



CHALMERS

Chalmers Publication Library

Tracking Visual and Infrared Objects using Joint Riemannian Manifold Appearance and Affine Shape Modeling

This document has been downloaded from Chalmers Publication Library (CPL). It is the author's version of a work that was accepted for publication in:

Proceedings of the IEEE International Conference on Computer Vision. 2011 IEEE International Conference on Computer Vision Workshops, ICCV Workshops 2011, Barcelona, 6 -13 November 2011

Citation for the published paper:

Khan, Z. ; Gu, I. (2011) "Tracking Visual and Infrared Objects using Joint Riemannian Manifold Appearance and Affine Shape Modeling". Proceedings of the IEEE International Conference on Computer Vision. 2011 IEEE International Conference on Computer Vision Workshops, ICCV Workshops 2011, Barcelona, 6 -13 November 2011 pp. 1847-1854.

<http://dx.doi.org/10.1109/ICCVW.2011.6130473>

Downloaded from: <http://publications.lib.chalmers.se/publication/145757>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.

Chalmers Publication Library (CPL) offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all types of publications: articles, dissertations, licentiate theses, masters theses, conference papers, reports etc. Since 2006 it is the official tool for Chalmers official publication statistics. To ensure that Chalmers research results are disseminated as widely as possible, an Open Access Policy has been adopted. The CPL service is administrated and maintained by Chalmers Library.

(article starts on next page)

Tracking Visual and Infrared Objects using Joint Riemannian Manifold Appearance and Affine Shape Modeling

Zulfiqar Hasan Khan, Irene Yu-Hua Gu

Dept. of Signals and Systems, Chalmers University of Technology, Gothenburg, 41296, Sweden

zulfiqak@chalmers.se, irenequ@chalmers.se

Abstract

This paper addresses the problem of object tracking from visual and infrared videos captured either by a dynamic or stationary camera where objects contain large pose changes. We propose a novel object tracking scheme that exploits the geometrical structure of Riemannian manifold and piecewise geodesics under a Bayesian framework. Two particle filters are alternately employed for tracking dynamic objects. One for online learning object appearances on Riemannian manifolds using tracked candidates, another for tracking object bounding box parameters with appearances on the manifold embedded. The rationale for obtaining this enhanced manifold tracker as compared with existing ones is to introduce an additional state variable, such that not only the manifold point representing the object is updated, but also the velocity of dynamic manifold point is estimated. Main contributions of the paper include: (a) propose an online appearance learning strategy by a particle filter on the manifold; (b) an object tracker that incorporates the manifold appearance for prediction under a particle filter framework; (c) use partitioned sub-regions of object bounding box that incorporates the spatial information in the appearance; (d) use Gabor features in different frequencies and orientations in partitioned sub-regions for IR (infrared) video objects. Hence, the proposed tracking scheme is applicable to both visual and IR videos. Experiments on videos where objects contain significant pose changes show very robust tracking results. The proposed scheme is also compared with two most relevant manifold tracking methods, results have shown much improved tracking performance in terms of tracking drift and tightness and accuracy of tracked boxes.

1. Introduction

Using covariance matrices of image features for object tracking has drawn increasing interest lately. [1] uses covariance matrices for characterizing the spatial features, sta-

tistical properties and correlations within the similar objects. It enables efficient fusion of different type of features while keeping small dimensionality, and is shown to be robust and versatile for variations in illuminations, views and poses at modest computational cost. The space of $n \times n$ non-singular covariance matrices of image features (or, Symmetric Positive Definite (SPD) matrices) can be formulated as connected points on the Riemannian manifold. The Log-Euclidean and affine invariant metrics [3, 2] provide a framework for generating the statistics on the Riemannian manifold. Numerical results of both metrics are similar, however, the first metric has a simpler form of distances and Riemannian means as compared with the second metric that has no closed form solution for Riemannian means.

[4] proposes a method of integrating the covariance matrix of image features in the visual tracking framework. It uses an exhaustive search and a distance measure proposed by [5] to find the best matching, and proposes the strategy for updating the model using Lie algebra on the SPD Riemannian manifold. The method may track objects with moderate pose changes however significant pose changes remain a challenging task. To improve this, variants of covariance tracking approaches are proposed, e.g. [6] uses a particle filter (PF) [7] and an affine invariant metric [2] on a SPD Riemannian manifold for finding the similarity of covariance matrices and tracking the location, width and height of an object bounding box. [8] incrementally learns the covariance matrix by a Log-Euclidean metric [3] on the SPD Riemannian manifold and a particle filter (PF) to track the central location and scale of object bounding box. However, none of these techniques uses simultaneous estimation of affine parameters of bounding box and online learning of the covariance descriptor. [9] employs the Log-Euclidean metric on the SPD Riemannian manifold for tracking the affine parameters of moving object box. It incrementally learns the eigen space representation of model in the tangent space of the SPD Riemannian manifold. Some good results are reported, however, it is very time consuming. Other recent work on manifold face tracking are reported,

e.g. using an off-line manifold training strategy from face datasets containing different poses [10], subsequently by online learning using local linearity of appearance manifolds using particle filters with a coarse-to-fine factorized sampling [11]. [12] proposes visual tracking by applying Kalman filters to the velocity of basis matrix in tangent planes of Grassmann manifolds. [13] proposes a Kalman filter on SPD Riemannian manifolds for visual tracking. [14] proposes a head pose estimation approach by using covariance matrices of object features and a nearest centroid classifier. Despite these, tracking visual objects containing significant pose changes remains an challenging issue.

To further tackle the problem, we propose to use several state variables under the proposed appearance model on the manifold and the affine shape model in the vector space. This enables to estimate both the objects and the change speed of objects, consequently, much more robust tracking performance. Details are described below.

2. General Description of the Scheme

The basic idea behind the proposed scheme is to exploit the geometrical structure of the Riemannian manifold and piecewise geodesics under a Bayesian framework. This is achieved by characterizing the object appearance by a covariance matrix of image features in partitioned sub-regions and representing it as a point on the Riemannian manifold. Further, the change of Riemannian manifold point is also estimated and updated.

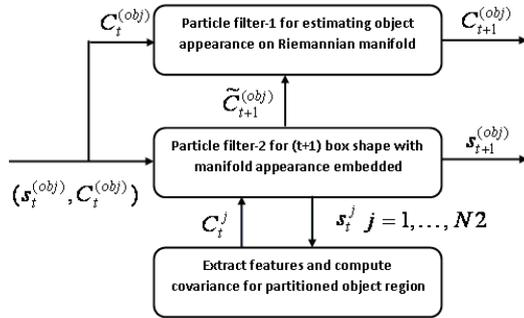


Figure 1. Block diagram of the proposed scheme. The notations C_t^{obj} , C_{t+1}^{obj} , s_t^{obj} , s_{t+1}^{obj} , C_t^j , s_t^j , denote: tracked object appearance covariance at t and $(t+1)$, tracked shape parameters of object bounding box at t and $(t+1)$, the candidate covariance matrices and candidate shape parameters, respectively.

Fig.1) shows the block diagram of the proposed tracking scheme. In Block 1 (top block in Fig.1) particle filter-1 is utilized for updating the object appearance C_{t+1}^{obj} on the Riemannian manifold using the tracked object appearance \tilde{C}_{t+1}^{obj} and the predicted manifold points \tilde{C}_{t+1}^i from the dynamic appearance model. The dynamic model is realized

by computing the velocity vector Δ_{t+1}^i of candidate manifold points under a constant velocity assumption and then mapping the velocity vector to the manifold originated from the previous manifold point \tilde{C}_t^i . In Block 2 (middle block), the shape parameters of objects bounding box is tracked by particle filter-2. The affine shape parameters are defined as the state vector, and the candidates box s_t^j are computed by the Brownian motion model, while the appearance computed by Block 3 (bottom block) is embedded as the likelihood, given a previous covariance matrix of object image features C_t^{obj} . It is known that 2D images of a 3D object with pose changes cannot be well described by a single vector space, rather, the trajectory of poses is more suitable to be described as moving points on a smoothed curved surface or on a differentiable manifold. The main motivations and novelties lie on online updating the covariance matrix on the SPD manifold by using not only the previous covariance matrix, but also the velocity indicating the change of covariance matrix. This may lead to more robust tracking.

3. SPD Riemannian Manifold

This section briefly reviews Riemannian geometry on the space of symmetric positive definite (SPD) matrices with a focus on the two mapping functions, distance metric and Riemannian mean. For simplifying the notation, throughout the text here after, we denote $Symm_n^+$ as the space of $n \times n$ SPD matrices on the Riemannian manifold, by \mathcal{M} as the Riemannian manifold and \mathcal{T} as the tangent space of the Riemannian manifold.

3.1. Riemannian geometry

The space of $Symm_n^+$ is not a vector space. Instead it lies on Riemannian manifold that constitutes a convex-half cone in the vector space of matrices. The derivative at a point on \mathcal{M} lies in the \mathcal{T} , which is a vector space formed by symmetric matrices, not necessarily $Symm_n^+$. Two Riemannian metrics, namely affine-invariant metric and Log-Euclidean metric, proposed by [2] and [3], are used for computing statistics on $Symm_n^+$. Numerical results of both Riemannian metrics are similar, however, the Log-Euclidean metric is computationally efficient, and calculation of mean points on \mathcal{M} is easy with a closed form. Readers are referred to [2, 3, 15] for more details.

3.2. Exponential mapping function ($\mathcal{T} \rightarrow \mathcal{M}$)

The exponential mapping function maps a tangent vector to a point on the manifold. Given a point P (i.e., a starting point $P(t=0)$) on the manifold \mathcal{M} and the corresponding tangent vector Δ in the tangent space \mathcal{T} , (1) maps the tangent vector along the geodesic to yield the end point Q on the manifold reached in unit time, i.e. $Q = P(1)$. The expo-

nential map [15] for the Log-Euclidean metric is given by:

$$\exp_P(\Delta) = \exp(\log P + \Delta) \quad (1)$$

3.3. Logarithmic mapping function ($\mathcal{M} \rightarrow \mathcal{T}$)

The logarithmic mapping function maps a manifold point to a vector in the tangent space. Given two points P, Q on \mathcal{M} , (2) results in a velocity vector Δ in \mathcal{T} corresponding to the geodesic from P to Q on \mathcal{M} . The logarithmic map [15] for the Log-Euclidean metric is given by:

$$\Delta = \log_P Q = \log Q - \log P \quad (2)$$

3.4. Geodesic

It is defined as the shortest distance between two points on the manifold. Given two points P, Q on \mathcal{M} , the geodesic for the Log-Euclidean metric is given by [15]:

$$D(P, Q) = \|\log_P Q\|_2 = \|\log Q - \log P\|_2 \quad (3)$$

3.5. The Riemannian mean

It is the expected value of a set of points on \mathcal{M} . Given a finite number of points $P_i, i = 1, \dots, N$, on \mathcal{M} , the expected value or the mean of the Log-Euclidean metric is given by:

$$E_{LE}(P_1, \dots, P_N) = \exp\left(\frac{1}{N} \sum_{i=1}^N \log P_i\right) \quad (4)$$

Computing the mean in (4) implies mapping the points on \mathcal{M} to the tangent space \mathcal{T} by using the log operator, followed by the mean in \mathcal{T} , and then mapping the result back to \mathcal{M} using the exp operator. In this work we use the Log-Euclidean metric for the Riemannian geometry.

4. Online Learning/Updating Object Appearances on the Riemannian Manifold

Online learning/updating of object appearance model is designed to mitigate the tracking drift due to object appearance changes through video frames. The basic idea is to consider the dynamic appearance of an object as a point moving on a nonlinear smoothing surface, or a Riemannian manifold, where novel dynamic models are applied to seek the trajectory of this moving point. Given a previous manifold point C_{t-1} at time $t-1$, a set of candidate manifold points C_t^j are generated by the dynamic model at current time t . The velocity vector for each individual manifold point is computed in the tangent plane, followed by computing their corresponding manifold candidate points. At time t , the best manifold point in (7) is then selected by finding the weighted Riemannian mean of the candidate points C_t^j under the Log-Euclidean metric, where the weights are

calculated by the geodesic to the tracked manifold point \tilde{C}_t (see Section 6). To realize this, two dynamic models, one is in the tangent plane of manifold point, and another is on the manifold itself, are formed as follows:

$$\begin{cases} \Delta_t = \Delta_{t-1} + V_1 \\ C_t = \exp_{C_{t-1}}(\Delta_t) \end{cases} \quad (5)$$

The first equation in (5) models the 1st order dynamics (i.e. velocity vectors) of manifold points under a constant velocity assumption, where V_1 is zero-mean white noise. The 2nd equation in (5) is the dynamic appearance model where two manifold points of successive time instants are related by mapping the velocity vector Δ_t in the tangent plane to the manifold with the origin as the previously tracked object point on the manifold C_{t-1} . Further, the covariance matrix C_t of feature vector f is defined by $C_t = E[ff^T]$, where f is defined in (8) for visual objects, and Gabor filter outputs at different orientations for infrared objects [19]. For taking object spatial information into account, each feature vector is formed from a set of partitioned areas (see Section 5). The above method can be further detailed as follows: A particle filter is applied on the Riemannian manifold to generate candidate points on the manifold C_t^j where $j = 1, \dots, N_1$ and N_1 is the number of particles. Let C_{t-1}^j be the previous manifold points at $t-1$ and Δ_{t-1}^j be the corresponding velocity vector that connects (C_{t-2}^j, C_{t-1}^j) , i.e. C_{t-1}^j is on the end point of the geodesic starting from C_{t-2}^j . The predicted velocity particles Δ_t^j are generated according to first equation of (5), with $\sigma_{V_1}^2$ ($\sigma_{V_1}^2 = .0001$ in our tests). Newly predicted manifold points C_t^j are then obtained by mapping Δ_t^j according to the second equation of (5). The corresponding weights for the particle filter (PF) is then obtained as:

$$w1_t^j = p(\tilde{C}_t | C_t^j) = \exp(-d(\tilde{C}_t, C_t^j)/\sigma_t^2) \quad (6)$$

where $d(\cdot)$ is the Log-Euclidean distance metric on the manifold, \tilde{C}_t is the manifold point using the tracked box (see Section 6) and σ_t^2 is the measurement noise ($\sigma_t^2 = 0.1$ in our tests). The weights are then normalized by $w1_t^j = w1_t^j / \sum_j w1_t^j$.

Finally, the MMSE estimate of the appearance covariance matrix C_t is obtained by weighted Log-Euclidean Riemannian mean of the predicted manifold points:

$$C_t = \exp\left(\frac{1}{N_1} \sum_{j=1}^{N_1} w1^j \log(C_t^j)\right) \quad (7)$$

5. Extracting Features from Partitioned Areas

This section describes details on forming feature vectors and covariance matrices from partitioned sub-regions. For

each pixel in the bounding box, a d -component feature vector $f(x, y)$ is defined, same as in [4], ($d = 9$ in our work)

$$f(x, y) = [x, y, I, |I_x|, |I_y|, I'_{mag}, I'_\theta, |I_{xx}|, |I_{yy}|]^T \quad (8)$$

where x, y is the pixel position, I is the image pixel intensity, $|I_x|$ and $|I_y|$ are the absolute values of 1st derivative of image intensity along the x and y directions, $I'_{mag} = \sqrt{I_x^2 + I_y^2}$ is the gradient magnitude, I'_θ is the unsigned gradient orientation, $I'_\theta = \begin{cases} \theta, & \text{if } \theta \geq 0 \\ \theta + \pi/2, & \text{if } \theta < 0 \end{cases}$ where $\theta = \tan^{-1}(\frac{I_y}{I_x})$, $|I_{xx}|$ and $|I_{yy}|$ are the absolute value of 2nd derivative of the intensity image along the x and y directions, respectively.

In the proposed tracker, feature vectors of object appearance are formed from partitioned sub-regions of bounding box. The rationale behind using partitioned areas is that two image regions having rather different contents can appear in similar statistics. By using spatially-dependent statistics from partitioned areas, such phenomenon is less likely to occur. Let the object appearance at t be described by a covariance matrix, representing a manifold point. Similar to [4], the covariance matrix of the object appearance is formed from the feature vector, however the difference is the covariance matrix consisting of L sub-covariance matrices as the result of partitioning object bounding box into L sub-regions. For the j th sub-region, $j = 1, \dots, L$, a sub-covariance matrix is formed from the sample average,

$$C^j = \frac{1}{M-1} \sum_{l=1}^M (f_j(l) - \mu_j)(f_j(l) - \mu_j)^T \quad (9)$$

where M and μ_j are the total number of samples and the sample mean of j th sub-region, respectively. For IR videos, we propose to use Gabor filter outputs from different bands and orientations in partitioned sub-regions to form the feature vector. The reason behind such choice is that IR images appear to be very different from visual band images (e.g. lack of sharp edges and image details). Selection effective features for IR images remains an open research issue. Using features from Gabor filter outputs would enable one to adjust filter bandwidths and select subbands of interest, hence to take IR image natures into account.

The Log-Euclidean metric on the Riemannian manifold can be explained as applying the logarithm to the the above sub-covariance matrix, resulting in $\log(C_j)$. Since the covariance matrix and its matrix logarithm are both symmetric, there are only $d \times (d+1)/2$ independent values. Therefore, $\log(C_j)$ is represented as vector of independent values, i.e. only by the upper triangular part of matrix

$$\text{vec}(\log(C^j)) = [\log(c_{1,1}^j), \log(c_{2,1}^j), \dots, \log(c_{d,d}^j)]^T \quad (10)$$

Finally, the vector representation of region ($\text{vec}(\log(C))$) is obtained by concatenating $\text{vec}(C^j)$ over all sub-regions:

$$\text{vec}(\log(C)) = [\text{vec}(\log(C^1)) \dots \text{vec}(\log(C^L))]^T \quad (11)$$

In our tests, $L=16$ (or, 4×4) partitioned sub-regions are used.

6. Tracking through Object Shape (Bounding Box) Modeling with Embedded Appearance Likelihood

The basic idea of tracking the object bounding box is to use a second particle filter for dynamic object shape modeling in the linear space, and to embed the object appearance from the previous frame as the likelihood. This is further detailed as follows: Let the state vector $s_t = [y_t^1 \ y_t^2 \ \beta_t \ \gamma_t \ \alpha_t \ \phi_t]^T$ be defined as the affine parameters of object bounding box shape, containing a six parameter vector (2D box center, scale, rotation, aspect ratio, and skew). The state space model of object shape is defined below, whose dynamic is the Brownian motion model (i.e., Gaussian distributed):

$$s_t = s_{t-1} + v_2, \quad s_t : p(s_t | s_{t-1}) \sim \mathcal{N}(0, \Omega) \quad (12)$$

where $\Omega = \text{diag}(\sigma_{y_1}^2, \sigma_{y_2}^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_\alpha^2, \sigma_\phi^2)$ and the diagonal elements correspond to the variances of shape parameters. These parameter values dictate the kind of motion of interest in a tracker and are determined empirically.

Predicted particles s_t^j are generated according to (12) where $j = 1, \dots, N_2$. For each candidate bounding box specified by the particle s_t^j , the corresponding image I_t^j is extracted and the covariance C^j and its vector form are calculated.

The likelihood is modeled as the Gaussian-distributed Log-Euclidean distance between the j th candidate appearance C_t^j and the previous tracked region appearance C_{t-1} :

$$p(d_{I_t^j} | s_t^j) = \exp(-(d(C_{t-1}, C_t^j)/\sigma^2)) \quad (13)$$

where σ^2 is the variance of measurement noise determined empirically. Noting that the likelihood is proportional to the dynamic prediction error on the manifold. The weight of the particle filter is then assigned as the likelihood,

$$w2_t^j = p(d_{I_t^j} | s_t^j) \quad (14)$$

with SIS re-sampling [7]. Finally, the maximum likelihood estimate of object bounding box is computed:

$$s_t = s_t^{j^*} \quad \text{where: } j^* = \text{argmax}_j(w2_t^j) \quad (15)$$

Based on the estimated state vector, the object appearance \tilde{C}_t is calculated and is used in the object update particle filter (PF-1).

7. Experiments and Results

For testing the effectiveness of the proposed scheme, several videos (visual and infrared) with significant object pose changes, captured by a nonstationary or a stationary camera, are used. The object bounding box in the first frame is manually marked, and the box is partitioned into $M = 16$ rectangular sub-regions. Each object region is normalized to 32×32 pixels; $N_1 = 600$, $\sigma^2 = 0.25$ are set for the particle filter PF_1 ; $N_2 = 400$, $\sigma_{v_2}^2 = 0.001$ are set for the particle filter PF_2 ; $\sigma_l^2 = 0.1$ is used. Fig.2-5 (Red box) shows the tracking results from 4 videos, where the first 3 videos are captured by visual-band camera and the fourth video is captured by infrared camera.

7.1. Comparisons

The proposed scheme is compared with two existing manifold trackers that are most relevant to our scheme: (a) *Tracker-1* uses covariance-based tracking in [4] (b) *Tracker-2* uses probabilistic tracking on the Riemannian manifold in [6]. We also compared with mean shift tracker and particle filter (or CONDENSATION)-based tracker. These trackers are not performed on the manifold. For videos with large pose changes, our trackers have shown superior performance over these trackers. Due to space limitations, comparisons of tracked objects in images using these methods are not included in this paper. Fig.2-5 shows the tracking results for four case studies.

In the first case (Fig.2), a human face is tracked from a visual-band video where the face has significant pose change accompanied by rotations, translations and scale changes. In the second case (Fig.3), face tracking is performed on a visual-band video containing walking man with appearance and pose changes of the face in different frames of video (captured by a dynamic camera). In the third case (Fig.4), car is tracked from a visual-band video captured by a moving camera and the car undergoes pose changes during the course of motion. In the fourth case (Fig.5), the video contains a human face captured by calibrated infrared (IR) camera and is rather challenging for tracking due to low contrast and strong thermal noise.

From the tracking results, one can see that *Tracker-1*, tracked areas have often drifted or lost from target objects due to its inability to follow the orientation changes. The error accumulates resulting in the loss of tracking eventually for video "Dudek". For *Tracker-2*, the performance is shown somewhat better, however the box size is often severely deviated from the real sizes may be due to lack of online learning to adapt object appearance change. The proposed method has clearly provided better tracking. The proposed method has successfully tracked target objects through videos, even during large pose change. This is due to embedding of the updated appearance (learned on the

Riemannian manifold) in likelihood for tracking bounding box shape affine parameters of moving object. The bounding box from the proposed method is shown to be relatively tight and accurate.

7.2. Performance Evaluation

To evaluate and compare the performance, the Euclidian distance is used to compute the distance between the 4 corners of tracked object box and the ground truth box (marked manually with visually acceptable orientation, size, width and height). Fig.6-a shows the resulting distances between the tracked region and the ground truth region as a function of image frames for 3 different methods on "Chia" face video sequence (Fig.2).

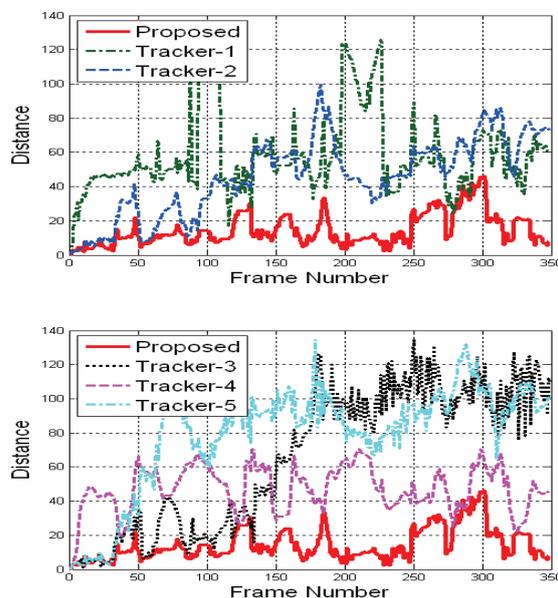


Figure 6. Results of Euclidian distances between the tracked and ground-truth regions for 2 video "chia". Left: Red curve: distances for the proposed tracker; Green curve: tracker-1 (i.e., the covariance tracker in [4]); Blue curve: tracker-2 (i.e., the probabilistic tracker on the Riemannian manifold in [6]); Right: Red curve: distances for the proposed tracker; Black curve: tracker-3 (i.e., the incremental subspace tracker in [16]); Magenta curve: tracker-4 (i.e., the mean shift tracker in [17]); Cyan curve: tracker-5 (i.e. the CONDENSATION tracker in [18])

To further evaluate the tracking performance, the proposed method has been compared with three existing state of art video tracking methods : (a) *Tracker-3* uses subspace-learning based tracking in [16] (b) *Tracker-4* uses mean shift tracking in [17] (c) *Tracker-5* uses condensation tracking in [18]. Fig.6-b shows the Euclidean distances of tracking results. *Tracker-3* initially performs well (first 130

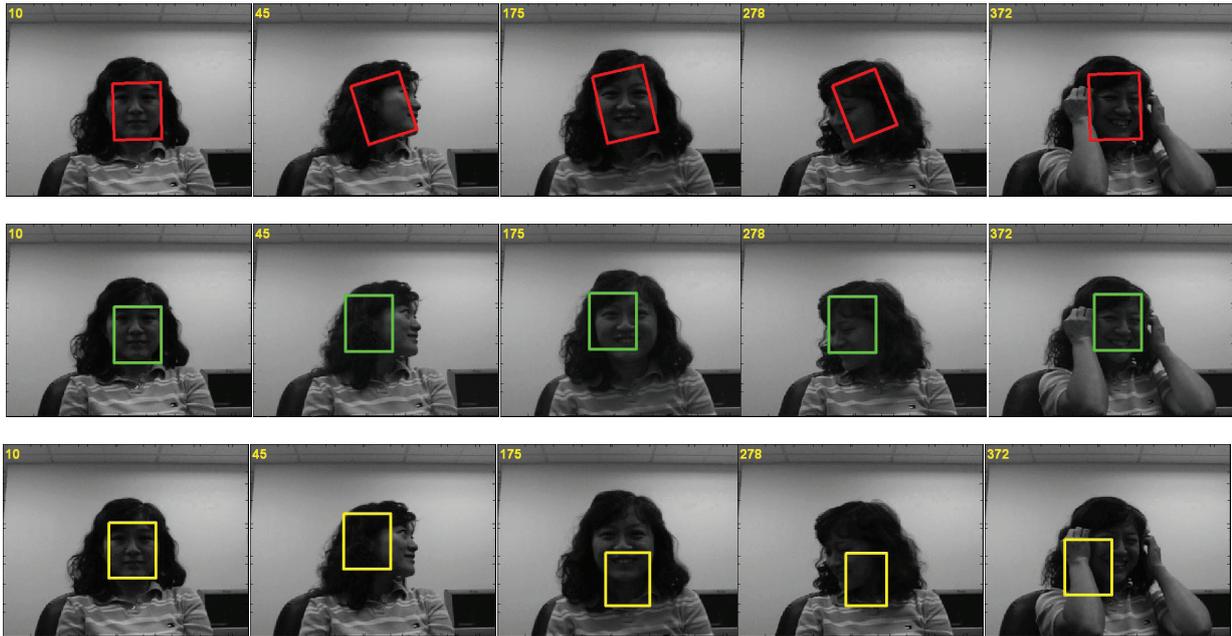


Figure 2. Tracking results from visual-band video frames in "Chia". Rows 1-2: from the proposed scheme (Red box); Rows 3-4: from *Tracker-1* (green box); Rows 5-6: from *Tracker-2* (Yellow box).



Figure 3. Tracking results from visual-band video frames in "dudek". Row-1: from the proposed scheme (Red box); Row-2: from *Tracker-1* (green box); Row-3: from *Tracker-2* (yellow box).

frames) and then gradually loses track during fast and significant pose change probably due to vector space learning of the appearance subspace. The reason for the drift and

loss of tracking for *Tracker-4* and *Tracker-5* is due to lack of online learning during the course of tracking. Comparison of the results in Fig.3-6 show that the proposed tracker



Figure 4. Tracking results from visual-band video frames in "Car". Row-1: from the proposed scheme (Red box); Row-2: from *Tracker-1* where the results are copied from the figure in [4] (Green box); Row-3: from *Tracker-2* (Yellow box).

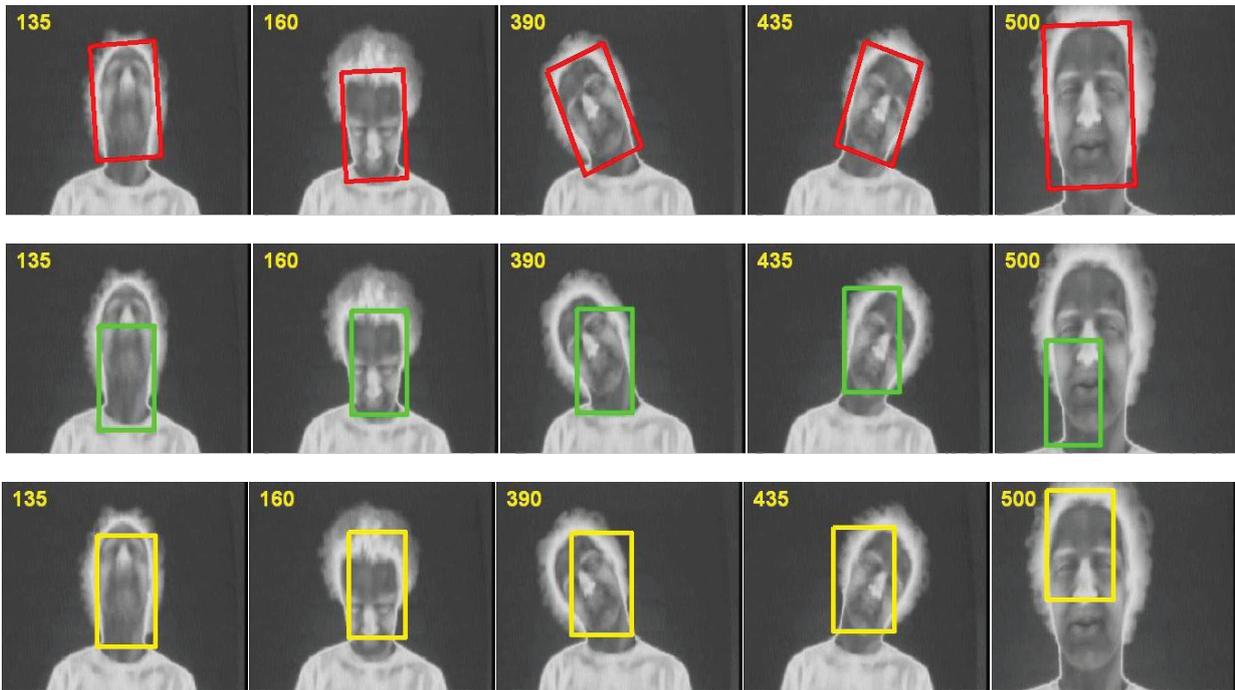


Figure 5. Tracking results from one infrared video. Row-1: from the proposed scheme (Red box); Row-2: from *Tracker-1* (Green box); Row-3: from *Tracker-2* (Yellow Box).

is able to provide better tracking performance.

7.3. Discussion

Computational time for the proposed tracking scheme (average 15 seconds for tracking one frame of image using our Matlab program in a PC with an Intel Xeon CPU 2GHz and 4 GB RAM) still poses a problem and limits the real-time applications. To overcome this, C-programs with optimized codes, or FPGA can be considered to replace the Matlab programs used in current tests.

8. Conclusion

Test results of the proposed tracking scheme on visual and infrared videos for tracking of object box shape parameters with embedded object appearance on the Riemannian manifold, as well as updating the object appearance on the manifold using tracked observations, have shown to be very robust tracking performance, especially for objects with large dynamic of pose changes. The online learning/updating of object appearance covariance matrix on the manifold by exploiting both the position and velocity is shown to be effective. The online tracking by integrating the dynamic appearance on the manifold and object shape in the vector space, and through employing two particle filters separately on the manifold and in the vector space, is shown to be very effective. Comparisons with two existing manifold-based tracking methods have provided further support to the robustness of the proposed scheme. Finally, using IR object appearance features from Gabor filter outputs from different subbands and orientations in partitioned areas is shown to be effective.

References

- [1] O.Tuzel, F.Porikli and P.Meer, "Region covariance: a fast descriptor for detection and classification", Proc. ECCV, pp. 589-600, 2006. [1](#)
- [2] X. Pennec, P. Fillard and N. Ayache, "A riemannian framework for tensor computing", Int. J. Comput. Vision, 66(1), pp. 41-66, 2006 [1](#), [2](#)
- [3] V. Arsigny, P. Fillard, X.Pennec and N. Ayache, "Geometric means in a novel vector space structure on symmetric-positive definite matrices", SIAM J. Matrix Anal. Appl., 66(1), pp. 328-347, 2008. [1](#), [2](#)
- [4] F.Porikli, O.Tuzel, P.Meer, "Covariance tracking using model update based on Lie algebra", Proc. IEEE CVPR, pp. 728-735, 2006. [1](#), [4](#), [5](#), [7](#)
- [5] W. Forstner and B. Moonen, "A metric for covariance matrices", Technical report, Dept. of Geodesy and Geoinformatics, Stuttgart University, 1999. [1](#)
- [6] Y. Wu, B. Wu, J. Liu and H. Lu, "Probabilistic tracking on Riemannian manifolds", Proc. ICPR, pp. 1-4, 2008. [1](#), [5](#)
- [7] A. Dore, M. Soto, C. Regazzoni, "Bayesian tracking for video analytics", IEEE Trans. Image Processing, 27(5), pp. 46-55, 2010. [1](#), [4](#)
- [8] Y. Wu, J. Cheng, J. Wang and H. Lu, "Real-time visual tracking via Incremental Covariance Tensor Learning", Proc. ICCV, pp. 1631-1638, 2009. [1](#)
- [9] X. Li, W. Hu, Z. Zhang, X. Zhang, M. Zhu and J. Cheng, "Visual tracking via incremental Log-Euclidean Riemannian subspace learning", Proc. IEEE CVPR, pp. 1-8, 2008. [1](#)
- [10] H. Qiao, P. Zhang, B. Zhang, S.Zheng, "Learning an intrinsic-variable preserving manifold for dynamic visual tracking", IEEE Trans. Syst., Man, Cybern., vol. 40, no. 3, pp. 868-880, 2010. [2](#)
- [11] Y.M.Lui, J.R.Beveridge, L.D.Whitley Subbarao, "Adaptive appearance model and condensation algorithm for robust face tracking", IEEE Trans. Syst., Man, Cybern., 40(3), pp. 437-448, 2010. [2](#)
- [12] T.Wang, A.G.Backhouse, I.Y-H.Gu, "Online subspace learning on Grassmann manifold for moving object tracking in video", Proc. ICASSP, pp. 969-972, 2008. [2](#)
- [13] A. Tyagi and J. W. Davis, "A recursive filter for linear systems on Riemannian manifolds", Proc. IEEE CVPR, pp. 1-8, 2008. [2](#)
- [14] L. Dong, L. Tao and G. Xu, "Head Pose Estimation Using Covariance of Oriented Gradients", Proc. ICASSP, pp. 1470-1473, 2010. [2](#)
- [15] R.Subbarao and P.Meer, "Nonlinear mean shift over Riemannian manifolds", Int. J. Comput. Vis., vol. 84, no. 1, pp. 1-20, 2009. [2](#), [3](#)
- [16] D.Ross, J.Lim, R.S.Lin, M.H.Yang, "incremental learning for robust visual tracking", Int. J. Comput. Vis., 77(1), pp.125-141, 2008. [5](#)
- [17] D.Comaniciu, V.Ramesh and P.Meer, "Kernel-based object tracking", IEEE Trans. Pattern Anal. Mach. Intell., vol. 5, pp. 564-577, 2003. [5](#)
- [18] M.Isard and A.Blake, "CONDENSATION - Conditional Density Propagation for Visual Tracking", International Journal of Computer Vision, vol. 29, no. 1, pp. 5-28, 1998. [5](#)
- [19] J.K. Kamarainen, V. Kyrki and H.Kalviainen, "Invariance properties of Gaborfilter-based features-Overview and applications", IEEE Trans. Image Processing, vol. 15, no. 5, pp. 1088-1098, 2006. [3](#)