ONLINE SUBSPACE LEARNING ON GRASSMANN MANIFOLD FOR MOVING OBJECT TRACKING IN VIDEO

Tiesheng Wang[(a)], *Andrew G. Backhouse*[(b)], *Irene Y.H. Gu*[(b)]

[(a)]Institute of IPPR, Shanghai Jiao Tong Univ., Shanghai, China
[(b)]Dept. of Signals and Systems, Chalmers Univ. of Technology, Sweden
tieshengw@sjtu.edu.cn, {andy, irenegu}@chalmers.se

ABSTRACT

This paper proposes a robust object tracking method in video where the time-varying principal components of object's appearance are updated online. Instead of directly updating the PCA-based subspace using matrix decomposition, the subspace is updated by tracking on the Grassmann manifold. The object tracker performs two alternating processes: (a) online learning of principal component subspace; (b) tracking a moving object using the learned subspace and a particle filter. Learning a PCA-based subspace is performed by treating principal component decompositions as noisy measurements. The measurements are mapped onto the Lie algebra of the Grassmann manifold. The direction of movement of the subspace is then tracked in the Lie algebra using a Kalman filter. The filtered output is then mapped back onto the Grassmann surface to update the principal component-based subspace. This produces a more reliable learning of the subspace.

Experiments have been conducted on face image sequences where heads were tilted in variable speed, partial face occlusion, along with changes in object depth and in illuminations. The results and evaluations have shown that the proposed method is robust against these changes when tracking moving objects.

Index Terms— time-varying subspace learning, Grassmann manifold, object tracking, Kalman filter, particle filter.

1. INTRODUCTION

Eigen-tracking [1], or tracking with Principal Component Analysis (PCA), is a frequently used method for tracking object regions in video. In the method, object tracking is accomplished by using lower-dimensional principal subspace of objects. While the principal subspace may well capture the linear variations in object appearances, it cannot describe the significant changes in appearance due to, e.g., illumination variations, shifting viewing angles, object deformation. Further, due to the lack of training data, the learning process is often done online incrementally by using the tracked object regions from the previous frames. In [2, 3] the object subspace is learned/updated directly from the sample covariance matrix, and then treats the obtained subspace as the true one. This will perform poorly when a small number of samples (object regions) are used, or when the statistics of samples are time-varying.

Motivated by this, we propose to learn/update the subspace of object appearance on the Grassmann manifold where the measurement noise is also taken into account. The time-evolving subspaces of object appearance correspond to the set of all k dimensional subspaces of an n dimensional Euclidean space. The set is the Grassmann manifold and is a curved surface. Learning and updating the subspace is therefore performed on this curved space. Description of the geometric properties of this space can be found in [4, 5] and the references therein.

Conventional learning/updating, e.g., using Kalman filters, is not designed for tracing movement on curved surfaces. It is worth noting that the tangent planes of curved surfaces are flat Euclidean spaces, furthermore, the velocity of motion at a point correspond to a vector in the point's tangent plane. Since tracking velocity is equivalent to tracking motion itself, it implies that efficient tracking and learning can be achieved in the tangent plane.

In [5], a subspace tracking of a point-wise signal was presented by using a particle filter on a Grassmann manifold. However, this cannot be directly applied to tracking video objects since the state vector characterizing the appearance of a moving 2D object has a rather high dimensionality. This makes a particle filter practically unfeasible due to the number of particles required. To overcome this problem, we propose iteratively online learning on the Grassmann surface for the PCA-based subspace of moving object by using a Kalman filter, and tracking a moving object by using a particle filter. This is performed by tracking the velocity of the subspace in the Lie algebra of the Grassmann manifold and mapping the velocity correctly into motion on the Grassmann surface.

It is worth noting that the mapping between velocity and motion on the Grassmann manifold can be computed owing

This work was partly supported by Swedish-Asian Research Links Program under Swedish Sida/VR grant number 348-2005-6095.

to the group structure of this manifold. Manifolds having a group structure are referred to as Lie groups. Lie groups have the additional benefit that all tangent planes of points in the Lie group are trivially related to the Lie algebra.

The rest of this paper is organized as follows. Section 2 reviews some basics on Lie groups and the Grassmann manifold. Subspace learning/updating is formulated in Section 3. Section 4 describes tracking visual object with time-evolving subspace. Section 5 describes the experiments with some results and evaluations included.

2. LIE GROUP AND GRASSMANN MANIFOLD

Lie Groups: Lie groups [6] are smooth surfaces (analytic manifolds) that have a group structure. This implies that there exists an operator $\oplus$ such that for any X,Y in the group $\mathcal{G}$, $X \oplus Y$ is also a member of $\mathcal{G}$. Furthermore, each element X has an inverse X^{-1} such that $X^{-1} \oplus X = E$ where E is the identity element.

A most important property of Lie groups used in this work is that the tangent planes for all points in a Lie group are related. If T_X is the tangent plane at X, and $f(t)$ is a path passing though X, then $X^{-1} \oplus f(t)$ passes through the identity and therefore has a derivative which lies in the tangent plane of the origin T_E. Similarly, all tangent planes can be related to the tangent plane at the origin due to this group structure. The direction and speed of two different paths are related in the tangent plane at the origin. A tangent plane T_E is called the Lie algebra $\mathfrak{g}$. Another important property of Lie groups exploited in this work is the exponential mapping, which maps points on the Lie algebra onto the Lie group. Since the Lie algebra is associated with a tangent plane that contains velocity vectors, the exponential mapping maps velocity vectors into the motion. It is worth mentioning that for Lie groups in general, the exponential map has a complex form.

Special Orthogonal Group: The special orthogonal group $SO(n)$ is relevant for the study of the Grassmann manifold. $SO(n)$ is the group of $n \times n$ orthogonal matrices. For this Lie group, the Lie algebra can be represented as the space of skew-symmetric matrices. Further, the exponential mapping $\mathfrak{g} \mapsto SO(n)$ for this space is given by

$$U = \exp(A) = I + \sum_{k=1}^{\infty} A^k/k! \tag{1}$$

where A is a skew-symmetric matrix.

Grassmann Manifold: The Grassmann manifold, denoted as $\mathcal{G}_{n,k}$, is the set of all k-dimensional subspaces of $\mathbb{R}^n$. Defining the group structure for $\mathcal{G}_{n,k}$ is not as easy as it is for the special orthogonal group. There exists no useful representation of the members of $\mathcal{G}_{n,k}$ which makes a simple expression of group addition possible.

The member in $\mathcal{G}_{n,k}$ can be represented non-uniquely as basis matrix $U_{n,k}$, or uniquely as a projection matrix $P = U_{n,k}^T(U_{n,k}U_{n,k}^T)^{-1}U_{n,k}$. No matrix multiplication can be used for group addition in these two types of representations.

The tangent space (or, the Lie algebra) for the Grassmann manifold can be obtained by studying $U_{n,k}$. Let $U \in SO(n)$ and the first k columns of U is formed by $U_{n,k}$. The rotation of the right $n-k$ columns of U does not affect $U_{n,k}$ and the rotation of the first k columns does not change the subspace. Therefore, $\mathcal{G}_{n,k}$ is equivalent to the quotient group $SO(n)/(SO(k) \times SO(n-k))$. This group has a Lie algebra corresponding to skew symmetric matrices A given by $A = \begin{bmatrix} 0 & X \\ -X^T & 0 \end{bmatrix}$, where X is a matrix of size $k \times (n-k)$, the identity element of this group is $Q = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}$.

The map from U to the projection matrix P is through $P = UQU^T$. The manifold formed by the projection matrices is diffeomorphic to $\mathcal{G}_{n,k}$ and the path $P(t) = e^{(At)}Qe^{(-At)}$ where $t = 0, \cdots, 1$, that starts at Q and ends at P is a geodesic which is the shortest path between Q and P. The matrix X in A can be seen as the geodesic direction and velocity matrix of the geodesic from Q to P.

Given two projection matrices P_{t-1} and P_t for two consecutive time intervals, the problem is to find the piece-wise geodesic direction matrix X_t. A similar method as in [5] is used. Initially, the geodesic direction matrix X_0 is obtained through an eigen decomposition of $Q - P_0$, where P_0 is the initial projection matrix. For a given time t, a new geodesic direction X_t between P_{t-1} and P_t is obtained through the following steps: using $U_{t-1} = e^{A_{t-1}}$ to rotate P_{t-1} to the identity element $Q = U_{t-1}^T P_{t-1} U_{t-1}$, and rotate P_t to $P = U_{t-1}^T P_t U_{t-1}$. Then, computing the corresponding geodesic direction X_t from $Q - P$ in a similar way as that in the initial step. Finally, obtaining U_t through $U_t = U_{t-1} e^{A_t} U_{t-1}^T$.

3. SUBSPACE LEARNING ON THE GRASSMANN MANIFOLD

Since the basis matrix $U_{n,k}$ of subspace is a point on the Grassmann manifold, updating subspace can be formulated as inference/filtering on the Grassmann manifold.

3.1. Define the State-Space on the Grassmann Manifold

Based on the geometrical structure of Grassmann manifold, updating the subspace can be briefly summarized as: first, the observed projection matrix is mapped into the associated Lie algebra and updated along the geodesic direction in the Lie algebra; then the updated geodesic direction is mapped back to the Grassmann surface to obtain the updated subspace.

We assume that the piece-wise geodesic direction of the subspace is a constant up to a difference of Gaussian noise between the two consecutive time interval. Then, estimation of the posterior subspace velocity can be converted into the Lie algebra and hence formulated by a state-space model. Let

X_t be the state vector (i.e., velocity along the true geodesic direction of the projection matrix P_t) at time t, and Z_t be the observed noisy velocity along the geodesic direction, $A_t = \begin{bmatrix} 0 & Z_t \\ -Z_t^T & 0 \end{bmatrix}$ be the corresponding element of the noisy projection matrix in Lie algebra. Since A_t is skew-symmetric block diagonal, there is only $k(n-k)$ degree of freedom. Therefore, we only consider the geodesic directions Z_t and X_t for updating/filtering.

Let the observations up to the time t be $Z_{1:t} = \{Z_1, ..., Z_t\}$, the observed training object samples containing m tracked image blobs from $t-m+1$ to t be $Y_t = \{y_{t-m+1}, \cdots, y_t\}$. The projection matrix P_t can be formulated by applying SVD to the sample covariance matrix S_t of Y_t and then the definition of projection. This projection matrix corresponds to the noisy geodesic direction Z_t in the Lie algebra. Assuming X_t and Z_t are i.i.d. distributed, we have the following state and observation equations,

$$\mathbf{x}_t = F\mathbf{x}_{t-1} + \mathbf{v}; \quad \mathbf{z}_t = H\mathbf{x}_t + \mathbf{w} \tag{2}$$

where $\mathbf{v} \sim \mathcal{N}(0, Q_v)$ is the model noise, $\mathbf{w}$ is the observation noise $\mathbf{w} \sim \mathcal{N}(0, Q_w)$, Q_v and Q_w are the covariance matrices, and $\mathbf{x}_t$ and $\mathbf{z}_t$ are the equivalent vector representation of the matrices X_t and Z_t.

The true geodesic direction of the projection matrix P_t (i.e. the posterior estimation of state vector) can be obtained by applying the rule of propagation of state density over time,

$$p(\mathbf{x}_t|\mathbf{z}_{1:t}) = \frac{1}{C} p(\mathbf{z}_t|\mathbf{x}_t) \int p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}) d\mathbf{x}_{t-1}$$

where C is a constant for normalization.

3.2. Kalman Subspace Updating on Grassmann Manifold

Since the changes of subspace between any two consecutive frames are usually smooth, it is reasonable to assume that the piece-wise geodesic directions between two consecutive time instants is a constant up to the difference of a Gaussian noise. Based on the above assumption, F and H in (1) become the identity matrices. Further, under i.i.d. white noise assumption of the state-space model, $Q_v = \sigma_v^2 I$ and $Q_w = \sigma_w^2 I$. From these, the Kalman filter updating formula can be simplified,

$$\begin{aligned} &\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_{t|t-1}; \;\; P_{t|t-1} = P_{t-1|t-1} + \sigma_v^2 I; \\ &K_t = P_{t|t-1}(P_{t|t-1} + \sigma_w^2 I)^{-1}; \\ &\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + K_t \left(\mathbf{z}_t - \hat{\mathbf{x}}_{t|t-1}\right); P_{t|t} = (I - K_t) P_{t|t-1} \end{aligned} \tag{3}$$

where K_t is the gain of Kalman filter, and $P_{t|t}$ is the posterior estimation error covariance matrix of $\mathbf{x}_t$. Given the filtered velocity vector $\hat{\mathbf{x}}_{t|t}$, the corresponding matrix form is $\hat{X}_{t|t}$. Mapping from $\hat{A}_t = \begin{bmatrix} 0 & \hat{X}_{t|t} \\ -\hat{X}_{t|t}^T & 0 \end{bmatrix}$ in Lie algebra to the member U_t in $SO(n)$ is through the exponential map in (1) and $U_t = U_{t-1} e^{A_t} U_{t-1}^T$. The basis matrix $U_{n,k,t}$ is formed by the first k columns of U_t.

4. OBJECT TRACKING

Online subspace learning and object region tracking are performed alternatively. First, object region in the frame t is tracked by eigen-tracking given the subspace associated with U_{t-1}. Then, the newly tracked object region is used to update the subspace associated with U_t (see Section 3). We shall briefly describe the object tracking method below due to space limitation.

For tracking, each region is normalized into a fixed size. The motion of object region is represented by affine transformation with a parameter vector $l_t = [d_{x,t}\, d_{y,t}\, \theta_t, s_t]^T$, where $d_{x,t}$ and $d_{y,t}$ are the translation of region center, θ_t is the rotation and s_t is the scaling. The conditional probability is $p(y_t|l_t) \propto \exp(||y_t - U_{n,k,t-1}^T y_t U_{n,k,t-1}||^2/\sigma^2)$, where y_t is the image region related to l_t and σ^2 is the variance (chosen empirically), $p(l_t|l_{t-1})$ is assumed i.i.d. Gaussian distributed. A particle filter (implemented by the CONDENSATION algorithm [7]) is then used to track the object region by sampling particles from $p(l_t|l_{t-1})$ and assigning the weights according to $p(y_t|l_t)$. The proposed algorithm, including online subspace learning/updating and object region tracking, is summarized in Table 1.

Table 1. The proposed algorithm for object tracking

1. *Initialization*
   Track the first f frames with template matching. Initialize P_0 using the tracked object blobs.

FOR $t = f + 1$ to (the total number of frames)

2. *Tracking the object region*: Given U_{t-1}, apply the particle filter for tracking, results in a tracked object region y_t.
3. *Learning/Updating the subspace of object appearances*
   - Mapping P_t in the Lie group (Grassmann manifold) to the geodesic direction Z_t in the Lie algebra (see Section 2).
   - Updating $\hat{X}_{t|t}$ with the Kalman filter using (3).
   - Mapping $\hat{A}_{t|t}$ back to the Lie group (Grassmann manifold) with the exponential mapping using (1).
   - Updating the subspace with U_t (see Section 3.2).

$t \leftarrow t + 1$

END (FOR)

5. EXPERIMENTS AND RESULTS

Image sequences containing tilted face images (with small out-of-plane rotation) with variable speed of tilting, with nearly constant or large illumination changes, partial occlusions, large object depth changes were used for our experiments. In the experiments, the size of object regions was normalized to 32×32, and the initial object regions in the first $f = 5$ frames were tracked by template matching. Online subspace learning and particle filter tracking were then run alternatively to the remaining frames. The time interval for updating the subspace was $m = 5$ frames. The subspace dimension (i.e., the

Fig. 1. Face tracking results with the proposed method (marked with a green box). The frame number is shown on the top left corner of each image. Rows 1 to 5: from video containing a tilting head and nearly constant lighting; from video containing a tilting head and significant illumination changes; from video containing a tilting head with partial occlusion; from video containing a tilting head moving forward and backward; from video where the speed of head tilting changes significantly.

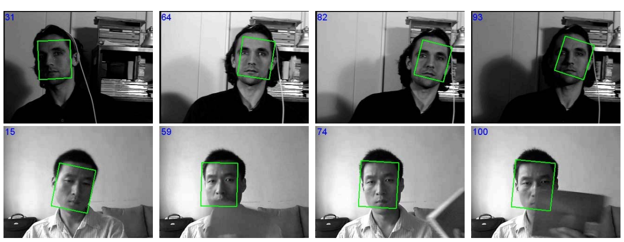

Fig. 2. Comparisons: tracking results by directly updating PCA-subspace. Rows 1 and 2 correspond to rows 2 and 3 in Fig.1.

number of basis vectors) was $k = 4$. The variances for the motion parameters were chosen before the tracking. P_t was computed from $Y_t^T Y_t$ instead of $Y_t Y_t^T$ for computational efficiency.

Fig.1 shows some examples of the test results. By visual inspection of the tracked object sequences as well as Fig.1, the method is shown to have successfully tracked moving faces, furthermore, the boxes for the tracked objects are rather tight in all tested cases. For comparisons, Fig.2 shows examples from tracking two videos, through directly updating PCA-subspace without tracking the subspace on the Grassmann manifold. The same parameter setting was used but without the Kalman filter in Section 3.2. Test results have shown that there is gradual loss of tracking probably due to illumination changes and occlusions. Our comparisons have shown that the proposed method is clearly more robust in tracking.

Fig.3 shows two evaluations to the proposed method. The first is the distance versus image frame, where the distance is defined as the number of pixels between the center of the tracked face region and the manually marked nose tip. One can observe some fluctuations of distances due to motion of face. In the 2nd evaluation, the squared error of object region reconstruction versus the image frame, where the error is defined between the original face in the tracked image region and the reconstructed face using the online learned subspace at the related time instant. One can observe some fluctuations in the squared errors although this fluctuations did not seem to directly impact the tracking results.

Overall, the tracking is shown to be rather good in all our tests and is robust to the types of changes (variable head tilting speed and head depth changes, occlusions, significant changes in illuminations).

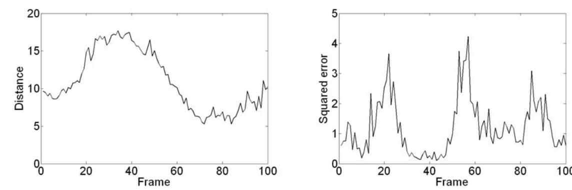

Fig. 3. Evaluation of the proposed method. Left: the distance between the nose tip to the tracked box center versus image frame; Right: the squared error between the original face and the reconstructed face. The video in the 2nd row of Fig.1 was used for the evaluations.

6. CONCLUSIONS

The proposed method of online subspace learning for object appearance through tracking the motion of the subspace on the Grassmann manifold has made online PCA-based subspace learning more reliable. The online subspace learning and object tracking, performed alternatively, has shown to generate robust tracking performance for tracking moving objects containing tilted faces with variable tilting speed, partial face occlusions, large face depth changes in video. Comparing with the method without using subspace updating on the Grassmann manifold, the proposed method is shown to have yielded more robust tracking performance.

7. REFERENCES

[1] M.J.Black, A.D.Jepson, Eigentracking: robust matching and tracking of articulated objects using a view-based representation, IJCV, vol.26, no.1, pp.63-84, 1998.

[2] D.Ross, J.Lim, M-H.Yang, Adaptive probabilistic visual tracking with incremental subspace update, ECCV'04, 2004.

[3] J.Lim, D.Ross, R-S.Lin, M-H.Yang, Incremental learning for visual tracking, Proc. NIPS'04, 2004.

[4] A.Edelman, T.A.Arias, S.T.Smith, The geometry of algorithms with orthogonality constraints, SIAM J. Matrix Anal. Appl. vol.20, no.2, pp.303-353, 1998.

[5] A.Srivastava, E.Klassen, Bayesian, geometric subspace tracking, Advances in Applied Probability, vol.36, no.1, pp.43-56, 2004.

[6] F. W. Warner, Foundations of differentiable manifolds and lie groups, Springer, 1983.

[7] M. Isard and A. Blake, CONDENSATION - conditional density propagation for visual tracking, IJCV, vol.29, no.1, pp.5-28, 1998.