2018 24th International Conference on Pattern Recognition (ICPR)
Beijing, China, August 20-24, 2018

Hybrid Sparse Subspace Clustering for Visual Tracking

Lin Ma*† and Zhihua Liu*
*Samsung Research Institute China - Beijing (SRC-B), China
Email: malin_u@126.com
†Department of Automation, Tsinghua University, Beijing, China

***Abstract*—In many conditions, the object samples are distributed in a number of different subspaces. By segmenting the subspaces with spectral clustering based subspace clustering, more accurate sample distribution is obtained. The LSR (Least Squares Regression) sparse subspace clustering method which fulfills the EBD (Enhance Block Diagonal) criterion and has closed-form solution, is an important spectral clustering based sparse subspace clustering method. However, LSR uses no discriminative information which is important to discriminate positive samples from the negative samples. Thus, we propose a new hybrid sparse subspace clustering method which makes the clustering discriminative by involving the discriminative information provided by graph embedding into LSR. The sub subspaces obtained based on the new subspace clustering method can both retain the object distribution information and also make the object samples less confused with surrounding environment. Experimental results on a set of challenging videos in visual tracking demonstrate the effectiveness of our method in discriminating the object from the background.**

I. Introduction

Visual object tracking can provide the basis for many vision applications, such as visual surveillance and motion analysis. To obtain the accurate object states at each time, many trackers are proposed based on various appearance models [1][2][3][4][5][6][7][8]. Subspace model is a widely used appearance model as subspace can decrease the sample dimension and also represent the sample distribution, e.g. the PCA (Principal Component Analysis) subspace [9][10]. Ross *et al.* [11] propose the IVT (Incremental Visual Tracking) method based on incremental PCA. IVT updates the subspace model adaptively and obtains robust tracking results. But IVT assumes that the samples are distributed in a single linear subspace which cannot represent the real sample distribution accurately sometimes. Thus, nonlinear subspace models have been proposed [12][13][14] to obtain more accurate sample distribution, such as kernel based methods [15] and Riemannian manifold based methods [16][17]. Ma *et al.* [18] project the samples into local linear subspace and represent the local nonlinear sample distribution with hypersphere in the projected linear subspace. With hypersphere, more accurate sample distribution is obtained. However, the nonlinear subspace models cannot tackle the conditions when samples from several different subspaces are mixed together.

To separate the mixed samples from different subspaces, subspace clustering method has been proposed. Spectral clustering based subspace clustering method is widely used for

Fig. 1. **Sample frames of the tracking results.** In the videos, the objects experience occlusion and background clutter challenges, which makes the background influence the determination of object state. Our method introduces the discriminative information into the generative subspace clustering process. And from the figure, we see that our method obtains better performance than SC1 which performs only the generative subspace clustering process.

its high ability to segment the subspaces [19][20][21]. Lu *et al.* [22] propose a least squares regression based subspace clustering (LSR) method and propose the EBD criterion. They have also demonstrated that LSR fulfills the EBD criterion. By using the Frobenious norm, LSR can have closed form solution. The previous subspace clustering methods only consider the positive samples and do not represent the discriminative information between positive samples and negative samples. However, the discriminative information is important in computer vision and machine learning fields. Thus, we propose a new hybrid subspace clustering method which involves the discriminative information into the generative subspace clustering.

Our method is based on the LSR method and uses an iterative way to introduce the discriminative information into the subspace clustering process. The discriminative information is computed based on graph embedding, and represented as a discriminative coefficient matrix. The generative information is represented as a generative coefficient matrix which is computed based on self representation. By combining the discriminative coefficient matrix and the generative coefficient matrix and performing spectral clustering on the combined coefficient matrix, the obtained multiple PCA subspaces can both represent the historical sample distribution and discriminate the object samples from the background samples (Figure 1).

The main contributions of this paper are summarized as follows.

1) We introduce discriminative information into the LSR

sparse subspace clustering method. With the discriminative information, the subspaces are able to represent the sample distributions and meanwhile discriminate the object samples from the background samples. The introduction of the discriminative information into LSR is achieved through a proposed iterative process. With this iterative process, other spectral clustering based subspace clustering methods also can be incorporated into the framework conveniently by computing the generative coefficient matrix accordingly.

2) We utilize this proposed subspace clustering method to perform visual tracking and obtain robust tracking performance.

II. Hybrid Sparse Subspace Clustering

Subspace clustering is important in representing the sample distributions. In many conditions, the object samples are distributed in a set of different subspaces. Researchers find that the convex Lambertian objects are distributed in nine linear subspaces approximately [23][24]. By segmenting the original subspace, the sample distribution is represented more accurately. Moreover, in visual tracking the object samples are distributed nonlinearly in many conditions [17][18]. And by segmenting the subspace into several linear subspaces, the following process is made more convenient. Besides, using linear subspaces to represent local distribution can avoid over-fitting problem.

To perform spectral clustering based subspace clustering, we need to obtain the reconstruction coefficient matrix. Let $x_i^+, i = 1, ..., N$ be N object samples or positive samples, and $A = [x_1^+, x_2^+, ..., x_N^+]$. Let W be the reconstruction coefficient matrix of the samples, $g(A)$ be a dictionary formed using A, $\Omega(A, W)$ be the regularization norm. In this paper, we define $g(A) = A$. Then the optimal reconstruction coefficient matrix is

$$W^* = arg\min_{W} ||A - g(A)W||_l + \lambda\Omega(A, W), \quad s.t. W \in C \tag{1}$$

where l represents the norm, λ is a regularizer constant, and C is the constraint of W [21]. By performing spectral clustering on $|W^*| + |W^{*T}|$, the sample subspace is segmented.

Many subspace clustering methods are proposed by defining various kinds of l, $\Omega(A, W)$ and C [19][21][22]. Among these methods, LSR is an important method which is convenient to be resolved [22]. In this paper, our method is based on LSR method. However, LSR is only used to segment the object samples, and uses no discriminative information which is important in object tracking. In contrast, we propose a new method which involves the discriminative information in the subspace clustering.

The proposed method is used to perform object tracking. We retain N representative object subimages as base subimages $B_i, i = 1, ..., N$. To tackle occlusion, the object subimage is divided into four half subimages, i.e. left, right, top and bottom. The new subspace clustering method is performed for each half subimage independently. Taking the left half subimage as example, in the tracking system the N left half subimages of N base subimages are taken as the positive

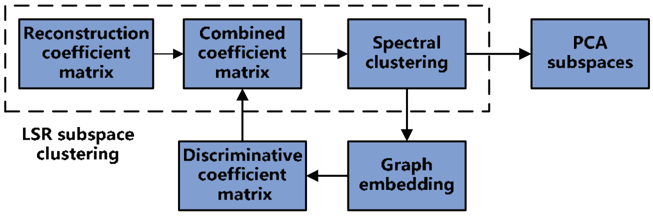

Fig. 2. **Process of the hybrid sparse subspace clustering method.** The new method is based on the LSR method, but differently we introduce discriminative information into the subspace clustering process. When the samples are classified into several groups, the graph embedding is performed to provide the discriminative information which is represented with the discriminative coefficient matrix. The discriminative coefficient matrix is combined with the generative coefficient matrix and the combined coefficient matrix is used for following spectral clustering. The process iterates until reaching the termination condition. Then we compute the PCA subspace for each group of samples to represent the object appearance.

samples. And we sample N' samples near the positive samples according to uniform distribution as the negative samples. The distance between the bounding boxes of the negative samples and the positive samples is within a given range for x, y position, height and width of the bounding box respectively. The proposed subspace clustering method is performed with these samples.

A. Segmenting the Subspace Involving Discriminative Information

Existing subspace clustering methods only consider the generative information of the object samples. However, in object tracking the discriminative information is also very important as the discriminative information can discriminate the object from the background. Thus, we involve the discriminative information in the subspace clustering. We base our method on LSR. Let $\widetilde{W}$ be the discriminative coefficient matrix indicating the sample relations provided by graph embedding. Then the objective function of W is revised as

$$f(W) = ||A - AW||_F^2 + \lambda||W||_F^2 + \lambda_1 g(W, \widetilde{W}), \tag{2}$$

where $g(W, \widetilde{W})$ defines the influence of the discriminative information in computing W, and λ_1 is a parameter determining the importance of the discriminative information in the subspace clustering. By minimizing $f(W)$, we obtain the optimal coefficient matrix representing the extent of two samples being in the same group.

To involve the discriminative information, we need to make W approximate $\widetilde{W}$ when defining $g(W, \widetilde{W})$. One way is defining $g(W, \widetilde{W}) = |||W| + |W^T| - \beta\widetilde{W}||_F^2$, where β is a parameter making the entry values of $|W| + |W^T|$ and $\widetilde{W}$ on the same order of magnitude. However, this kind of strategy cannot obtain closed-form solution as absolution operator exists. Besides, this method is not convenient to be used in other spectral based subspace clustering model as new special optimization algorithms need to be proposed for different subspace clustering methods. To avoid the two problems, we obtain the optimal combined coefficient matrix H in another way which contains two steps. First, according to LSR, we minimize

$$f(W) = ||A - AW||_F^2 + \lambda||W||_F^2 \tag{3}$$

through $df(W)/dW = 0$ and obtain the reconstruction coefficient matrix

$$W^* = \left[A^T A + \lambda I\right]^{-1} A^T A. \quad (4)$$

The generative information represents the sample distributions, and is useful to delineate outliers. In another part, the discriminative projection direction indicates the discriminative information with which the object samples are easier to be separated from the background. Thus, we balance between the generative information and the discriminative information, and then obtain the coefficient matrix H for spectral clustering as

$$H = \alpha(|W^*| + |W^{*T}|) + (1-\alpha)\widetilde{W}, \quad (5)$$

where α is a constant within (0, 1). The coefficient matrix is able to represent the relations between samples, and then can be used to segment the samples into several groups. By performing spectral clustering on H, the samples are classified into several groups. When new segmented subspaces are obtained, the discriminative information is changed. Then we perform spectral clustering again with the new obtained discriminative coefficient matrix. The process iterates until reaching the maximum iteration times N_T (Figure 2).

With this proposed method, we obtain closed-form resolution of generative coefficient. And in another part, various spectral based subspace clustering methods can be incorporated into the framework conveniently by computing the generative coefficient matrix accordingly. When computing the subspaces for each group of positive samples, we add samples to a group if this group has too few samples. Let $\widetilde{p}_k$ and $\bar{I}_k$ be the discriminative projection vector and mean of samples of group k respectively. Let the minimum sample number of each group be N_0. If the sample number of group k is smaller than N_0, for sample x_i^+ not belonging to group k, we define the affinity between x_i^+ and $\bar{I}_k$ as $d_i = \exp(-|\widetilde{p}_k^T(x_i^+ - \bar{I}_k)|)$, then we add samples with the largest affinity values into group k until the sample number of the group reaches N_0.

B. Computing Discriminative Coefficient Matrix

Graph embedding is a framework of dimensionality reduction. PCA, LDA, etc. can all be combined in this framework [25]. We perform graph embedding for each group using the object samples and all the background samples to obtain discriminative information. Let the negative samples be $x_i^-, i = 1, ..., N'$, and let the positive samples of a group be $\widetilde{x}_i^+, i = 1, ..., N_1$, while N_1 is the positive sample number of group k and N' is the number of all the negative samples. Let u be the mean of these samples, positive and negative, and we define $X = [\widetilde{x}_1^+ - u, ..., \widetilde{x}_{N_1}^+ - u, x_1^- - u, ..., x_{N'}^- - u]$. In graph embedding, we define the weights in the Lapalician matrix according to Euclidean distance between the samples. Let the Euclidean distance between sample $\widetilde{x}_i^+$ and x_j^- be $d_{i,j}$, then the weight about the two samples is defined as $\exp(-d_{i,j})$. The weights about the samples which are both positive samples or both negative samples are defined as 0. According to the weight, we define the Laplacian matrix $\mathcal{L}$, and the discriminative projection vector $\widetilde{p}_k$ is obtained by

$$\widetilde{p}_k = \arg\min_p p^T X \mathcal{L} X^T p, \quad (6)$$

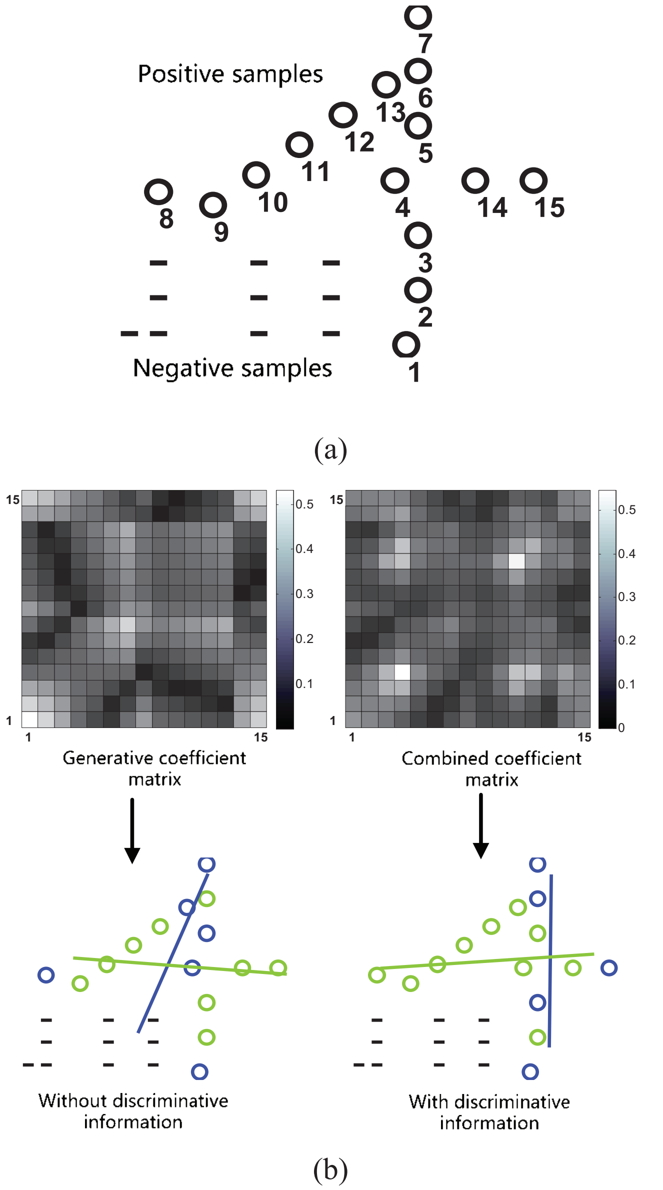

Fig. 3. **Effectiveness of the hybrid subspace clustering method.** For showing clarity, here the proposed method is performed on 2D samples. (a) The positive samples and the negative samples. (b) The clustering results of the positive samples represented with two kinds of colors. We also show the obtained coefficient matrix used for following spectral clustering. From the figure, we see that by involving discriminative information the subspaces are less confused with background. Here, the generative coefficient matrix on the left of (b) is represented as $|W^*| + |W^{*T}|$. In this figure, we define $\alpha = 0.5$, $N' = 10$, $N_T = 5$ and no samples are added into a group with too few samples. The two lines represent the corresponding subspaces of the samples of the same color.

$$s.t. p^T p = 1$$

The details of the graph embedding can be found in [26].

By making the subspace orthogonal to the projection direction, the object samples are less likely to be confused with the background samples (Figure 3). Here, the subspace of each sample group is computed corresponding to the mean sample of the group. To make the new subspace orthogonal to the projection direction, we make the samples having similar projection values to the mean sample classified into one group.

Algorithm 1: Hybrid sparse subspace clustering.

Input: $x_i^+, i = 1, ..., N$; $x_i^-, i = 1, ..., N'$; α.
Output: $U_k, k = 1, ..., K$.
1 Compute W^* with (4).
2 Obtain the combined coefficient matrix H with (5).
3 Perform spectral clustering on H.
4 **if** *reaching the iteration times* N_T **then**
5   goto line 9.
6 **end**
7 Perform graph embedding for each sample group.
8 Compute $\widetilde{W}$, back to line 2.
9 Add samples to a group if the number of the samples in the group is smaller than N_0.
10 Compute the PCA subspaces.

Algorithm 2: The flowchart of the tracking system.

//Tracking
Obtain coarse state with corner tracking.
Sample $X_i^S, i = 1, ..., N$.
Evaluate each candidate state with (11).
Select the optimal state.
//Updating
if *current frame fulfills storing constraint* **then**
    Store the frame image.
    Select the optimal base sub image for corner tracking at the next frame.
end
if *the saved sample number reaches a defined number* **then**
    Update base subimages.
    Update hybrid sparse subspace clustering model as Algorithm 1.
end

This requirement is achieved by defining the discriminative coefficient matrix $\widetilde{W}$ according to the projection value. Given one sample x_i^+, the similarity between x_i^+ and $\bar{I}_k$ is defined as

$$l_i^k = \exp\{-|\widetilde{p}_k^T(x_i^+ - \bar{I}_k)|\}. \tag{7}$$

Then sample similarity in terms of the discriminative projection between x_i^+ and x_j^+ is defined as

$$\widetilde{w}_{ij} \propto \max_k l_i^k l_j^k. \tag{8}$$

With the sample similarity, we define the discriminative matrix $\widetilde{W}$. With $\widetilde{W}$, we perform spectral clustering again. If two samples both have similar projection values to the same mean samples, they have large corresponding values in $\widetilde{W}$. Then they tend to be classified into the same group as the mean sample. As the mean samples are changed after performing spectral clustering, the $\widetilde{W}$ is also changed accordingly. Thus, the subspace clustering process iterates. Let the subspaces of the sample groups be $U_k, k = 1, ..., K$ where K is the number of sample groups. Then the proposed subspace clustering learning method is summarized as Algorithm 1. The $\widetilde{W}$ is defined as all 0 matrix initially.

III. Proposed Tracking System

Our aim is to obtain the optimal object state at each frame with the proposed subspace clustering method. The object state X_t at time t is represented as the x, y position and the width and height scales of the object. During tracking, at a frame we first perform corner tracking (only edge points are considered as candidate matching corners here) and obtain a coarse state [18], and then sample N_S candidate states around the coarse state according to Gaussian distribution. The candidate state is evaluated based on the proposed subspace clustering method. The state with the largest evaluation is selected as the optimal object state. The object sample is saved if it is not polluted. When M object samples are saved, we update the base subimages and the multiple subspaces obtained with the proposed subspace clustering method. The sample saving process and the base subimage updating process are as the method in [18]. The whole process of the tracking system is summarized as Algorithm 2.

State evaluation. To deal with occlusion problem, we divide each object sub image into four half subimages, and perform state evaluation for each half subimage separately. The subspace clustering is also performed for each half subimage separately. We take the left half subimage as an example, and it is the same for other half subimages. Given $\bar{I}_k$ and U_k which are the mean and subspace of the features about the left half subimages of group $k, k = 1, ..., K$, let N_D be the dimension of the subspace, then the evaluation of left half subimage I_t at frame t is defined as

$$L = \max_k L^k, \tag{9}$$

where

$$L^k = \exp\{-||(I_t - \bar{I}_k) - U_k U_k^T(I_t - \bar{I}_k)||_2^2\}. \tag{10}$$

Let O_t be the observation at frame t, and $L_r, r = 1, ..., 4$ be the evaluations about the four half subimages. By combining these evaluations, we obtain the evaluation of state X_t as

$$p(O_t|X_t) = \sum_r L_r. \tag{11}$$

IV. Experiments

In this section, we present the experimental results of our method. This section contains three parts. Firstly, we present the implementation details of our method. Secondly, we demonstrate the validity of the proposed subspace clustering method by comparing with the method not involving discriminative information. Finally, we compare our method with five state-of-the-art methods. Both quantitative and qualitative analysis are presented to show the effectiveness of our method when comparing with the five methods.

A. Implementation Details

We test our method on the OTB benchmark [27] containing fifty one videos which involving various challenges. The experiments are conducted on a PC with a 2.5 GHz Intel CPU with 12 GB RAM. The object subimage is warped to 32×32 gray subimage. We set K=2, $N_S = 150$, N=15, $N_0 = 5$, $N' = 20$, $N_D = 4$, $\lambda = 1$, $\alpha = 0.2$, $N_T = 3$ and update the system every 5 saved samples. The state at the first frame is manually set. The run time of our method is around 0.03 sec/frame. We compare our method with five state-of-the-art methods: MEEM [28], Staple [29], MUSTer [30], LSCT [18] and KCF [31]. For the comparing methods, we utilize the source codes provided by the authors. When performing corner tracking, a large number of corners exist on the object boundary in many times. To make the corners on the boundary able to be detected within the object area, the initial object sizes are relatively larger than in [27]. We utilize the pixel center error and overlap evaluation to evaluate the tracking performance.

When extracting the feature for a given bounding box, we divide the bounding box into 8×8 grids and for each grid we extract a vector composed of three values. By concatenating the vectors of the grids, we obtain the object feature. We define the weight of a grid as $a_{i,j} \propto \exp\{-0.1 \times ((i-4.5)^2 + (j-4.5)^2)\}, i,j = 1,...,8$. For a grid, let $b_{i,j}$, $dx_{i,j}$, $dy_{i,j}$ be the intensity, derivatives with respect to x and y of pixel $i,j = 1,...,3$ of the grid and a' be the weight of the grid, then the three values are defined as $\sum_{i,j} a' b_{i,j}$, $\sum_{i,j} a' dx_{i,j}$ and $\sum_{i,j} a' dy_{i,j}$.

B. Effectiveness of Involving Discriminative Information in Sparse Subspace Clustering

We compare our method with the subspace clustering method not using discriminative information (SC1) on fifty one video sequences to test the effectiveness of involving the discriminative information in the sparse subspace clustering process. When not using discriminative information, the subspaces of the object samples are confused with the background. Then the background sample also has small distance to the subspace which causes the tracking not robust any more. In contrast, by utilizing the discriminative information in the subspace clustering, our method obtains better performance. The results of SC1 and ours are shown in Table 1. From Table 1, we see that on both the two criterion, our method achieves better performance. Figure 1 shows some sample frames of SC1 and our method on *David3* and *Shaking*. The object experiences occlusion and background clutter in the two videos, and the background disturbs the sample evaluation in

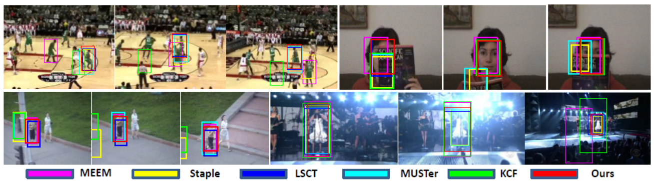

Fig. 4. **Sample frames** of our method and five other methods, MEEM [28], Staple [29], MUSTer [30], LSCT [18] and KCF [31] on four videos.

SC1 and reduces the tracking performance, while our method still obtains good tracking results.

C. Comparison with State-of-the-arts

Quantitative analysis. We compare our method with five state-of-the-art methods on the fifty one benchmark video sequences. The evaluations according to pixel center error and overlap of the six methods are shown in Table 2. From the table, we see that our method ranks the first when center error < 40 and overlap > 0 and > 0.25. And we rank the first according to mean percentage of frames about overlap, and rank the second according to mean percentage of frames about pixel center error. LSCT [18] represents the local distributions of the samples with a set of hyperspheres which represent the nonlinear sample distribution effectively. However, they use no discriminative information, which makes the object sample confused with the background sometimes. KCF [31] uses circulate matrix to represent the appearance at each location and use Gaussian distribution to initialize the evaluations. In this way, foreground and background area can be separated. However, it is not able to represent the object distributions and the relations between different frames effectively. In contrast, by involving discriminative information and representing appearance distribution, our method obtains better results. Staple [29], which combines correlation filter and color information, obtains good performances on both the two criterions. However, when the object is occluded severely by camouflage objects, Staple is influenced largely and fails to determine the accurate object states, such as Basketball, FaceOcc1 and Jogging.1, in Figure 4. But our method is still able to obtain the object states robustly by corner tracking and utilizing half subimages. And also by involving discriminative information, our method can discriminate the object from the camouflage occluders. Figure 5 shows the pixel center error map of the six methods on two video sequences.

Qualitative analysis. Our method obtains promising results when large challenges, such as occlusion and illumination

TABLE I
PERCENTAGE OF FRAMES ACCORDING TO PIXEL CENTER ERROR AND OVERLAP OF SC1 AND OURS.

		SC1	Ours
center	<20	0.683	**0.745**
	<30	0.773	**0.844**
	<40	0.819	**0.899**
overlap	>0	0.877	**0.956**
	>0.25	0.818	**0.898**
	>0.5	0.617	**0.657**

TABLE II
PERCENTAGE OF FRAMES ACCORDING TO PIXEL CENTER ERROR AND OVERLAP VALUES OF OURS AND THE COMPARING METHODS.

		MEEM	Staple	MUSTer	LSCT	KCF	Ours
center	<20	0.725	**0.813**	0.765	0.688	0.717	0.745
	<30	0.814	**0.853**	0.842	0.800	0.811	0.844
	<40	0.873	0.874	0.866	0.855	0.843	**0.899**
	Mean	0.804	**0.847**	0.824	0.781	0.791	0.829
overlap	>0	0.928	0.921	0.922	0.937	0.901	**0.956**
	>0.25	0.858	0.895	0.873	0.877	0.831	**0.898**
	>0.5	0.619	**0.683**	0.679	0.588	0.611	0.657
	Mean	0.801	0.833	0.825	0.801	0.781	**0.837**

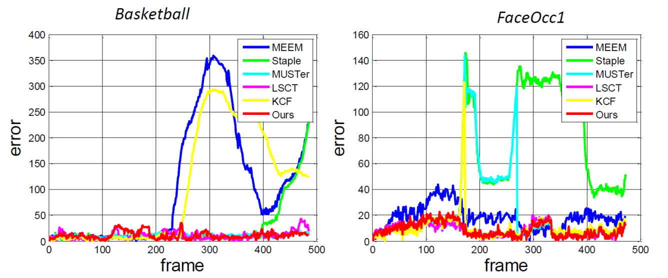

Fig. 5. **Pixel center error** of our method and five comparing methods on two videos at each frame.

variation, exist. Figure 4 shows some sample frames of our method tackling various challenges. We divide the object appearance into four half subimages. By evaluating each half subimage independently, our method is able to tackle the occlusion problem as the non-occluded area is not influenced by the occluded area. For example *FaceOcc1* experiences severe occlusion, the non-occluded half subimages still have high evaluation values. And by combining the evaluations of the four half subimages together, our method still achieves robust performance. When the object has similar appearance to the background, e.g. *Basketball*, as the segmented subspaces contains discriminative information. The object sample is not easy to be confused with the background. Thus our method can obtain accurate object state in this condition. When the object experiences drastic illumination variation, e.g. *Basketball*, the object appearances are distributed on a linear subspace approximately. As the PCA subspaces of each group of the object samples represent the linear distribution of samples, our method tackles the illumination variation effectively.

Our method is also able to tackle the pose variation problem. When the object pose has large variation, the object samples are distributed nonlinearly, e.g. *Jogging.1*. By segmenting the subspace into several linear subspaces, the nonlinear subspace is approximated and the following process is made more convenient. And then our method is able to deal with the pose variation. However, when both the object appearance and the background changes drastically, e.g. *Ironman*, our method is not able to discriminate the object from the background accurately.

V. Conclusion and Future Work

In this paper, we have proposed a novel subspace clustering method which involves discriminative information. The discriminative information is provided by graph embedding. By combining the generative information and the discriminative information, the multiple subspaces of the classified object samples can both represent the sample historical distribution and discriminate the object samples from the background samples. In the future, we will continue our researches in combining the generative information and discriminative information effectively in subspace clustering.

References

[1] C. Ma, J. B. Huang, X. Yang, and M.-H. Yang, "Hierarchical convolutional features for visual tracking," *ICCV*, 2015. 1

[2] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," *ICCV*, 2015. 1

[3] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," *CVPR*, 2016. 1

[4] A. Lukezic, T. Vojir, L. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," *CVPR*, 2017. 1

[5] H. Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," *ICCV*, 2017. 1

[6] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, "Eco: Efficient convolution operators for tracking," *CVPR*, 2017. 1

[7] J. Gao, H. Ling, W. Hu, and J. Xing, "Transfer learning based visual tracking with gaussian processes regression," *ECCV*, 2014. 1

[8] S. Hare, A. Saffari, and P. H. S. Torr, "Struck: Structured output tracking with kernels," *ICCV*, 2011. 1

[9] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991. 1

[10] J. Yang, D. Zhang, A. F. Frangi, and J. Y. Yang, "Two-dimensional pca: a new approach to appearance-based face representation and recognition," *TPAMI*, vol. 26, no. 1, pp. 131–137, 2004. 1

[11] D. A. Ross, J. Lim, R. Lin, and M. Yang, "Incremental learning for robust visual tracking," *IJCV*, vol. 77, no. 1, pp. 125–141, 2008. 1

[12] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000. 1

[13] J. B. Tenenbaum, V. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000. 1

[14] Z. Y. Zhang and H. Y. Zha, "Principal manifolds and nonlinear dimensionality reduction via local tangent space alignment," *SIAM J. Sci. Comput.*, vol. 26, no. 1, pp. 313–338, 2004. 1

[15] C. Alzate and J. A. K. Suykens, "Multiway spectral clustering with out-of-sample extensions through weighted kernel pca," *TPAMI*, vol. 32, no. 2, pp. 335–347, 2010. 1

[16] X. Li, W. Hu, Z. Zhang, and X. Zhang, "Robust visual tracking based on an effective appearance model," *ECCV*, 2008. 1

[17] X. Li, W. Hu, Z. Zhang, X. Zhang, and G. Luo, "Visual tracking via incremental log-euclidean riemannian subspace learning," *CVPR*, 2008. 1, 2

[18] L. Ma, X. Zhang, W. Hu, J. Xing, J. Lu, and J. Zhou, "Local subspace collaborative tracking," *ICCV*, 2015. 1, 2, 4, 5

[19] E. Elhamifar and R. Vidal, "Sparse subspace clustering," *CVPR*, 2009. 1, 2

[20] J. S. Feng, Z. C. Lin, H. Xu, and S. Yan, "Robust subspace segmentation with block-diagonal prior," *CVPR*, 2014. 1

[21] H. Hu, Z. Lin, J. Feng, and J. Zhou, "Smooth representation clustering," *CVPR*, 2014. 1, 2

[22] C. Y. Lu, H. Min, Z. Q. Zhao, L. Zhu, D. S. Huang, and S. C. Yan, "Robust and efficient subspace segmentation via least squares regression," *ECCV*, 2012. 1, 2

[23] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," *ICCV*, 2001. 2

[24] R. Ra and P. Hanrahan, "On the relationship between radiance and irradiance: Determining the illumination from images of a convex lambertian object," *Journal of the Optical Society of America A*, vol. 18, no. 10, pp. 2448 –2459, 2001. 2

[25] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *TPAMI*, vol. 29, no. 1, pp. 40–51, 2007. 3

[26] L. Ma, J. Xing, X. Zhang, and W. Hu, "Adaptive cooperative tracking based on multi-graph embedding and markov random field," *ICASSP*, 2013. 3

[27] Y. Wu, J. Lim, and M. Yang, "Online object tracking: A benchmark," *CVPR*, 2013. 4, 5

[28] J. Zhang, S. Ma, and S. Sclaroff, "Meem: Robust tracking via multiple experts using entropy minimization," *ECCV*, 2014. 5

[29] L. Bertinetto, O. M. J. Valmadre, S. Golodetz, and P. Torr, "Staple: Complementary learners for real-time tracking," *CVPR*, 2016. 5

[30] Z. Hong, Z. Chen, C. Wang, X. Mei, D. Prokhorov, and D. Tao, "Multi-store tracker (muster): A cognitive psychology inspired approach to object tracking," *CVPR*, 2015. 5

[31] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *TPAMI*, vol. 24, no. 5, pp. 1–14, 2015. 5