# OBJECT TRACKING USING INCREMENTAL 2D-PCA LEARNING AND ML ESTIMATION

*Tiesheng Wang [a]*, *Irene Y.H. Gu [b]*, *Pengfei Shi [a]*

[a]Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China
[b]Dept. of Signals and Systems, Chalmers University of Technology, Sweden
{tieshengw, pfshi}@sjtu.edu.cn, irenegu@chalmers.se

## ABSTRACT

Video surveillance has drawn increasing interests in recent years. This paper addresses the issue of moving object tracking from videos. A two-step processing procedure is proposed: an incremental 2DPCA (two-dimensional Principal Component Analysis)-based method for characterizing objects given the tracked regions, and a ML (Maximum Likelihood) blob-tracking process given the object characterization and the previous blob sequence. The proposed incremental 2DPCA updates the row- and column-projected covariance matrices recursively, and is computationally more efficient for online learning of dynamic objects. The proposed ML blob-tracking takes into account both the shape information and object characteristics. Tests and evaluations were performed on indoor and outdoor image sequences containing a range of single moving object in dynamic backgrounds, which have shown good tracking results. Comparisons with the method using the conventional PCA were also made.

***Index Terms*—** object tracking, incremental 2DPCA, Maximum Likelihood estimation, video surveillance

## 1. INTRODUCTION

Video surveillance has drawn increasing interests in recent years. Among them, eigen-tracking is one of attractive techniques. In [1] eigen-tracking using multiple view representation was proposed. [2] introduced Rao-Blackwellized particle filter for eigen-tracking. To adapt to the variation of object appearance, [3] presented object tracking through incremental subspace learning from KL transformation, and [4] further proposed incremental updating of the mean. While eigen-tracking has achieved relatively good performance, its application is still limited due to the intensive computation.

In the eigen image representation, Principal Component Analysis (PCA) is used which requires heavy computation. This is due to computing eigenvectors from the covariance matrix associated with a set of 2D images. To reduce the computation, [5] proposed 2DPCA by using the covariance matrix through right side image projection. Further extended to using both side projections can be found in [6]. In [7], GLRAM was introduced for the low rank approximation of matrices on both sides. The two methods were shown to be mathematically equivalent [8]. However, further developing suitable for online learning and reduced PCA computation, in combination of efficient tracking model, are desirable for real-time tracking applications.

Motivated by the above, this paper proposes a two stage blob object tracking procedure, by using an incremental 2D-PCA for recursive object characterization and ML estimation for blob tracking. The former procedure is suitable for online learning and efficient computation, and the latter one is based on the probability of blob shape information and probability of object region characteristics.

The rest of the paper is organized as follows. In Section 2, the proposed 2-step object tracking is briefly described. In Section 3, after reviewing several PCA-based methods, we propose an incremental 2DPCA method which is more efficient for online learning. Section 4 describes the proposed ML blob-tracking of foreground moving objects. Section 5 describes the experiments, criteria for evaluation, and comparisons, with some results included. Finally, the conclusion is given in Section 6.

## 2. TRACKING BASED ON 2DPCA OBJECT CHARACTERIZATION AND ML ESTIMATION

For tracking foreground moving objects in the presence of dynamic background, we divide the problem into 2 separate, however, inter-related sub-issues. One is target object characterization through dynamic learning. Another is the object tracking assuming that the foreground object regions are provided together with previous learned objects.

For target object characterization, we propose to use incremental 2DPCA to characterize the image regions containing target objects. Under the assumption of separable kernels along rows and columns, recursive formulae are presented for an incremental 2DPCA algorithm. Comparing with the conventional PCA for 2D images, this provides a significant reduction in computation.

For blob-object tracking, we propose the use of ML (Maximum Likelihood) estimation for associating a current blob,

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

given the information of the previously tracked blob sequence of object (containing shape parameters), and the learned basis images from the 2DPCA of each objects.

## 3. 2DPCA-BASED IMAGE CHARACTERIZATION

This section briefly reviews the conventional PCA and 2DPCA. An incremental 2DPCA is then proposed for images.

### 3.1. 2DPCA for Image Characterization

A conventional PCA for 2D images (or, image regions) can be described as finding the principal eigenvectors from the covariance matrix of images [9]. For notational simplicity, an equivalent solution to the conventional PCA of 2D images $\mathbf{A}_i$ can be expressed by vector forms as follows. For 2D images sized $r \times c$, the sample covariance is first computed from $l$ images. Denoting 2D images $\mathbf{A}_i$ (size $r \times c$) by column-scanned vectors $\mathbf{a}_i$ (size $rc \times 1$). The sample covariance matrix can be computed from

$$\mathbf{C} = \frac{1}{l}\sum_{i=1}^{l}(\mathbf{a}_i - \overline{\mathbf{a}})(\mathbf{a}_i - \overline{\mathbf{a}})^T = \frac{1}{l}\mathbf{D}\mathbf{D}^T \quad (1)$$

where $\overline{\mathbf{a}} = 1/l\sum_{i=1}^{l}\mathbf{a}_i$ is the mean vector, $\mathbf{d}_i = \mathbf{a}_i - \overline{\mathbf{a}}$ is the mean subtracted image in vector, and $\mathbf{D} = [\mathbf{d}_1\ \mathbf{d}_2\ ...\ \mathbf{d}_l]^T$ is the sample data matrix . PCA requires finding eigenvectors associated with the first few largest eigenvalues of $\mathbf{C}$. Each eigenvector is sized $(rc \times 1)$ corresponds to a 2D basis image. Eigenvectors can be computed from using SVD, $\mathbf{D} = \mathbf{U}\Sigma\mathbf{V}^T$, where $\mathbf{U}$ contains eigenvectors of $\mathbf{D}\mathbf{D}^T$. Directly computing eigenvectors in the conventional PCA is computationally intensive since the size of $\mathbf{C}$ is large $(rc \times rc)$. In an online system, training images are usually obtained sequentially, and the image number $l$ is also large.

By assuming that 2D transform kernels are separable, PCA can be performed through row and column projected covariance matrices of 2D images. The method can be briefly described as follows. For the given 2D training images $\mathbf{A}_i$, $i = 1, ..., l$, combining right image projection $\mathbf{M}_i = \mathbf{A}_i\mathbf{R}$ and left image projection $\mathbf{M}_i = \mathbf{L}^T\mathbf{A}_i$ yields:

$$\mathbf{M}_i = \mathbf{L}^T\mathbf{A}_i\mathbf{R} \quad (2)$$

In the 2DPCA, the principal components are computed from the column- and row-projected covariance matrices:

$$\mathbf{G} = \frac{1}{l}\sum_{i=1}^{l}\mathbf{B}_i^T\mathbf{B}_i,\ \mathbf{F} = \frac{1}{l}\sum_{i=1}^{l}\mathbf{B}_i\mathbf{B}_i^T \quad (3)$$

where $\overline{\mathbf{A}} = 1/l\sum_{i=1}^{l}\mathbf{A}_i$ and $\mathbf{B}_i = \mathbf{A}_i - \overline{\mathbf{A}}$. The matrix $\mathbf{L}$ is computed from $(s)$ principal eigenvectors of $\mathbf{F}$, and $\mathbf{R}$ from $(t)$ principal eigenvectors of $\mathbf{G}$. Since the sizes of $\mathbf{G}$ and $\mathbf{F}$ are small, computing 2DPCA is efficient.

**Table 1**. Training using Incremental 2DPCA

| |
|---|
| Step-1: Recursively updating 2D mean image, row-projected covariance matrix using (6), and column-projected covariance matrix using (7). |
| Step-2: Computing the row and column projected principal components from $\mathbf{F}_r$ and $\mathbf{G}_r$. |
| Step-3: An outer product between each pair of row and column-eigenvectors corresponds to a basis image. |

### 3.2. Incremental 2DPCA Learning from Videos

To further reduce the computation and to make the 2DPCA more suitable for online processing, we introduce an efficient updating algorithm for the row and column projected covariance matrices. Suppose the previous 2DPCA is obtained by using $n$ training images, while 2DPCA is being learned after adding $m$ newly input images. In an incremental 2DPCA, the mean image, the row- and column-projected covariance matrices are recursively updated as follows. Assuming the original 2D image set is $\mathcal{A}_p = \{\mathbf{A}_1, ..., \mathbf{A}_n\}$, the newly added 2D image set is $\mathcal{A}_q = \{\mathbf{A}_{n+1}, ..., \mathbf{A}_{n+m}\}$, and the total image set at current time is $\mathcal{A}_r = \{\mathbf{A}_1, ...\mathbf{A}_{n+m}\}$. Denoting the corresponding mean images as $\overline{\mathbf{A}}_p$, $\overline{\mathbf{A}}_q$, $\overline{\mathbf{A}}_r$, and the row-projected covariance matrices for the above 3 sets as,

$$\mathbf{F}_p = \frac{1}{n}\sum_{i=1}^{n}\mathbf{B}_i\mathbf{B}_i^T, \mathbf{F}_q = \frac{1}{m}\sum_{i=1}^{m}\mathbf{B}_i\mathbf{B}_i^T, \mathbf{F}_r = \frac{1}{n+m}\sum_{i=1}^{n+m}\mathbf{B}_i\mathbf{B}_i^T \quad (4)$$

and the column-projected covariance matrices:

$$\mathbf{G}_p = \frac{1}{n}\sum_{i=1}^{n}\mathbf{B}_i^T\mathbf{B}_i, \mathbf{G}_q = \frac{1}{m}\sum_{i=1}^{m}\mathbf{B}_i^T\mathbf{B}_i, \mathbf{G}_r = \frac{1}{n+m}\sum_{i=1}^{n+m}\mathbf{B}_i^T\mathbf{B}_i \quad (5)$$

Then, it can be shown that the row- and column-projected covariance matrices can be updated recursively after $m$ new input images as follows:

$$\mathbf{F}_r = \frac{1}{m+n}(n\mathbf{F}_p + m\mathbf{F}_q + nm(\overline{\mathbf{A}}_p - \overline{\mathbf{A}}_q)(\overline{\mathbf{A}}_p - \overline{\mathbf{A}}_q)^T) \quad (6)$$

$$\mathbf{G}_r = \frac{1}{m+n}(n\mathbf{G}_p + m\mathbf{G}_q + nm(\overline{\mathbf{A}}_p - \overline{\mathbf{A}}_q)^T(\overline{\mathbf{A}}_p - \overline{\mathbf{A}}_q)) \quad (7)$$

This leads to the following incremental 2DPCA algorithm, summarized in Table 1. The incremental 2DPCA offers a significant reduction in computation as compared with the PCA of 2D images directly (using (1)) and an efficient sequential processing. In fact, one can obtain a basis image from 2DPCA by performing outer product between a row eigenvector and a column eigenvector. Performing outer products over all pairs of row and column eigenvectors generate the entire basis images. The only added assumption in the (incremental) 2DPCA is that the row and column transform kernels are separable.

## 4. STATISTICALLY-BASED BLOB-TRACKING

Once the characteristics of each specific object region in the previous image sequence is described by the incremental 2D-PCA, tracking is then performed for estimating the new blob position of the object. This can be formulated as the ML estimation problem. For the given sequence of $j$-th object blobs ($L_{1:t-1,j}$) and the 2DPCA description of learned $j$-th object $O_j^{Lib} \approx O_{1:t-1,j}$ from the previous frames $(1, ..., t-1)$, the conditional probability (or the likelihood) of the $i$-th candidate object $\tilde{O}_i$ in the blob region $L_{t,i}$ of the current frame $t$ (assuming isolated and non-overlapping objects) is described as,

$$p(L_{t,i}, \tilde{O}_i | L_{1:t-1,j}, O_j^{Lib}) \approx p(L_{t,i}|L_{1:t-1,j}) p(\tilde{O}_i|O_{1:t-1,j}) \quad (8)$$

where $(1, \cdots t-1)$ is simplified as $(1 : t-1)$. The first term is the probability of $i$-th blob which is based on the shape of $j$-th object blob sequence. The second term in (8) is related to the probability of error where $\mathbf{A}_{t,i}$, the $i$-th region in the current $t$ frame, is projected to the previous learned basis images of the $j$-th object characterized by $O_j^{Lib}$ from the 2DPCA. Assuming the error image $\mathbf{E}_{ij}^t = \mathbf{A}_{t,i} - \mathbf{L}_j \mathbf{L}_j^T \mathbf{A}_{t,i} \mathbf{R}_j \mathbf{R}_j^T$ is i.i.d.Gaussian distributed, the following relation is obtained,

$$p(\tilde{O}_i|O_j^{Lib}) \approx p(\tilde{O}_i|O_{1:t-1,j}) = p(\mathbf{E}_{ij}^t) \quad (9)$$

For the first term in the right hand side of (8), each blob shape is described by a parameter vector $[x_t, y_t, \theta_t, s_t, \alpha_t, \phi_t]^T$ , i.e., the horizontal and vertical coordinates of object center, rotation, scaling, aspect ratio of the box, and skew direction. Assuming the difference of each individual parameter in consecutive frames is Gaussian distributed, the conditional probability in the first term of (8) can be computed in a similar way as that for the second term. Finally, locating, or, associating the best blob index $i$ with the previous sequence of blob $j$ can be formed as the following ML estimate,

$$i^* = \arg\max_i p(L_{t,i}|L_{t-1,j}) p(\tilde{O}_i|O_{1:t-1,j}) \quad (10)$$

where the assumption of 1st-order Markov processes is used.

## 5. EXPERIMENTAL RESULTS

The proposed method has been tested for several image sequences, containing indoor and outdoor scenarios with dynamic backgrounds and object changes, e.g., in pose, expression and motion. Variations in these sequences require object appearance be effectively modeled. However, for simplicity, only one global moving object (e.g. car, face) with variety of moving speed was tracked in each image sequence. The images size is $320 \times 240$, frame rate is 30 fps, and all images are in gray scales. Initially, the position of each object was manually marked. Each object region was scaled to $32 \times 32$. 2DPCA was first computed using 5 frames from the selected region. It was then updated incrementally in every 5 frames. In the incremental 2DPCA, the numbers of row and column eigenvectors were set as $s = 4$ and $t = 4$. This is equivalent to 16 basis images (from outer products of all possible pairs).

Fig.1 shows the resulting tracked blobs of object from 2 videos using the proposed method. From Fig.1 one can observe that most object regions were correctly and accurately tracked, except in a few cases we observed that the speed of tracker was lagged behind when the object motion was fast. This is probably due to the relatively large interval in incremental learning (5 frames in our tests), which is a compromise between the computation and tracking speed. To obtain more insight, Fig.2 shows two original and reconstructed object regions for the tracked blobs, using the proposed incremental 2DPCA.

**Fig. 1**. Results obtained from the proposed method. Each row contains several frames of video shown by large images, where the tracked blob is marked by the green box. Row 1: frames from the video 'davidin300', (frames #120,339,449); Row 2: from 'car4', (#10,195,510).

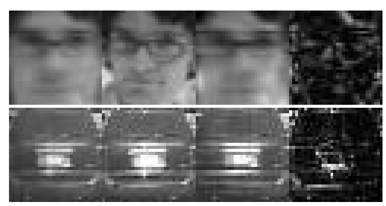

**Fig. 2**. Characterization of the tracked region using 2DPCA. From left to right: the mean, the original and the reconstructed region using 2DPCA, and the residuals between the original and reconstructed regions. Rows 1: for tracked face region (from #120); Rows 2: for tracked car region (from #10).

*Evaluation:* Since ML blob-tracking and 2DPCA object characterization are inter-affected, the errors in tracked-blobs will reflect the overall errors caused by these two factors. For evaluation, we applied 2 methods. The first evaluation method is based on the distance between the center of tracked-blob and a pre-selected salient point on the object. We compare this distance with the "ground truth" which was manually marked through image frames. The 2nd evaluation method is based on the Square Reconstruction Error (SRE) in each frame, defined as $\|\mathbf{A}_i - \mathbf{L}\mathbf{L}^T \mathbf{A}_i \mathbf{R}\mathbf{R}^T\|_F^2$. It is worth noting that this error value can be affected by many factors, rather than being a pure indicator on whether the target object is correctly tracked. For example, sudden changes in the object, e.g. lighting and pose, can lead to large SREs. Fig.3 (left) shows the distance to the salient object point versus the frame number, and Fig.4

(1st row) shows the coordinates of salient object point versus the frame number. Comparing the ground truth and the results from the 2DPCA (i.e., the green curve vs. the red curve) in Fig.3 (left) and Fig.4 (1st row), showed that the proposed method has a reasonably good tracking performance. For those frames with relatively large errors (in Fig.3 (left)), our visual inspection indicated that the errors were mainly due to the small size of the bounding box as comparing with those in the ground truth, however, the faces were fully tracked. It is observed from Fig.4 (2nd row) that SREs appeared large between the frames #185 and #230. Visual inspection of the resulting sequence showed that the target car was correctly tracked by a tight bounding box in these frames, however, the lighting had suddenly changed as the car entered a dark tunnel. Those frames having relatively large SREs in Fig.3 (right) were due to the same reasons, i.e., the deviations in the size or the angle of the bounding box.

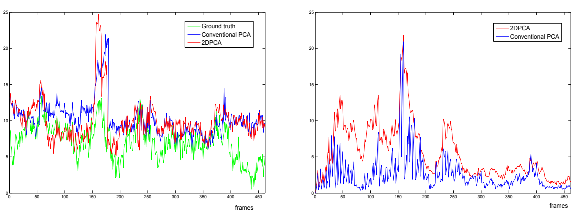


**Fig. 3**. Evaluation results for the face sequence 'davidin300'. Left: distances between the bounding box center and the nose tip as a function of image frames; Right: corresponding SREs as a function of frames. Green: the ground truth; Red: from the proposed method; Blue: from the conventional PCA.

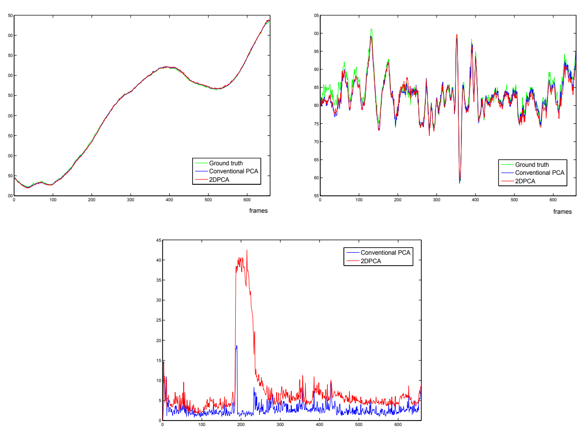


**Fig. 4**. Evaluation results for the car sequence 'car4'. Row 1: $x$ and $y$ coordinates from the centers of the target car vs. frames. Row 2: SREs vs. frames. Green: the ground truth; Red: from the proposed method; Blue: from the conventional PCA.

*Comparisons:* The two evaluation methods used above are then applied to the image sequences resulted from using the conventional PCA obtained from (1). The comparisons were made with [4] where the conventional PCA was set, and also the tracking probabilities coincide when a single object is handled. Comparisons can therefore indicate the impact caused by using the conventional PCA versus the incremental 2DPCA. For the conventional PCA, image in 32×32 region was *column–scanned* to a 1D vector of $1024 \times 1$. The number of principal eigenvectors was set as 16, equivalent to 16 basis images for each object. The SRE used in the 2nd evaluation method for the conventional PCA is defined as $||\mathbf{a}_i - \mathbf{X}\mathbf{X}^T\mathbf{a}_i||^2$, where columns of $\mathbf{X}$ are the principal eigenvectors of $\mathbf{C}$ in (1).

The results from these two evaluation methods associated with the conventional PCA are then included in Fig.3 and Fig.4 (in blue curves) for comparison. Comparing the distance curves (or, the salient point position curves) between the two methods, the results from the proposed incremental 2DPCA are shown somewhat better than the conventional PCA (which is more obvious in the face sequence). Comparing with the SRE curves, the results from the conventional PCA method are shown to be slightly better. It is worth noting that large SREs may only partially indicate whether the bounding box has correctly tracked the target.

## 6. CONCLUSIONS

The proposed method of using incremental 2DPCA object characterization combining with ML blob object region tracking has been tested on a variety of videos containing moving objects in dynamic backgrounds. The experimental results have shown that the proposed method has successfully tracked moving objects in most cases, however, for fast object moving, adjustment to a smaller updating interval in incremental learning is required. Comparisons with the conventional PCA which requires more computations for learning the basis images, the proposed method has shown similar, or slightly reduced tracking performance.

## 7. REFERENCES


[1] M.J.Black, A.D.Jepson, "Eigentracking: Robust matching and tracking of articulated objects using a view-based representation", Int'l Journal of Computer Vision, 26(1), 63-84, 1998.

[2] Z.Khan, T.Balch, F.Dellaert, "A rao-blackwellized particle filter for eigentracking", CVPR'04, 2004.

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

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

[5] J.Yang, D.Zhang, A.F.Frangi, J-Y.Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition", IEEE Trans. PAMI, 26(1), 131-137, 2004.

[6] D.Zhang, Z-H.Zhou, "(2d)$^2$ PCA: Two-directional two-dimensional pca for efficient face representation and recognition", Neurocomputing, 69(1-3), 224-231, 2005.

[7] J.Ye, "Generalized low rank approximations of matrices", Machine Learning, 61(1-3), 167-191, 2005.

[8] C. Ding, J. Ye, "Two-dimensional singular value decomposition (2dsvd) for 2d maps and images", in Proc. SIAM Int'l Conf. on Data Mining (SDM'05), 32-43, 2005.

[9] M.Turk, A.Pentland, "Eigenfaces for recognition", Journal of Cognitive Neuroscience, 3(1), 71-86, 1991.