Co-Learned Multi-View Spectral Clustering for Face Recognition Based on Image Sets

Huang, Likun; Lu, Jiwen; Tan, Yap Peng

2014

Huang, L., Lu, J., & Tan, Y.-P. (2014). Co-Learned Multi-View Spectral Clustering for Face Recognition Based on Image Sets. IEEE Signal Processing Letters, 21(7), 875-879.

https://hdl.handle.net/10356/81735

https://doi.org/10.1109/LSP.2014.2319817

Co-Learned Multi-View Spectral Clustering for Face Recognition Based on Image Sets

Likun Huang*, *Student Member, IEEE,* Jiwen Lu, *Member, IEEE,* and Yap-Peng Tan, *Senior Member, IEEE*

***Abstract*—Different from the existing approaches that usually utilize single view information of image sets to recognize persons, multi-view information of image sets is exploited in this paper, where a novel method called Co-Learned Multi-View Spectral Clustering (CMSC) is proposed to recognize faces based on image sets. In order to make sure that a data point under different views is assigned to the same cluster, we propose an objective function that optimizes the approximations of the cluster indicator vectors for each view and meanwhile maximizes the correlations among different views. Instead of using an iterative method, we relax the constraints such that the objective function can be solved immediately. Experiments are conducted to demonstrate the efficiency and accuracy of the proposed CMSC method.**

***Index Terms*—Set-based face recognition, spectral clustering, multi-view, co-learning.**

I. Introduction

THERE has been a growing research interest in face recognition that utilizes sets of images. An image set contains variation information that is unavailable in an isolated image, which helps improve classification performance under challenging conditions, e.g., large variations in pose, illumination, low resolution, and etc., where the conventional face recognition systems based on single-shot images often fail to perform well. Generally, an image set can be a collection of unordered still images or frames from a video clip of a person. With rapid development of social networks, hundreds of millions of people have shared their personal photos and videos through Facebook, YouTube, Flickr and etc., thus image sets are nowadays much easier to be obtained than before. For the set-based face recognition system, an input is an image set that needs to be classified to one of the classes in the gallery dataset, where each reference is also an image set.

Previous Work: For the set-based face recognition techniques, how to appropriately describe an image set is very important. From this perspective, the existing work can be roughly grouped into two categories: single-model based methods [1]–[11], [27] and multi-model based methods [12], [28], [29]. Most existing work belong to the single-model based category where a cluster, a linear subspace or a probability density is employed to describe each image set. However, these methods have limited capability to capture the structures of image sets that are acquired under real-life scenarios with arbitrary qualities and nonlinear distributions. Thus more flexible and appropriate models are required to describe image sets obtained under practical conditions. Recently, Wang *et al.* proposed a multi-model based method known as Manifold-to-Manifold Distance (MMD) [12]. In this work, a Maximal Linear Patch (MLP) method was proposed to divide each image set into several clusters, in which the data points were linearly distributed. On this basis, every image set can be considered as a nonlinear manifold and described by multiple linear subspaces. To measure the similarity between two different manifolds, a MMD was designed. Through a series of comparisons with existing single-model based methods, the multi-model based method demonstrated an inherent advantage in describing an image set with nonlinear structure, which leads to a better recognition performance.

Motivations: Inspired by [12], we investigate the problem of face recognition based on image sets within the multi-model based framework. As mentioned above, a key step is how to describe an image set by using an appropriate model that exploits the information of this set as much as possible. Apparently, there is a severe under-utilization of image sets in the existing work, which only make use of single view information of each image set, e.g., the intensity information. In our daily lives, it is common that many real-world datasets naturally consist of multiple views, e.g., multiple different languages of the same article, web pages including contents and hyperlinks, videos including frames and audio information, and etc. Even for an image set that only consists of a collection of images, plenty information is still able to be exploited by using some image preprocessing steps, e.g., SIFT feature [21] and LBP feature [19] [20] can be extracted from an image set and provide complementary information to each other. In view of this, we propose a Co-learned Multi-view Spectral Clustering (CMSC) method to cluster each image set into several clusters by using the optimal embeddings learned from multiple views simultaneously. This leads to two contributions:

1) The existing set-based methods only utilize a single view of a dataset to recognize individuals. In contrast, our proposed CMSC method integrates multiple views of a dataset to enhance the performance.

2) Different from the existing multi-view clustering algorithms that divide each image set in an iterative manner [15], the proposed CMSC method solves the optimization function immediately after a relaxation of the constraints. This provides an efficient and robust division of each image set according

Likun Huang* (corresponding author) and Yap-Peng Tan are with the School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore, 639798. E-mail: huan0181@e.ntu.edu.sg, eyptan@ntu.edu.sg.

Jiwen Lu is with the Advanced Digital Sciences Center, Singapore, 138632. E-mail: jiwen.lu@adsc.com.sg.

EDICS: IMD-PATT: Pattern recognition.

to its multi-view information.

II. The Proposed CMSC Methods

As preliminaries, we first introduce the spectral clustering algorithms that exploit single view and multi-view information of image sets respectively.

Single View Spectral Clustering [13], [14]: Assume there is a dataset $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]$, from which an adjacency matrix $\mathbf{S}$ that stores the similarities between all pairs of points in $\mathbf{X}$ can be computed, i.e., $\mathbf{S}_{i,j} = \exp\left(-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{\sigma^2}\right)$. The Laplacian matrix $\mathbf{L} = \mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2}$, where $\mathbf{D}_{i,i} = \sum_{j=1}^{n} \mathbf{S}_{i,j}$ and $\mathbf{D}_{i,j} = 0$ for $i \neq j$. The single view spectral clustering algorithm aims to find the optimal embedding matrix $\mathbf{V}$, which consists of the approximations of the cluster indicator vectors:

$$\max_{\mathbf{V} \in \mathbf{R}^{n \times k}} tr\left(\mathbf{V}^T \mathbf{L} \mathbf{V}\right), \qquad s.t. \quad \mathbf{V}^T\mathbf{V} = \mathbf{I}. \tag{1}$$

The optimal $\mathbf{V}$ contains the first k eigenvectors of $\mathbf{L}$ as columns. Each row vector of $\mathbf{V}$ can be normalized and used to assign a data point to one of the k clusters by using the k-means algorithm.

Multi-View Spectral Clustering: A co-regularized multi-view spectral clustering (MSC) approach was recently proposed in [15] to extend the single view spectral clustering to the multi-view case. However, Co-regularized MSC adopts an iterative procedure to seek an embedding matrix $\mathbf{V}_i$ for each view i, $i \in 1, 2, \cdots, m$ of the dataset, which results in an unspecified initialization and a dependence on the number of iterations. Moreover, with a kernel matrix pre-defined for each $\mathbf{V}_i$, the geometric intuition of the multiplication among these kernel matrices in the objective function is not clear. Furthermore, analysis on the convergence of this iterative algorithm was absent.

Motivated by the above work, we propose a CMSC approach as shown in Fig. 1, which has the following key features:

1) To expose the geometric intuition explicitly, we use between-view correlation as a constraint such that the objective function enables the embeddings obtained from different views to get closer to each other.

2) Instead of an iterative learning procedure, our proposed CMSC method solves the optimization function immediately after a relaxation of constraints, such that all solutions $\mathbf{V}_1$, $\mathbf{V}_2$, $\cdots$, $\mathbf{V}_m$ can be obtained simultaneously. This is more efficient and accurate than updating only one matrix in each iteration.

In the following, two methods for multi-view spectral clustering are proposed, namely pairwise-based CMSC and centroid-based CMSC.

A. Pairwise-based CMSC

We assume that there are representations of an image set in m views, i.e., $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_m$, from which the Laplacian matrices $\mathbf{L}_1, \mathbf{L}_2, \cdots, \mathbf{L}_m$ can be calculated respectively. Since the representation in each single view of an image set is sufficient for clustering independently, we attempt to maximize $tr(\mathbf{V}_i^T\mathbf{L}_i\mathbf{V}_i), i \in \{1, 2, \cdots, m\}$ such that the embedding matrix $\mathbf{V}_i$, i.e., the approximations of the cluster indicator vectors, can be optimized for each view i. Meanwhile, we incorporate the between-view correlations into the objective function to make sure that the collaboratively learned embeddings $\mathbf{V}_i, i \in \{1, 2, \cdots, m\}$ get closer to each other. This enables the representations of the same data point in different views to be assigned to the same cluster. To achieve this goal, we define the between-view correlation as $tr(\mathbf{V}_i^T\mathbf{V}_j)$ for any two embeddings $\mathbf{V}_i$ and $\mathbf{V}_j$.

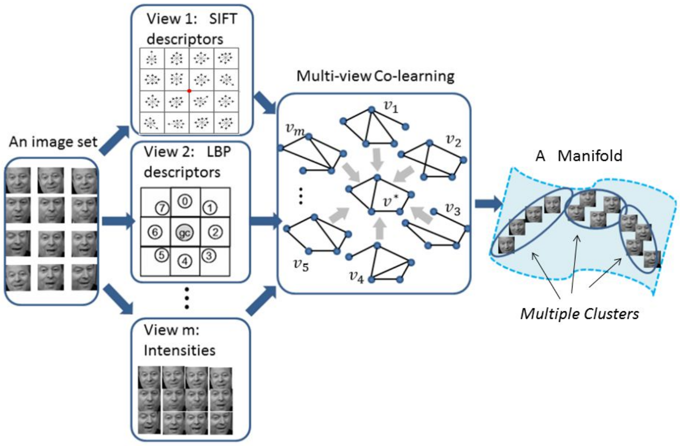

Fig. 1. The proposed CMSC method clusters an image set into several subsets (i.e., clusters) through a co-learning stage, which enforces the graphs from multiple views to be consistent with each other.

To integrate the above objectives, we propose an objective function that aims to seek a collection of embeddings $\mathbf{V}_1$, $\mathbf{V}_2$, $\cdots$, $\mathbf{V}_m$ to maximize both the individual spectral clustering terms and their correlations:

$$\begin{aligned} \max\nolimits_{\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_m \in \mathbf{R}^{n \times k}} \quad & \Bigg\{ \lambda_1 \sum_{i=1}^{m} tr(\mathbf{V}_i^T\mathbf{L}_i\mathbf{V}_i) \\ & + \lambda_2 \sum_{1 \leq i,j \leq m, i \neq j} tr(\mathbf{V}_i^T\mathbf{V}_j) \Bigg\}, \\ s.t. \quad & \mathbf{V}_i^T\mathbf{V}_i = \mathbf{I} \; \forall \; i \in \{1, 2, \cdots, m\}, \end{aligned} \tag{2}$$

where λ_1 and λ_2 are introduced to balance the weights of individual spectral clustering terms and their correlations, and $\sum_{1 \leq i,j \leq m, i \neq j} tr(\mathbf{V}_i^T\mathbf{V}_j)$ denotes the summation of all pairwise correlations among multiple views.

However, all the constrains in (2) are nonlinear, which makes it intractable to obtain closed-form solutions. To solve this problem, we relax these constraints to a single one as:

$$s.t. \qquad \sum_{i=1}^{m} \beta_i \mathbf{V}_i^T\mathbf{V}_i = \mathbf{I}, \tag{3}$$

where $\beta_1, \beta_2, \cdots, \beta_m$ are introduced to balance the powers of constraints in m different views.

For ease of exposition, we let $\mathbf{V}^T = \begin{bmatrix} \mathbf{V}_1^T & \mathbf{V}_2^T & \cdots & \mathbf{V}_m^T \end{bmatrix}$, then the objective function in (2) and (3) can be rewritten in a matrix form as:

$$\max_{\mathbf{V} \in \mathbf{R}^{(mn) \times k}} tr(\mathbf{V}^T\mathbf{L}\mathbf{V}), \qquad s.t. \quad \mathbf{V}^T\mathbf{B}\mathbf{V} = \mathbf{I}, \tag{4}$$

where

$$\mathbf{L} = \begin{bmatrix} \lambda_1\mathbf{L}_1 & \lambda_2\mathbf{I}^{n\times n} & \cdots & \lambda_2\mathbf{I}^{n\times n} \\ 0 & \lambda_1\mathbf{L}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \lambda_2\mathbf{I}^{n\times n} \\ 0 & \cdots & 0 & \lambda_1\mathbf{L}_m \end{bmatrix}, \quad (5)$$

$$\mathbf{B} = \begin{bmatrix} \beta_1\mathbf{I}^{n\times n} & 0 & \cdots & 0 \\ 0 & \beta_2\mathbf{I}^{n\times n} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \beta_m\mathbf{I}^{n\times n} \end{bmatrix}, \quad (6)$$

and $\mathbf{I}^{n\times n}$ denotes an identity matrix of size $n\times n$. The constraint $\mathbf{V}^T\mathbf{B}\mathbf{V} = \mathbf{I}$ imposes a scale normalization on the objective function such that trivial solutions can be avoided. Thus the optimization problem in (4) becomes a generalized eigen-decomposition of matrices $\mathbf{L}$ and $\mathbf{B}$, i.e., $\mathbf{L}\mathbf{V} = \lambda\mathbf{B}\mathbf{V}$.

B. Centroid-based CMSC

In the pairwise-based CMSC method, we assume that all the embedding matrices $\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_m$ tend to each other, which requires computation of the pairwise correlations among m embedding matrices in different views. From (2), it is noted that altogether $\frac{m(m-1)}{2}$ multiplications are conducted, which is very time-consuming. In view of this, a centroid-based scenario is considered where we assume that all embedding matrices $\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_m$ move towards a centroid matrix $\mathbf{V}^*$. Thus, there are only m comparisons between each embedding matrix $\mathbf{V}_i$ and the centroid matrix $\mathbf{V}^*$ that need to be calculated, which significantly reduces the computational complexity. Based on this idea, we propose a centroid-based CMSC method where a centroid embedding matrix $\mathbf{V}^*$ is involved for computing the between-view correlations. By defining the between-view correlation as $tr(\mathbf{V}_i^T\mathbf{V}^*), i \in \{1, 2, \cdots, m\}$, the proposed objective function aims to seek $m+1$ embedding matrices $\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_m, \mathbf{V}^*$ to maximize the individual spectral clustering terms and their correlations:

$$\max_{\mathbf{V}_1,\mathbf{V}_2,\cdots,\mathbf{V}_m,\mathbf{V}^*\in\mathbf{R}^{n\times k}} \left\{ \lambda_1 \sum_{i=1}^{m} tr(\mathbf{V}_i^T\mathbf{L}_i\mathbf{V}_i) + \lambda_2 \sum_{i=1}^{m} tr(\mathbf{V}_i^T\mathbf{V}^*), \right\}$$
$$s.t.\ \mathbf{V}_i^T\mathbf{V}_i = \mathbf{I}\ \forall\ i \in \{1, 2, \cdots, m\}, \quad \text{and } \mathbf{V}^{*T}\mathbf{V}^* = \mathbf{I}. \quad (7)$$

Again, λ_1 and λ_2 are introduced to balance the weights of the individual spectral clustering terms and their correlations, and $\sum_{i=1}^{m} tr(\mathbf{V}_i^T\mathbf{V}^*)$ denotes the summation of all the correlations between each $\mathbf{V}_i$ and $\mathbf{V}^*$.

Similar to the pairwise-based case, in order to obtain closed-form solutions, we relax the $m+1$ nonlinear constraints in (7) by combining them into a single one:

$$s.t. \quad \sum_{i=1}^{m} \beta_i\mathbf{V}_i^T\mathbf{V}_i + \beta^*\mathbf{V}^{*T}\mathbf{V}^* = \mathbf{I}, \quad (8)$$

where $\beta_1, \beta_2, \cdots, \beta_m$, β^*, pre-defined according to the prior knowledge, are introduced to balance the powers of constraints in m views and the centroid embedding. This constraint imposes a scale normalization on the objective function such that a trivial solution can be avoided.

Similarly, to facilitate the solving of the objective function, we let $\mathbf{V}^T = \begin{bmatrix}\mathbf{V}_1^T & \mathbf{V}_2^T & \cdots & \mathbf{V}_m^T & \mathbf{V}^{*T}\end{bmatrix}$, then the objective function in (7) restricted by the relaxed constraint in (8) can be rewritten in a matrix form as:

$$\max_{\mathbf{V}\in\mathbf{R}^{(m+1)n\times k}} tr(\mathbf{V}^T\mathbf{L}\mathbf{V}), \qquad s.t.\ \mathbf{V}^T\mathbf{B}\mathbf{V} = \mathbf{I}, \quad (9)$$

where

$$\mathbf{L} = \begin{bmatrix} \lambda_1\mathbf{L}_1 & 0 & \cdots & 0 & \lambda_2\mathbf{I}^{n\times n} \\ 0 & \lambda_1\mathbf{L}_2 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ \vdots & \ddots & \ddots & \lambda_1\mathbf{L}_m & \lambda_2\mathbf{I}^{n\times n} \\ 0 & \cdots & \cdots & 0 & \lambda_1\mathbf{L}^* \end{bmatrix}, \quad (10)$$

$$\mathbf{B} = \begin{bmatrix} \beta_1\mathbf{I}^{n\times n} & 0 & \cdots & \cdots & 0 \\ 0 & \beta_2\mathbf{I}^{n\times n} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \beta_m\mathbf{I}^{n\times n} & 0 \\ 0 & \cdots & \cdots & 0 & \beta^*\mathbf{I}^{n\times n} \end{bmatrix}, (11)$$

and $\mathbf{L}^*$ is the Laplacian matrix of the centroid graph $\mathbf{S}^*$. To reasonably initialize the centroid graph $\mathbf{S}^*$, we let $\mathbf{S}^* = \frac{1}{m}\sum_{i=1}^{m}\mathbf{S}_i$, then we have the Laplacian matrix $\mathbf{L}^* = \mathbf{D}^{*-1/2}\mathbf{S}^*\mathbf{D}^{*-1/2}$. Similarly, this optimization problem becomes a generalized eigen-decomposition where $\mathbf{L}\mathbf{V} = \lambda\mathbf{B}\mathbf{V}$.

C. The Whole Procedure

To recognize persons based on image sets, the whole procedure of the proposed CMSC includes two steps: image set modeling and manifolds matching.

Image Set Modeling: Assume there are N face image sets stored in the gallery dataset. We consider each image set as a manifold $\mathbf{M}$ and describe each manifold by using a collection of local models. To achieve this, each image set is first divided into several subsets through the proposed CMSC method, then each subset is modeled by its mean vector $\mathbf{m}_i$. In this way, a manifold $\mathbf{M}$ can be modeled by using a collection of mean vectors $\{\mathbf{m}_1, \mathbf{m}_2, \cdots, \mathbf{m}_r\}$.

Manifolds Matching: Based on [12], the nearest distance between local models from two manifolds $\mathbf{M}^X = \{\mathbf{m}_1^X, \mathbf{m}_2^X, \cdots, \mathbf{m}_{n_X}^X\}$ and $\mathbf{M}^Y = \{\mathbf{m}_1^Y, \mathbf{m}_2^Y, \cdots, \mathbf{m}_{n_Y}^Y\}$ is defined as the manifold-to-manifold distance:

$$d(\mathbf{M}^X, \mathbf{M}^Y) = \min\left\{\min_i d(\mathbf{m}_i^X, \mathbf{M}^Y), \min_j d(\mathbf{m}_j^Y, \mathbf{M}^X)\right\}, \quad (12)$$

where $i \in \{1, 2, \cdots, n_X\}$, $j \in \{1, 2, \cdots, n_Y\}$, and

$$\begin{aligned} d(\mathbf{m}_i^X, \mathbf{M}^Y) &= \min_{1\leq j\leq n_Y} d(\mathbf{m}_i^X, \mathbf{m}_j^Y) \\ &= \min_{1\leq j\leq n_Y} \|\mathbf{m}_i^X - \mathbf{m}_j^Y\|_2^2, \end{aligned} \quad (13)$$

where $d(\mathbf{m}_i^X, \mathbf{M}^Y)$ denotes the distance between local model $\mathbf{m}_i^X$ and manifold $\mathbf{M}^Y$, and $d(\mathbf{m}_i^X, \mathbf{m}_j^Y)$ denotes the distance

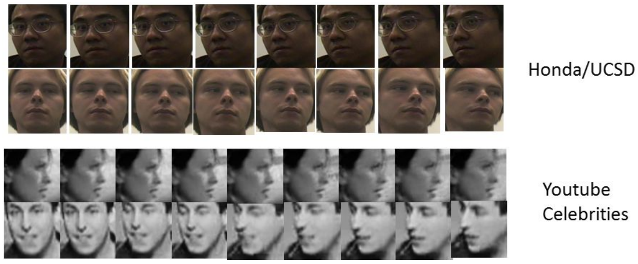

Fig. 2. Some examples of the cropped face images from the Honda/UCSD database and the Youtube Celebrities Database.

between local models $\mathbf{m}_i^X$ and $\mathbf{m}_j^Y$.

III. Experiments

A. Experimental Settings

We evaluate the performance of the proposed pairwise-based CMSC method (PW-CMSC) and centroid-based CMSC method (CT-CMSC) under the following configurations.

Image Set Modeling: For comparison purposes, we use 5 different methods to model each face image set.

1) MLP [12]: As a clustering method, MLP is able to divide each image set into multiple linearly distributed subsets;

2) k-means Clustering [16];

3) Single View Spectral Clustering (Single-view-SC) [13], [14];

4) Co-regularized Multi-view Spectral Clustering (Co-regularized-MSC) [15];

5) Proposed PW-CMSC and CT-CMSC.

Manifold-to-manifold Distance Measure: For fair comparisons, we use the manifold-to-manifold distance defined in (12) for all these five methods.

Multiple Views: We extract the intensity values, the local LBP features and SIFT features as the multi-view information of each image set, for which we first divide it into multiple 50% overlapped local patches. Then we extract LBP features and SIFT features from the patches of 10×10 pixels for the Honda/UCSD database [25] and 20×20 pixels for the Youtube Celebrities Database [18] respectively, as shown in Fig. 2.

B. Experimental Results

Experimental Results on Honda/UCSD Database [25]: There are 59 video sequences from 20 persons in this database. Each video sequence is fragmented into a set of frames, based on which we apply the Viola-Jones face detection algorithm [22] to detect face regions. Then we evaluate different methods under the same configurations as in [23]– [25] where 20 video sequences are selected from 20 different subjects as the gallery image sets, with the remaining 39 video sequences for testing. Then we randomly test each method 50 times and the averaged recognition rates are shown in Table I.

TABLE I
Averaged results on the Honda/UCSD database.

Methods	Recognition Rates
MLP [12]	$94.87\% \pm 0.00\%$
K-means [16]	$92.74\% \pm 2.53\%$
Single-view-SC [14]	$92.74\% \pm 2.16\%$
Co-regularized-MSC [15]	$96.41\% \pm 1.79\%$
Proposed PW-CMSC	$97.12\% \pm 1.64\%$
Proposed CT-CMSC	$\mathbf{98.97}\% \pm 1.37\%$

We observe in Table I that the proposed CT-CMSC method reaches 98.97% that outperforms all the other methods. It is noted that the variance of recognition rate achieved by MLP is zero. Different from the k-means based methods that start the clustering from k randomly selected points, the MLP model starts from only one seed point to construct local models. This makes MLP much more robust than the k-means based methods. Hence the other 5 clustering methods, which are based on the k-means clustering algorithm, can not achieve a recognition performance with zero-variance.

Experimental Results on Youtube Celebrities Database [18]: In this database, 1910 video sequences are collected from 47 celebrities from the Youtube website. Due to the low quality of this Database, we apply a tracking algorithm [26] to crop face regions across frames of each video. The same configurations as in [23] are considered here. We conduct 5-fold cross validation experiments, where the whole database is divided into 5 folds and each of them contains 423 videos from 47 persons, with 9 videos per person. In each fold, 3 image sets per person are randomly selected to constitute the gallery sets, with the remaining 6 videos per person as the probe sets. We repeat the random division 50 times and summarize the averaged recognition rates in Table II.

TABLE II
Averaged results on the Youtube Celebrities database.

Methods	Recognition Rates
MLP [12]	$54.78\% \pm 2.16\%$
K-means [16]	$55.52\% \pm 2.33\%$
Single-view-SC [14]	$55.32\% \pm 1.25\%$
Co-regularized-MSC [15]	$56.38\% \pm 0.94\%$
Proposed PW-CMSC	$\mathbf{65.98}\% \pm 0.79\%$
Proposed CT-CMSC	$65.40\% \pm 0.57\%$

In Table II, the proposed PW-CMSC method achieves the best recognition rate of 65.98%, which outperform all the other methods. Since the results shown in Table II are averaged over 5 folds where the training and testing data are randomly selected, the recognition rate of MLP is no longer of zero-variance.

IV. Conclusion

In this paper, we proposed two novel multi-view spectral clustering methods that are designed for solving the problem of face recognition based on image sets. The proposed CMSC methods learned the optimal embeddings from multiple views simultaneously such that each image set can be divided into subsets more precisely and robust. The efficiency and accuracy of the proposed methods were demonstrated through experiments.

References

[1] S.K. Zhou and R. Chellappa, *From sample similarity to ensemble similarity: Probabilistic distance measures in reproducing kernel hilbert space*, IEEE Transactions on PAMI, pp. 917–929, 2006.

[2] O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cipolla and T. Darrell, *Face recognition with image sets using manifold density divergence*, CVPR, pp. 581–588, 2005.

[3] G. Shakhnarovich, J. Fisher, and T. Darrell, *Face recognition from long-term observations*, ECCV, pp. 851–865, 2002.

[4] T.K. Kim, J. Kittler and R. Cipolla, *Discriminative learning and recognition of image set classes using canonical correlations*, IEEE Transactions on PAMI, pp. 1005–1018, 2007.

[5] L. Wang, X. Wang, and J. Feng, *Subspace distance analysis with application to adaptive bayesian algorithm for face recognition*, Pattern Recognition, pp. 456–464, 2006.

[6] F. Li, Q. Dai, W. Xu, and G. Er, *Weighted subspace distance and its applications to object recognition and retrieval with image sets*, Signal Processing Letters, pp. 227–230, 2009.

[7] O. Yamaguchi, K. Fukui, and K. Maeda, *Face recognition using temporal image sequence*, in FG, pp. 318–323, 1998.

[8] K. Fukui and O. Yamaguchi, *Face recognition using multiviewpoint patterns for robot vision*, Robotics Research, pp. 192–201, 2005.

[9] L.Wolf and A. Shashua, *Learning over sets using kernel principal angles*, The Journal of Machine Learning Research, pp. 913–931, 2003.

[10] T. Kozakaya, O. Yamaguchi, and K. Fukui, *Development and evaluation of face recognition system using constrained mutual subspace method*, IPSJ, pp. 951–959, 2004.

[11] J. Hamm and D.D. Lee, *Grassmann discriminant analysis: a unifying view on subspace-based learning*, in ICML, pp. 376–383, 2008.

[12] R. Wang, S. Shan, X. Chen, and W. Gao, *Manifold-manifold distance with application to face recognition based on image set*, in CVPR, pp. 1–8, 2008.

[13] U. Von Luxburg, *A tutorial on spectral clustering, Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[14] A. Y. Ng, M. I. Jordan, Y. Weiss, et al., *On spectral clustering: Analysis and an algorithm*, Advances in neural information processing systems, vol. 2, pp. 849–856, 2002.

[15] A. Kumar, P. Rai, and H. Daumé III, *Co-regularized multi-view spectral clustering*, Advances in Neural Information Processing Systems, vol. 24, pp. 1413–1421, 2011.

[16] J. MacQueen et al., *Some methods for classification and analysis of multivariate observations*, in Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol. 1, pp. 14, California, USA, 1967.

[17] K. Lee, J. Ho, M. Yang, and D. Kriegman, *Video-based face recognition using probabilistic appearance manifolds*, in CVPR, pp. I–313, 2003.

[18] M. Kim, S. Kumar, V. Pavlovic, and H. Rowley, *Face tracking and recognition with visual constraints in real-world videos*, in CVPR, pp. 1–8, 2008.

[19] T. Ojala, M. Pietikainen, and D. Harwood, *Evaluation of texture measures with classification based on kullback discrimination of distributions*, in Pattern Recognition, 1994.

[20] T. Ojala, M. Pietikainen, and D. Harwood, *Comparative study of texture measures with classification based on featured distributions*, in Pattern recognition, vol. 29, no. 1, pp. 51–59, 1996.

[21] D. G. Lowe, *Recognition from local scale-invariant features*, in the proceedings of the seventh IEEE international conference on computer vision, vol. 2, pp. 1150–1157, 1999.

[22] P. Viola and M. J. Jones, *Robust real-time face detection*, International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, 2004.

[23] Y. Hu, A. Mian, and R. Owens. *Sparse approximated nearest points for image set classification*. In CVPR, pp. 121–128, 2011.

[24] H. Cevikalp and B. Triggs. *Face recognition based on image sets*. In CVPR, pp. 2567–2573, 2010.

[25] K. Lee, J. Ho, M. Yang, and D. Kriegman. *Video-based face recognition using probabilistic appearance manifolds*. In CVPR, pp. 313–220, 2003.

[26] D. Ross, J. Lim, R. Lin, and M. Yang.*Incremental learning for robust visual tracking*. IJCV, pp. 125–141, 2008.

[27] Q. Qiu, V. M. Patel, P. Turagay and R. Chellappa, *Domain Adaptive Dictionary Learning*, ECCV, 2012.

[28] R. Wang and X. Chen, *Manifold discriminant analysis*, In CVPR, pp. 429–436, 2009.

[29] Z. Cui, S. Shan, H. Zhang, S. Lao and X. Chen, *Image Sets Alignment for Video-based Face Recognition*, CVPR 2012.