the destination view in terms of pixel coordinate, scale and 2D orientation, so the likelihood measure of mapped object positions can be applied. For each individual view, $(M-1)$ mapped estimates and one self-estimate of the object position are collected to jointly improve the object position estimation in that particular view, based on the ML criterion.

A. Warping Object Vertical Axis

Under the assumption that objects move or stand uprightly on a dominant planar ground, which is usually the case in outdoor scenes, the planar homography, epipolar geometry and vertical vanishing point constraints are combined to warp the vertical axis of tracked object in one view to remaining views [11]. For the sake of completeness, we briefly describe this method. The vertical axis of object is the line segment connecting its top and ground points, see the dotted line segment in Fig.2.

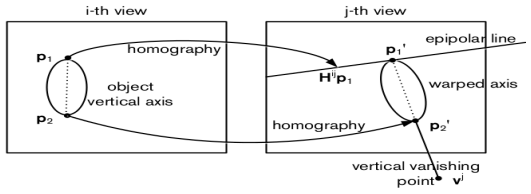

Fig. 2. Warping object vertical axis from i-th to j-th view by using planar homography, epipolar geometry and vertical vanishing point constraints.

1) Planar Homography: Let 2D homogeneous points $\mathbf{p}_1 \leftrightarrow \mathbf{p}'_1$ and $\mathbf{p}_2 \leftrightarrow \mathbf{p}'_2$ denote the corresponding top and ground points of object between i- and j-th view. Given the homography $\mathbf{H}^{ij}$ induced by the plane Π from the i-th view to j-th view, the correspondences of object ground positions are related by $\mathbf{p}'_2 = \mathbf{H}^{ij}\mathbf{p}_2$. However, for the top point $\mathbf{p}_1$ which is off Π, $\mathbf{p}'_1 \neq \mathbf{H}^{ij}\mathbf{p}_1$, as shown in Fig.2. Hence, homography is not sufficient for warping object vertical axis, other geometric constraints must be sought and added.

2) Epipolar Geometry: Given $\mathbf{p}_1$ in the i-th view, its corresponding point in j-th view $\mathbf{p}'_1$ lies on the projection of the preimage of $\mathbf{p}_1$ onto the j-th view. This relation is expressed by using the fundamental matrix $\mathbf{F}^{ij}$ satisfying $\mathbf{p}'_1\mathbf{F}^{ij}\mathbf{p}_1 = 0$. Since the preimage of $\mathbf{p}_1$ is a line, the projection of this line onto the j-th view gives the line $L(\mathbf{p}_1) = \mathbf{F}^{ij}\mathbf{p}_1$, which is the epipolar line associated with $\mathbf{p}_1$, as illustrated in Fig.2. Thus, the epipolar geometry constrains the corresponding points that lie on the conjugate pairs of epipolar lines.

3) Vertical Vanishing Point: To correctly obtain the warped axis inclination, the vertical vanishing point $\mathbf{v}^j$ of j-th view is used. As depicted in Fig.2, the warped axis lies on a straight line passing through $\mathbf{v}^j$ and $\mathbf{p}'_2$, and the top point $\mathbf{p}'_1$ is obtained as the intersection between the epipolar line and the straight line of the axis, $\mathbf{p}'_1 = (\mathbf{F}^{ij}\mathbf{p}_1) \times (\mathbf{v}^j \times \mathbf{p}'_2)$, where $\times$ is the homogeneous cross product operation.

Using the same process the vertical axis of tracked object in j-th view may be warped onto i-th view.

B. Mapping Position and Appearance of Tracked Object

In each individual view, a tracked object region is tightly bounded by an ellipse shape, then the region of tracked object at t in i-th view ($i = 1, \cdots, M$) is described by the shape parameters (or, state vector) of the ellipse as

$$\mathbf{s}_t^i = \left[x_0^i,\ y_0^i,\ h_x^i,\ h_y^i,\ \theta^i\right]^T \tag{1}$$

where (x_0^i, y_0^i) is the 2D center position, h_x^i and h_y^i the half lengths of major and minor axes, and θ^i the rotation angle.

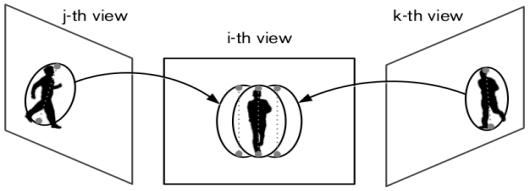

Fig. 3. Mapping estimated positions and appearances of tracked object from j-th and k-th view to i-th view, based on warped vertical axis.

As shown in Fig.3, the upper and lower vertices are connected by the major axis of ellipse. We approximate the top point of object by the upper vertex, and the ground point of object by the lower vertex, so the vertical axis of object is represented by the major axis of its bounding ellipse.

Given the warped vertical axis of tracked object from j-th to i-th view, tracking results in j-th view including the MAP estimation of object position $\mathbf{s}_t^j$ and its corresponding region R_t^j are projected to the i-th view, which become $\mathbf{s}_t^{ji}$ and R_t^{ji}. Assuming a constant aspect ratio for object across all camera views, the mapping of tracked object approximately obeys a 2D similarity transformation:

$$\mathbf{u}' = \lambda^{ji}\mathbf{R}(\psi^{ji})\mathbf{u} + \mathbf{t} \tag{2}$$

where $\mathbf{u}$ is a pixel at (x, y) in the j-th view and $\mathbf{u}'$ is the transformed pixel in the i-th view. λ^{ji} is a scaling factor, $\mathbf{R}(\psi^{ji})$ is a rotation matrix $\mathbf{R}(\psi^{ji}) = \begin{bmatrix} \cos\psi^{ji} & -\sin\psi^{ji} \\ \sin\psi^{ji} & \cos\psi^{ji} \end{bmatrix}$, and $\mathbf{t}$ is a translation vector $\mathbf{t} = \left[t_x^{ji},\ t_y^{ji}\right]^T$. The parameters λ^{ji}, ψ^{ji}, t_x^{ji} and t_y^{ji} can be derived from at least two pairs of corresponding points, so the upper and lower vertex correspondences between the vertical axis in j-th view and its warped axis in i-th view are sufficient to estimate the transformation parameters.

The parameters of $\mathbf{s}_t^{ji}$ describing the mapped position of tracked object from j-th view to i-th view are computed by

$$h_x^{ji} = \lambda^{ji}h_x^j, \quad h_y^{ji} = \lambda^{ji}h_y^j, \quad \theta^{ji} = \theta^j - \psi^{ji} \tag{3}$$

and (x_0^{ji}, y_0^{ji}) is obtained by (2), where $j = 1, \cdots, M,\ j \neq i$. Similarly, the appearance of tracked object in j-th view is mapped to i-th view by transformation of pixels in R_t^j using (2). In this way, positions and appearances of tracked objects from $(M-1)$ views are mapped onto the i-th view.

C. Multi-View ML Estimation of Object Position

Before applying ML, a reference model containing multi-view information for each view is formed. To obtain the reference model of i-th view, reference models in all views are used. Since object appearances in different views are not consistent with each other in terms of pixel coordinates, scale and 2D orientation, they are mapped to i-th view before computing the appearance feature descriptor. The mapping is achieved by using the method in Section III-B, resulting in a mapped reference object region R_{ref}^{ji} in i-th view from j-th view. For notational convenience, $R_{\text{ref}}^{ii} = R_{\text{ref}}^i$.

We choose region covariance as the feature descriptor of object appearance [12]. Let the covariance matrix of reference object in j-th view mapped to i-th view be $\mathbf{C}_{R_{\text{ref}}^{ji}}$. Then the reference model for i-th view $\mathcal{C}_{\text{ref}}^i$ is formed by M distinct view components $\mathcal{C}_{\text{ref}}^i = \left\{\mathbf{C}_{R_{\text{ref}}^{1i}},\ \cdots,\ \mathbf{C}_{R_{\text{ref}}^{Mi}}\right\}$. The dissimilarity measure is based on the geodesic on a Riemannian manifold computed between a candidate object image in R_t^{ji} and the reference model $\mathcal{C}_{\text{ref}}^i$ by

$$d(\mathcal{C}_{\text{ref}}^i, \mathbf{C}_{R_t^{ji}}) = \min_{k=1,\cdots,M} \sqrt{\sum_l \ln^2 \lambda_l(\mathbf{C}_{R_{\text{ref}}^{ki}}, \mathbf{C}_{R_t^{ji}})} \tag{4}$$

where λ_l is the generalized eigenvalue of $\mathbf{C}_{R_{\text{ref}}^{ki}}$ and $\mathbf{C}_{R_t^{ji}}$.

Given the ellipse region of tracked object R_t^{ji} (i.e., mapped from j-th view to i-th view), $\mathbf{C}_{R_t^{ji}}$ is computed. Similarly, denote $R_t^{ii} = R_t^i$ and $\mathbf{s}_t^{ii} = \mathbf{s}_t^i$. The likelihood is computed from the Gaussian-distributed *geodesic* between $\mathcal{C}_{\text{ref}}^i$ and $\mathbf{C}_{R_t^{ji}}$:

$$p(\mathbf{C}_{R_t^{ji}}|\mathcal{C}_{\text{ref}}^i) \propto \exp\left(-\frac{d^2(\mathcal{C}_{\text{ref}}^i, \mathbf{C}_{R_t^{ji}})}{2\sigma_i^2}\right) \tag{5}$$

where $d(\cdot)$ is the geodesic defined in (4) and σ_i is empirically determined. Then, the ML estimate in i-th view is obtained by:

$$\hat{\mathbf{s}}_t^i = \mathbf{s}_t^{j^*i}, \quad j^* = \arg\max_{j=1,\ldots,M} p(\mathbf{C}_{R_t^{ji}}|\mathcal{C}_{\text{ref}}^i) \tag{6}$$

If $j^* \neq i$, then, individual tracker in i-th view is re-initialized.

IV. Object Tracking in Individual Views

In this scheme, appearance-based tracking in Layer-1 (as shown in Fig.1) is performed independently in each view, using the previous multi-view tracking results. This corresponds to a single-view object tracker if one camera is used.

Similarly, region covariance is used to model the object appearance. The state vector $\mathbf{s}_t^i$ describing the parameters of ellipse which tightly bounds a tracked object region at time t in i-th view is given in (1).

By assuming that parameters smoothly change in video sequence, Brownian motion is used to model the dynamic between states: $\mathbf{s}_t^i = \mathbf{s}_{t-1}^i + \mathbf{w}_t^i$, where $\mathbf{w}_t^i \sim N(0,\ \mathbf{Q}^i)$ and $\mathbf{Q}^i = \text{diag}\left\{(\sigma_{x_0}^i)^2, (\sigma_{y_0}^i)^2, (\sigma_w^i)^2, (\sigma_h^i)^2, (\sigma_\theta^i)^2\right\}$ is the covariance containing diagonal elements, each corresponding to the variance of individual parameters of $\mathbf{s}_t^i$. These variances are determined empirically. The observation model measures the likelihood of the observation in candidate object regions given $\mathbf{s}_t^i$, same as (5).

A sequential importance sampling (SIS) particle filter with re-sampling is used to approximate the recursive Bayesian estimation. Anisotropic mean shift [13] which is adaptive to object shape, scale and orientation is employed to guide the particle filter.

V. Experimental Results

A. Experimental Setup

The proposed tracking method has been tested on PETS 2001, 2006 and 2009 [14] [15] [16], TU Graz Multi-Camera datasets [17]. Each dataset contains synchronized video from multiple cameras, where we chose videos containing full occlusions for our experiments, as described in Table I.

Dataset	No. camera views	No. tested frames	No. full occlusions	full occlusion frames: (shortest, longest)
PETS'01/S3.Tr	2	60	1	33
PETS'01/S3.Ts	2	129	1	50
PETS'06/S7	3	300	1	182
PETS'09/S2.L1	2	39	3	(1,7)
TU Graz /Easy	3	1000	14	(34,135)
TU Graz /Hard	3	719	19	(7,168)

TABLE I
Information on tested multi-view videos.

The planar homography and epipolar constraints are obtained by manually marking the corresponding points between each pair of views, and estimating the homography and fundamental matrices from these point correspondences using the Gold Standard algorithm [18]. The vertical vanishing points in individual views are obtained by detecting vertical lines using the Hough transform and estimating their convergence points with RANSAC. Fig.4 shows an example of axis-based mapping of object position and appearance using the above approach. We can see that the position, scale and 2D orientation of the mapped object conform well to the ground truth.

Fig. 4. Mapping estimated position and object appearance between a pair of views, using warped vertical axis. For each row, Column 1: detected vertical axis (dotted line) and tracked region in one view. '$\circ$' and '$\times$': top and ground points of object. Column 2: warped vertical axis in each destination view, where the vertical line passes through the warped axis, the inclined line is epipolar line associated with object top point from the other view in Column 1. Column 3: mapped object appearance in each destination view.

B. Test for the Proposed Tracking Scheme

Fig.5 shows several frames of a person running beside a lawn. The person is visible in the 1st view, however, gets long-term fully occluded by a tree in the later part of frames for the 2nd view. It can be clearly seen from the 2nd view (the occluded view) in Row 2 that our tracker still keeps track of the running person even when he is fully occluded by the tree, while the single-view tracker loses track of the person completely.

Fig.6 shows the sequence in a train station where a person with a suitcase is visible in the 3rd view, but gets partially occluded by a trolley in some frames and reappears afterwards in the 2nd view, and is long-term fully occluded by the same trolley in the later part of frames and reappears in the end for the 1st view. As we observed, even when the person is fully occluded as in frames # 775, 828 in Row 1 and partially occluded as in frame # 592 in Row 2, our tracker still follows the target as long as it is visible in other views. As a comparison shown in Row 1 and 2 of Fig.6, the single-view tracker drifts from the beginning of occlusion and is not able to recover the tracking afterwards.

Fig. 7. Results of three-view tracking for an object experiencing long-term full occlusion. Row 1: results from proposed multi-view ML tracking (yellow ellipse) for the 1st, 2nd and 3rd view, respectively. Row 2: the ground truth for each view. The frame is # 533. Multi-view videos are obtained from Easy scenario Set 1 of TU Graz Multi-Camera datasets.

Fig.7 shows one typical frame of the sequence in an indoor environment where a person walks around, gets fully occluded by a synthetically added book from time to time in each view. Fig.8 shows several frames of multiple people walking around, where frequent intersections occur. It is observed in both figures that though the object is invisible in some views due to occlusion, the trackers still follow the object with the tracking result mapped from the un-occluded view.

C. Comparisons

The proposed tracking scheme is then compared with two existing methods [9] [10] that also uses multiple cameras for occlusion handling. In Fig.9, where only the occluded view is shown, similar results are obtained by both methods. In Fig.10, where both views contain occlusions, it can be seen that our proposed tracking scheme adapts better to the object shape.

D. Performance Evaluation

To further evaluate the proposed tracking scheme, the dataset with synthetically added full occlusions (Fig.7) are used so the ground truth is available when long-term full occlusion occurs. Two objective measures are utilized:

- Euclidean distance between the ground/top point $(\hat{x}, \hat{y})$ of the tracked object and manually marked Ground Truth (GT): $d_E = \sqrt{(\hat{x} - x^{GT})^2 + (\hat{y} - y^{GT})^2}$;
- Bhattacharyya distance $\rho(p, q)$ between tracked object and GT object: $d_B = \sqrt{1 - \rho(p, q)}$, where $\rho(p, q)$ is defined between the color histograms p and q from tracked and GT regions.

Under each criterion, good performance is indicated by small values. The results are averaged over all views, as illustrated

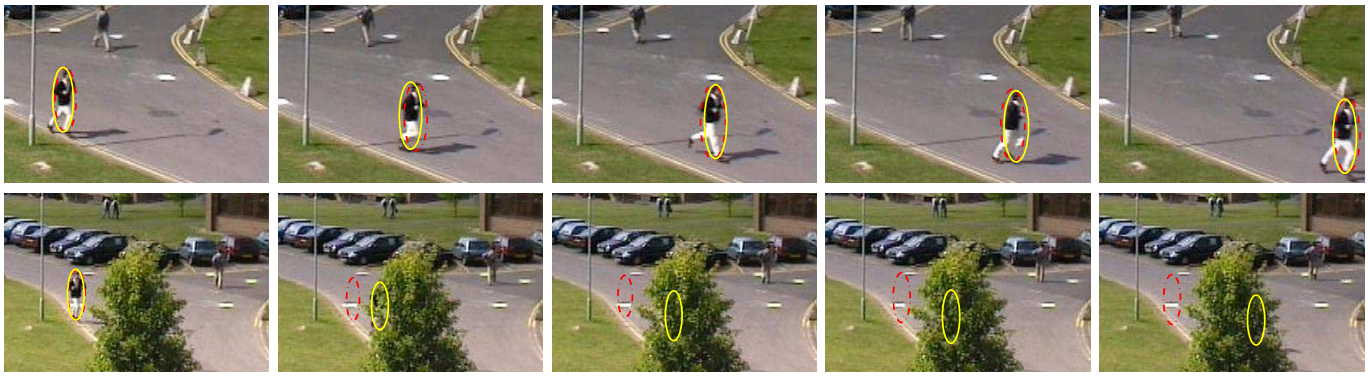

Fig. 5. Results of two-view tracking for an object experiencing long-term full occlusion. Rows 1-2: from single-view tracker (red dash-dot line) and proposed multi-view ML tracking (yellow solid line) for the 1st and 2nd view. Frames for each column are: # 2884, 2899, 2904, 2909, 2919. Multi-view videos are from Training set of PETS'01 S3.

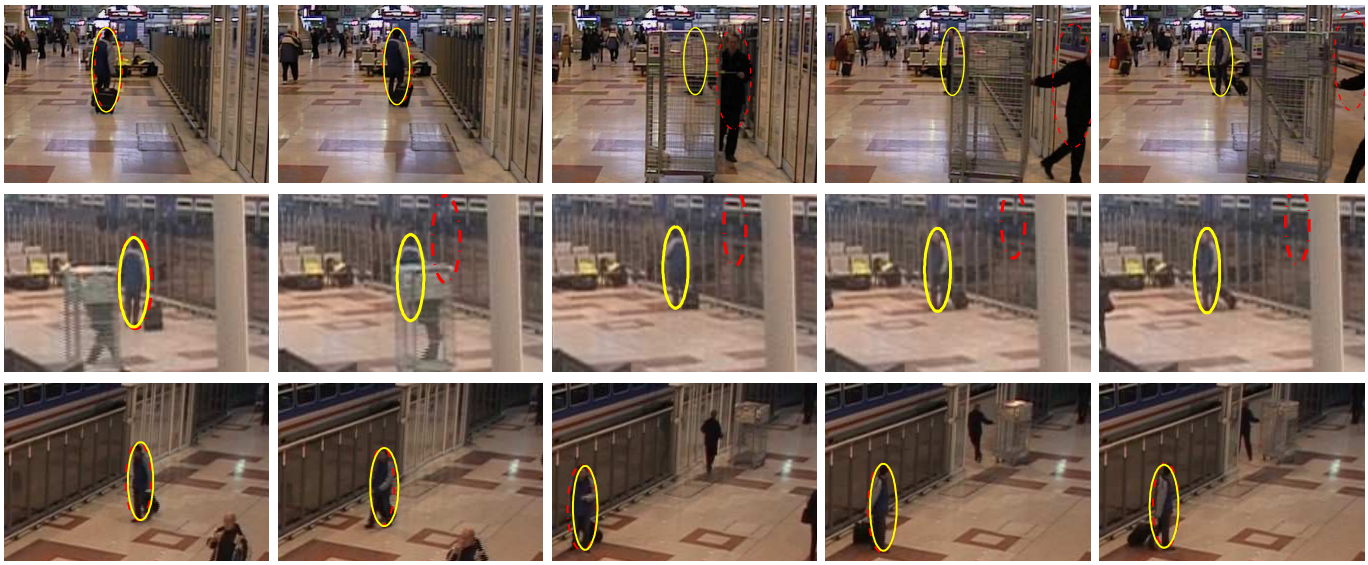

Fig. 6. Results of three-view tracking for an object experiencing partial and long-term full occlusions. Rows 1-3: results from single-view tracker (red dash-dot line) and proposed multi-view ML tracking (yellow solid line) for the 1st, 2nd, and 3rd view, respectively. Frames for each column: # 574, 592, 775, 828, 839. Multi-view videos are obtained from View 1, 2 and 4 of PETS'06 S7.

Fig. 8. Results of three-view tracking for an object experiencing frequent intersections. Row 1-3: results from proposed multi-view ML tracking (yellow ellipse) for the 1st, 2nd and 3rd view, respectively. Frames are: # 3201, 3272, 3369, 3702, 3835. Multi-view videos are from Hard scenario Set 1 of TU Graz Multi-Camera datasets.

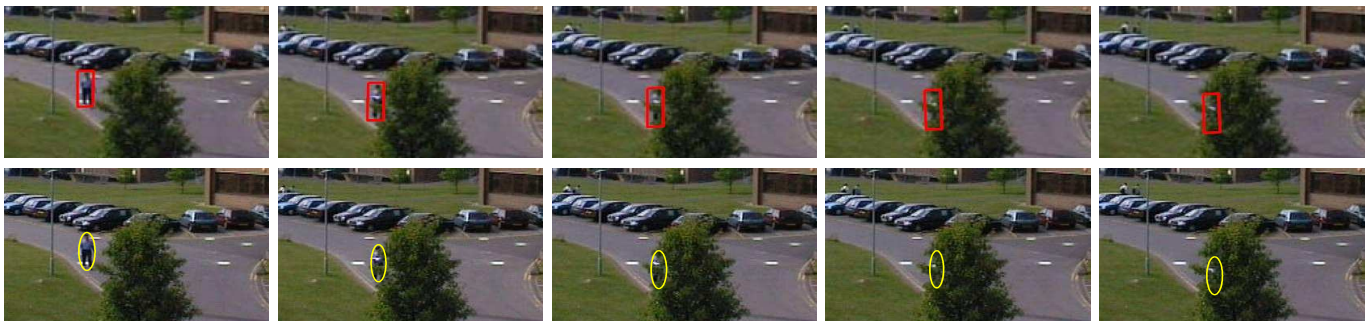

Fig. 9. Results of two-view tracking for an object experiencing long-term full occlusion. Row 1-2: results from [9] (red rectangle) for the 1st and 2nd view, respectively. Row 3-4: results from proposed multi-view ML tracking (yellow ellipse) for the 1st and 2nd view, respectively. Frames for each column are: # 5019, 5074, 5095, 5105, 5118. Multi-view videos are obtained from Testing set of PETS'01 S3.

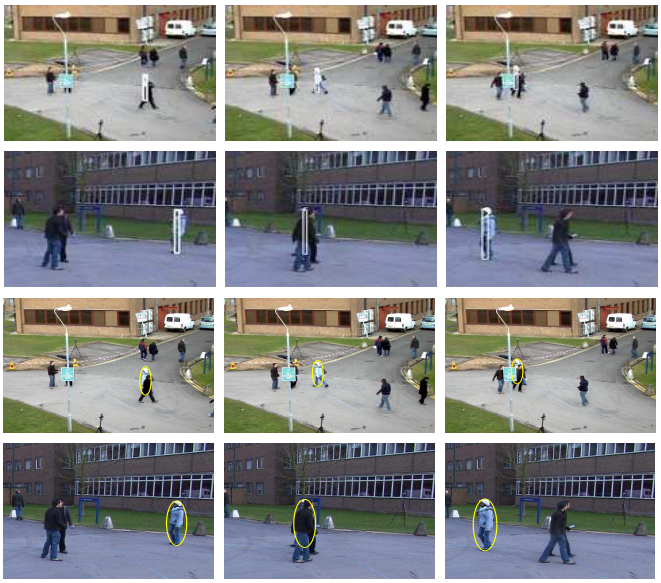

Fig. 10. Results of two-view tracking for an object experiencing intersections. Row 1-2: results from [10] (white rectangle) for the 1st and 2nd view, respectively. Row 3-4: results from proposed scheme (yellow ellipse) for the 1st and 2nd view, respectively. Frames for each column are: # 130, 145, 151. Multi-view videos are obtained from View 1 and 5 of PETS'09 S2.L1.

in Fig.11. It is observed that both Euclidean distance of ground/top points and Bhattacharyya distance of object region remain low with small variances for all frames, despite the frequent occurrences of full occlusions.

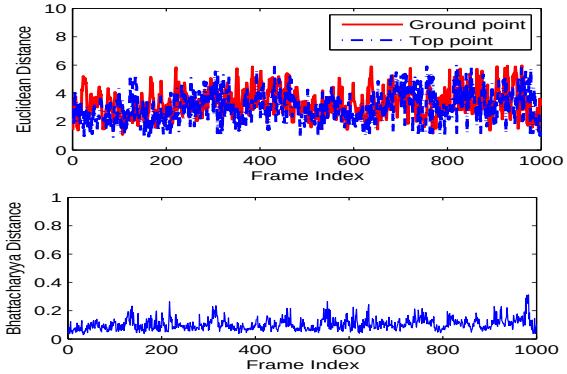

Fig. 11. Top: Euclidean distance between the ground/top points of tracked object and the ground truth. Bottom: Bhattacharyya distance between the tracked and the ground truth regions. The results are averaged over all views.

VI. Conclusion

The proposed multi-view tracker, through mapping positions and appearances of tracked object between camera views, and maximum likelihood estimation built upon geodesics on the Riemannian manifold, is tested with videos containing full occlusion. Results have shown the effectiveness of the proposed tracker, in terms of robustness against tracking drifts caused by long term full occlusions or object intersections. Performance evaluated using two criteria and comparisons with two existing methods have provided further support to the robustness of the proposed scheme. Future work will be conducted on extensive testing and evaluation on videos where appearances of occluding and target objects are more similar.

References

[1] J. Pan, B. Hu, "Robust occlusion handling in object tracking," IEEE Proc. CVPR, 2007.

[2] N. Papadakis, A. Bugeau, "Tracking with occlusions via graph cuts," IEEE Trans. PAMI, vol. 33, issue 1, pp. 144 - 157, 2011.

[3] G. Chao, S. Jeng, S. Lee, "An improved occlusion handling for appearance-based tracking," IEEE Proc. ICIP, 2011.

[4] S. Kwak, W. Nam, B. Han, J.H. Han, "Learning occlusion with likelihoods for visual tracking," IEEE Proc. ICCV, 2011.

[5] W. Du, J. Piater, "Multi-camera people tracking by collaborative particle filters and principal axis-based integration," Proc. ACCV, 2007.

[6] A.C. Sankaranarayanan, R. Chellappa, "Optimal multi-view fusion of object locations," IEEE Int'l Wksp Motion and Video Computing, 2008.

[7] Y. Zhou, H. Nicolas, J. Benois-Pineau, "A multi-resolution particle filter tracking in a multi-camera environment," IEEE Proc. ICIP, 2009.

[8] J. Fan, et al. "Distributed multi-camera object tracking with Bayesian inference," IEEE Int'l Symposium on Circuits and Systems, 2011.

[9] Z. Yue, S.K. Zhou, R. Chellappa, "Robust two-camera tracking using homography," IEEE Proc. ICASSP, 2004.

[10] B. Kwolek, "Multi camera-based person tracking using region covariance and homography constraint," IEEE Proc. AVSS, 2010.

[11] S. Calderara, A. Prati, R. Cucchiara, "HECOL: homography and epipolar-based consistent labeling for outdoor park surveillance," Int'l J. CVIU, vol. 111, issue 1, pp. 21 - 42, 2008.

[12] O. Tuzel, F. Porikli, P. Meer, "Region covariance: a fast descriptor for detection and classification," Proc. ECCV, 2006.

[13] Z.H. Khan, I.Y.H. Gu, A.G. Backhouse, "Robust visual object tracking using multi-mode anisotropic mean shift and particle filters," IEEE Trans. CSVT, vol. 21, issue 1, pp. 74-87, 2011.

[14] http://www.cvg.cs.rdg.ac.uk/PETS2001/, IEEE Int'l Wksp PETS, 2001.

[15] http://www.cvg.rdg.ac.uk/PETS2006/, IEEE Int'l Wksp PETS, 2006.

[16] http://www.cvg.rdg.ac.uk/PETS2009/, IEEE Int'l Wksp PETS, 2009.

[17] http://lrs.icg.tugraz.at/download.php, Multi-camera datasets, T.U. Graz.

[18] R. Hartley, A. Zisserman, "Multiple View Geometry in Computer Vision," Cambridge University Press, 2000.