Chalmers Publication Library

Visual Tracking and Dynamic Learning on the Grassmann Manifold with Inference from a Bayesian Framework and State Space Models

(article starts on next page)

VISUAL TRACKING AND DYNAMIC LEARNING ON THE GRASSMANN MANIFOLD WITH INFERENCE FROM A BAYESIAN FRAMEWORK AND STATE SPACE MODELS

Zulfiqar Hasan Khan,   Irene Yu-Hua Gu

Dept. of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden
{zulfiqak, irenegu}@chalmers.se

ABSTRACT

We propose a novel visual tracking scheme that exploits both the geometrical structure of Grassmann manifold and piecewise geodesics under a Bayesian framework. Two particle filters are alternatingly employed on the manifold. One is used for online updating the appearance subspace on the manifold using sliding-window observations, and the other is for tracking moving objects on the manifold based on the dynamic shape and appearance models. Main contributions of the paper include: (a) proposing an online manifold learning strategy by a particle filter, where a mixture of dynamic models is used for both the changes of manifold bases in the tangent plane and the piecewise geodesics on the manifold, (b) proposing a manifold object tracker by incorporating object shape in the tangent plane and the manifold prediction error of object appearance jointly in a particle filter framework. Experiments performed on videos containing significant object pose changes show very robust tracking results. The proposed scheme also shows better performance as comparing with three existing trackers in terms of tracking drift and the tightness and accuracy of tracked boxes.

***Index Terms*—** visual tracking, manifold tracking, manifold learning, Grassmann manifold, piecewise geodesics, particle filter, state space modeling

1. INTRODUCTION

Tracking visual objects on curved surfaces or manifolds has drawn increasing interests recently. [1, 2] shows that subspaces actually reside on a curved surface or differentiable manifold, and online learning of subspace can be better characterized by considering the temporal consistency of time-varying subspaces and the geometrical structure of the manifold. [1] uses conjugate gradient and Newton's method for subspace tracking on the Grassmann and Stiefel manifolds and applied to orthogonal procrustes. [2] proposes piecewise geodesics on the complex Grassmann manifold using projection matrices for subspace tracking with simulations to synthetic signals from an array of sensors. [7] utilizes particle filters (PFs) on the Riemannian manifold to estimate the target position and time-varying noise covariance with simulations on trajectories of 2D point targets. [3] proposes visual

tracking by applying a Kalman filter to the velocity of basis matrix in the tangent plane of Grassmann manifold. The method can track objects with moderate pose changes however significant pose changes remain a challenging issue. [4] proposes covariance tracking on the symmetric manifold. An exhaustive search which is computationally expensive is used to find the best match and strategy for model update is proposed. Some good results are reported. Other variants of covariance tracking are proposed, for example, [5] employs a Log-Euclidean metric on a symmetric manifold for covariance tracking of visual objects. [8] proposed nonlinear mean shift on Riemannian manifolds for image segmentation and nonlinear filtering. Other recent work on manifold face tracking are reported, e.g. using an offline manifold training strategy from face datasets containing different poses [9], and online learning using local linearity of the appearance manifold using particle filters with a coarse-to-fine factorized sampling [10]. Despite these promising work, challenges remain in tracking visual objects that contain significant pose changes. Computational time is also an issue of concern in terms of manifold visual tracking.

Motivated by these issues, we propose a novel scheme for robust visual object tracking on the Grassmann manifold. The main contributions of the proposed tracking scheme include: (a) propose an online manifold learning strategy by a particle filter, where a mixture of dynamic models is used for both the changes of manifold bases in the tangent plane and the piecewise geodesics on the manifold. (b) propose a manifold object tracker by incorporating object shape in the tangent plane and the manifold prediction error of object appearance jointly in a particle filter framework.

2. PROBLEM FORMULATION: THE BIG PICTURE

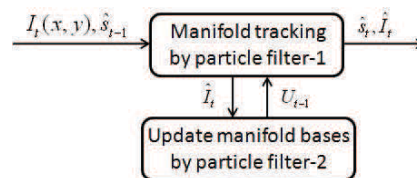**Fig. 1**. Block diagram of proposed scheme. The notations $I_t(x, y)$, $\hat{s}_t$, $\hat{I}_t$ denote the image frame, bounding box parameters, and tracked object image at time t, respectively.

The proposed scheme consists of two major parts: (a) vi-

sual tracking on the Grassmann manifold, (b) updating the manifold basis matrix in a fixed time interval. The essence for visual tracking in (a) (top block in Fig.1) is to apply a dynamic object shape model in the tangent plane of the manifold, as well as an object appearance model for the prediction error on the manifold. A particle filter is utilized where the affine shape parameters are defined as the state vector while the appearance are embedded as the likelihood, given a previous manifold basis matrix $\mathbf{U}_{t-1}$. The basic idea for updating the manifold basis matrix $\mathbf{U}_t$ at t in (b) (bottom block in Fig.1) is to define a state space model containing two mixtures: the velocity of manifold basis matrix lies in the tangent plane whose dynamic is described by a constant velocity model, the basis matrix is on the manifold whose dynamic is modeled by piecewise geodesics on the manifold. To realize the above mixture model, another particle filter is utilized. It is known that the pose changes of a 3D object in 2D image plane cannot be well described by a single vector space, rather, the trajectory of pose changes is more suitable to be described as points on a smoothed curved surface or a differentiable manifold. The main motivations and novelties lie on online updating the basis matrix on the manifold by using not only the previous manifold basis matrix, but also the velocity indicating the change of basis matrices, in addition to using manifold-based tracking rather than linear vector space-based tracking. This would lead to more robust tracking.

3. MANIFOLDS AND PARTICLE FILTERS: REVIEW

This section briefly reviews the Grassmann manifold, two mapping functions and particle filters (PFs) that are used in the proposed tracking scheme.

3.1. Grassmann Manifold

A Grassmann manifold $\mathbf{g}_{n,k}$ is the collection of all k-dimensional subspaces in the n-dimensional $\mathcal{R}^n$. $\mathbf{g}_{n,k}$ can be viewed as the quotient space $SO(n)/(SO(k) \times SO(n-k))$, where $SO(n)$ is the Special Orthogonal Group whose elements are $n \times n$ matrices. The Grassmann manifold $\mathbf{g}_{n,k}$ can be considered as a smoothing curved surface. Each point on $\mathbf{g}_{n,k}$ can be uniquely represented by the projection matrix $U(U^TU)^{-1}U^T$, where U is an orthonormal basis of size $n \times k$. Alternatively, one may use $U \in \mathcal{R}^{n \times k}$ contained in the set $\{UH : H \in SO(k)\}$ as the equivalent, since the orbit of U spans the same subspaces [12]. The main advantage of using $U = \{UH : H \in SO(k)\}$ is the computational efficiency compared to projection matrices.

Two important mapping functions [8] performed between the manifolds to the tangent planes are briefly summarized below.

Exponential map ($\mathcal{T}_{\mathbf{x}} \rightarrow \mathbf{g}_{n,k}$): Given $\mathbf{x}$, the starting point (t=0), on the manifold and the corresponding tangent Δ in the tangent space $\mathcal{T}_{\mathbf{x}}$, the exponential function in (1) maps the tangent vector along the geodesic to yield the end point $\mathbf{y} = \mathbf{x}(1)$ on the manifold,

$$\exp_{\mathbf{x}}(\Delta) = \mathbf{x}\mathbf{v}\cos(\mathbf{s})\mathbf{v}^T + \mathbf{u}\sin(\mathbf{s})\mathbf{v}^T \qquad (1)$$

where, $\mathbf{u}\mathbf{s}\mathbf{v}^T$ is the compact SVD of Δ and the sin and cos act element-by-element along the diagonal $\mathbf{s}$.

Logarithmic map ($\mathbf{g}_{n,k} \rightarrow \mathcal{T}_{\mathbf{x}}$): Given two points $\mathbf{x}$, $\mathbf{y}$ on the manifold, it find the velocity vector Δ in the tangent space on the point $\mathbf{x}$ as follows:

$$\Delta = \log_{\mathbf{x}}(\mathbf{y}) = \mathbf{u}\sin^{-1}(\mathbf{s})\mathbf{v}^T \qquad (2)$$

where $\mathbf{u}\mathbf{s}\mathbf{d}^T = \mathbf{y} - \mathbf{x}\mathbf{x}^T\mathbf{y}$ and $\mathbf{v}\mathbf{c}\mathbf{d}^T = \mathbf{x}^T\mathbf{y}$ is the generalized SVD with $\mathbf{c}^T\mathbf{c} + \mathbf{s}^T\mathbf{s} = \mathbf{I}$ and the $\sin^{-1}$ acts element-by-element along the diagonal of $\mathbf{s}$. The two mapping functions satisfy $\mathbf{x}^T \log_{\mathbf{x}}(\mathbf{y}) = 0$ and $\exp_{\mathbf{x}}(\log_{\mathbf{x}}(\mathbf{y})) = \mathbf{y}$.

Distance: The distance between $\mathbf{x}$ and $\mathbf{y}$ on the manifold is defined by the principal angle between the two subspaces [1]:

$$D(\mathbf{x}, \mathbf{y}) = \|\theta\| \qquad (3)$$

where $\theta = [\theta_1 \cdots \theta_k]^T$ is the principal angles between the subspaces associated with the manifold points $\mathbf{x}$ and $\mathbf{y}$. Detailed information can be found in [1].

3.2. Particle Filters (PFs)

PFs tracking, as a recursive Bayesian estimation, is formulated through estimating the posterior probability of state vector using the rule of propagation of state density over time,

$$p(s_t|z_{0:t}) \propto p(z_t|s_t) \int p(s_t|s_{t-1})p(s_{t-1}|z_{0:t-1})ds_{t-1} \qquad (4)$$

where s_t is the state vector at time t, $z_{0:t}$ is the observations (image pixels with the bounding box) up to t. Using a weighted sum of randomly generated samples or particles drawn from a proposal distribution q, the posterior pdf estimate can be approximated as:

$$p(s_t|z_{0:t}) \approx \sum_i \omega_t^i \delta(s_t - s_t^i) \qquad (5)$$

where s_t^i is the ith particle, w_t^i is the weight, $\sum_i \omega_t^i = 1$, $i = 1, \cdots, N_p$ is the total number of particles.

4. OBJECT TRACKING BASED ON PARTICLE FILTERS AND GRASSMANN MANIFOLD

The basic idea for tracking is to apply a dynamic object shape model in the tangent plane of the manifold by using particle filters, an object appearance model by the prediction error on the manifold, and by embedding this as the likelihood in the PFs. This model is different from [6] in terms of utilizing the manifold and the inference of its tangent plane.

Object shape model in the tangent plane: Let the state vector $s_t = [y_t^1 \ y_t^2 \ \beta_t \ \gamma_t \ \alpha_t \ \phi_t]^T$ be defined as vector with six parameters (2D box center, scale, rotation, aspect ratio, and skew) of affine transformed object bounding box, and assuming the object shape in the state space is modeled by the Brownian motion, then the dynamic shape between the two states is Gaussian distributed, that is:

$$\mathbf{s}_t = \mathbf{s}_{t-1} + \mathbf{v}, \qquad \mathbf{s}_t : \ p(\mathbf{s}_t|\mathbf{s}_{t-1}) \sim N(0, \mathbf{Q}) \qquad (6)$$

where $\mathbf{Q} = diag(\sigma_{y_1}^2, \sigma_{y_2}^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_\alpha^2, \sigma_\phi^2)$ whose diagonal elements correspond to the variances of individual 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 (6).

Object appearance model on the Grassmann manifold:

For each candidate bounding box specified by the particle $\mathbf{s}_t^j$, the corresponding image $\mathbf{I}_t^j$ is extracted. The likelihood is modeled as the Gaussian distributed dynamic prediction error on the manifold bases by,

$$p(\mathbf{d}_{\mathbf{I}_t^j}|s_t^j) = \exp(-\|\mathbf{d}_{\mathbf{I}_{tj}} - \mathbf{U}_{t-1}\mathbf{U}_{t-1}^T\mathbf{d}_{\mathbf{I}_t^j}\|^2/\sigma^2) \quad (7)$$

where $\mathbf{d}_{\mathbf{I}_t^j} = (\mathbf{I}_t^j - \hat{\mathbf{I}}_{t-1})$ is a column scanned vector defined as the difference between the candidate image region $\mathbf{I}_t^j$ and the reference image region $\hat{\mathbf{I}}_{t-1}$ at time $(t-1)$, $\mathbf{U}_{t-1}$ is the bases for the manifold at $(t-1)$, and σ^2 is the variance (empirically determined). Noting that this likelihood is proportional to the prediction error dynamics on the manifold bases. The PF weight is then assigned as the likelihood, $w1_t^j = p(\mathbf{d}_{\mathbf{I}_t^j}|s_t^j)$, with SIS resampling [11]. Finally, the ML (maximum likelihood) estimate of object bounding box is:

$$\hat{\mathbf{s}}_t = \mathbf{s}_t^{j^*} \qquad \text{where: } j^* = \text{argmax}_j(w1_t^j) \quad (8)$$

5. UPDATING MANIFOLD BASES USING DYNAMIC MODELS IN TANGENT PLANES AND MANIFOLDS

The basic idea for updating the manifold bases $\mathbf{U}_t$ at t is to define a state space model containing a mixture of 2 separate ones: the velocity of manifold bases lies in the tangent plane whose dynamic is model by a constant velocity model, the basis matrix is on the manifold whose dynamic is modeled by piecewise geodesics on the manifold. To realize the above model, a second PF is utilized for estimating the posterior pdf. Let the state vector be defined as $\mathbf{s2}_t = [\mathbf{U}_t, \mathbf{\Delta}_t]^T$, $\mathbf{U}_t$ be the basis matrix on the Grassmann manifold $\mathbf{g}_{n,k}$, $\mathbf{\Delta}_t$ be the corresponding velocity for $(\mathbf{U}_{t-1}, \mathbf{U}_t)$ where $\mathbf{U}_t$ is on the end point of the geodesic starting from $\mathbf{U}_{t-1}$, and the observation $\mathbf{Y}_t$ at t be defined as a sliding window (size L) of tracked object images $\mathbf{Y}_t = [\hat{\mathbf{I}}_{t-L+1} \cdots \hat{\mathbf{I}}_t]$ (where $\hat{\mathbf{I}}_t$ is tracked image area (see Section 4)). The following mixture model is applied to the state vector $\mathbf{s2}_t = [\mathbf{U}_t, \mathbf{\Delta}_t]^T$:

$$\begin{aligned} \text{Constant velocity:} & \quad \Delta_t = \Delta_{t-1} + \mathbf{V}_2 \\ \text{Piecewise geodesic:} & \quad \mathbf{U}_t = \mathbf{g}_{n,k}(\mathbf{U}_{t-1}, \Delta_t) \end{aligned} \quad (9)$$

where $\mathbf{V}_2$ is zero-mean white noise with a Gaussian distribution $\mathcal{N}(0, \sigma_{v_2}^2)$. Noting that the first model is defined in the tangent plane, and the 2nd on the Grassmann manifold. For the first state variable, predicted particles Δ_t^i are generated on the manifold point at (t-1) according to the 1st equation in (9) with $\sigma_{v_2}^2$ ($\sigma_{v_2}^2 = .01$ in our tests). For the 2nd state variable, particles $\mathbf{U}_t^i$ are obtained by mapping Δ_t^i to the manifold, using $\mathbf{U}_{t-1}$ and (2) under the piecewise-geodesic model. The particle weight $w2_t^i$ is assigned as the likelihood:

$$w2_t^i = p(\mathbf{U}_{Y_t}|\mathbf{U}_t^i) = exp(-(D_t^i/\sigma_l)^2) \quad (10)$$

where D_t^i is the distance between the principal angles of $\mathbf{U}_{Y_t}$ (i.e. bases for the new observation $\mathbf{U}_{Y_t}$) and $\mathbf{U}_t^i$ (i.e. predicted bases from particles), as defined in (3). The MMSE estimate of the basis matrix $\mathbf{U}_t$ is obtained by first computing $\Delta_{\mathbf{U}_{Y_t}}^i$ using $\log_{\mathbf{U}_{Y_t}}(\mathbf{U}_t^i)$ in (2), then estimating the sample mean $\hat{\Delta}_t$ and finally the new basis matrix $\mathbf{U}_t$ as follows:

$$\hat{\Delta}_t = \sum_{i=1}^{N_2} w2_t^i \Delta_{\mathbf{U}_{Y_t}}^i / N_2, \qquad \mathbf{U}_t = exp_{\mathbf{U}_{Y_t}}(\hat{\Delta}_t) \quad (11)$$

It is worth mentioning that one may decide to update the basis matrix in each m sample interval rather than the unit time interval as described above. The advantage is the reduced computation, however, the error for the piecewise geodesic model increases. Table 1 summarizes the pseudo algorithm.

Table 1. Pseudo algorithm for the proposed tracking scheme

Initialization:: choose an object region (for frame t=1), Track first L
 frames by template matching and compute U_{Y_t}; Generate $\mathbf{s}_t^j$,
 $j = 1 \cdots N_1$, and $\mathbf{s2}_t^i$, $i = 1 \cdots N_2$ using prior information.
 for frame $t = L, L+1, \cdots$, do:
   (*particle filter-1 for object tracking*)
   for particle $j = 1, \cdots, N_1$ do:
   1. Predict particles s_t^j using (6);
   2. Update particle weights $w1_t^j$ by using (7);
   end{j}
   3. Compute the posterior state vector $\hat{s}_t$ using (8) and resampling;
   (*particle filter-2 for updating manifold basis matrix*)
 6. Form a new observation $\mathbf{Y}_t$ and compute $\mathbf{U}_{Y_t}$;
 for particle $i = 1, \cdots, N_2$ in PF_2 do:
   7. Predict particles $[\mathbf{U}_t^i, \Delta_t^i]$ by using (9);
   8. Update particle weights $w2_t^i$ by using (10);
 end{i}
 9. Compute $\hat{\Delta}_t$ and $\mathbf{U}_t$ using (11) and resampling;
 end {t}

6. EXPERIMENTAL RESULTS

Setup: Seven videos with significant object pose changes are selected for the tests. For all these tests, object regions are normalized to 32×32 pixels. $N_1 = 600$, $\sigma^2 = 0.25$ are set for PF_1. $N_2 = 400$, $\sigma_{v_2}^2 = 0.01$ are set for PF_2. Sliding window size is set to $L = 5$, $\sigma_l^2 = 0.1$ are used in (10). $\mathbf{Q}$ in (6) is set empirically, $[1, 1, .05, .05, .005, .001]$ for "chia" video, $= [5, 5, .01, .02, .002, .001]$ for "david", and $[9,5,.05,.05,.005,.001]$ for "dudek", "danny" and "MeerCar". **Results and comparisons:** Figs. 2, 3 4 show some tracking results in 5 videos (key frames) by the proposed scheme. Further, the proposed scheme is compared with 3 trackers: *tracker-a*: linear subspace tracking without online learning, *tracker-b*: covariance tracking in [4], and *tracker-c*: manifold face tracking [10]. The results from these trackers are included in the same figures for comparisons. Observing the resulting images, the proposed scheme has shown robust tracking for videos containing large pose changing objects with varying changing speed, also better tracking performance with more accurate boxes in terms of tightness and orientation as compared the 3 existing methods.

7. CONCLUSIONS

Tests on the proposed tracking scheme, consisting of visual tracking on the manifold and online manifold basis updating, has shown very robust tracking performance for objects containing moderate to large pose changes. The online updating of basis matrices of manifold by exploiting both the position and velocity is shown to be effective. The online tracking by integrating dynamic appearance and shape on the manifold and its tangent plane in single particle filter is efficient. Comparisons with three existing methods have provided further support to the robustness of the proposed scheme, and
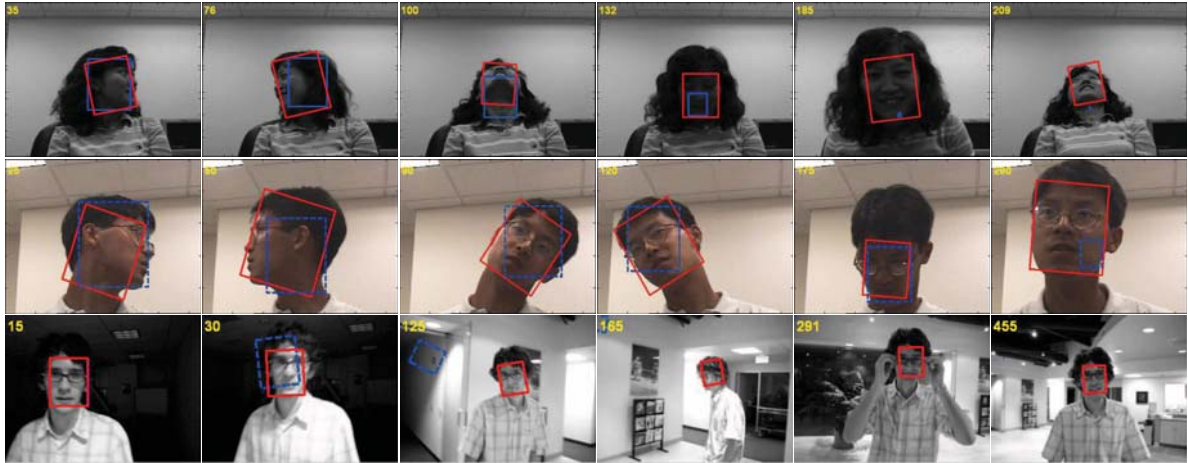
Fig. 2. Tracking results from video frames in "chia", "danny", and "david". Red box: from the proposed scheme, Blue box: from *tracker-a*.
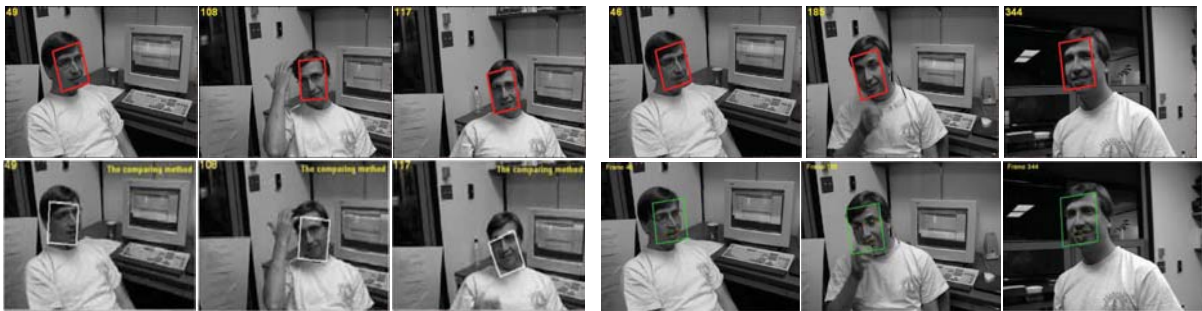

Fig. 3. Tracking results for video frames in "dudek". Row-1: from the proposed scheme, Row 2: *tracker-b* (columns 1-3) and *tracker-C* (columns 4-6).

Fig. 4. Tracking results for video frames "MeerCar". Top row: from the proposed scheme, Bottom row: from *tracker-b*.

relatively high speed (approx. 1 frame/sec in our Matlab program).

8. REFERENCES

[1] A.Edelman, T.A.Arias, S.T,Smith, "The geometry of algorithms with orthogonality constraints", *SIAM J. Matrix Anal. Appl.*, 20(2), 1998.

[2] A.Srivvastava and E.Klassen, "Bayesian and geometric subspace tracking", *Adv. Appl. Prob. (SGSA)*, vol. 36, pp. 43-56, 2004.

[3] T.Wang, A.G.Backhouse, I.Y-H.Gu, "Online subspace learning on Grassmann manifold for moving object tracking in video", *Proc. ICASSP* , 2008.

[4] F.Porikli, O.Tuzel, P.Meer, "Covariance tracking using model update based on Lie algebra", *Proc. CVPR*, pp. 728-735, 2006.

[5] X. Li, W. Hu, Z. Zhang et al, "Visual tracking via incremental Log-Eulcidean Riemannian subspace learning", *Proc. ICASSP*, 2008.

[6] 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.

[7] H. Snoussi and C. Richard, "Monte Carlo tracking on the Riemannian manifold of multivariate normal distributions", *Proc. Digital Signal Procesing Workshop*, pp.280-285, 2009.

[8] R.Subbarao and P.Meer, "Nonlinear mean shift over Riemannian manifolds", *Int. J. Comput. Vis.*, vol. 84, no. 1, pp. 1-20, 2009.

[9] 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.

[10] 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.

[11] A. Dore, M. Soto, C. Regazzoni, "Bayesian tracking for video analytics", *IEEE Trans. Image Process.*, 27(5), pp. 46-55, 2010.

[12] K. A. Gallivan, A. Srivastave, X. Liu, "Efficient algorithms for inferences on Grassmann manifolds", *Proc. IEEE workshop on Statistical Signal Processing*, pp.315-318, 2003.