A NOVEL AND ROBUST FACE CLUSTERING METHOD VIA ADAPTIVE DIFFERENCE DICTIONARY

Jiaxiang Ren[1], *Shengjie Zhao*[*1], *Kai Yang*[1], *Brian Nlong Zhao*[2]

[1]Key Laboratory of Embedded System and Service Computing, Tongji University, Shanghai, China
[2]Shanghai High School International Division, No. 989, Baise Road, Shanghai, China
{renjiaxiang,shengjiezhao,kaiyang}@tongji.edu.cn, briannlongzhao@qq.com

ABSTRACT

High-dimensional data are ubiquitous in most real-world research areas, such as machine learning, image processing and so on. Actually, high-dimensional data that belong to the same classes tend to gather in their own low-dimensional subspaces. Recently, many subspace recovery algorithms based on the sparse representation such as Sparse Subspace Clustering (SSC), Low-Rank Representation (LRR) and their variants are proposed to address the subspace clustering problems.

In this paper, we propose a novel clustering method based on SSC, called Enhanced Sparse Subspace Clustering (ESSC), to deal with complicated face images under variant expressions, illuminations or disguises. Assuming that the variant expressions or disguises of the face images are sharable, we introduce the adaptive difference dictionary to extend the simple linear combination of face images in SSC with the combination of the specific features and the common features. Thus both of the accuracy and generalization of clustering are improved simultaneously.The experimental results on the AR face databases show that the proposed ESSC not only makes up to 9.0% improvements on accuracy compared with SSC, but also is more robust and scalable for dealing with larger face clustering problems (up to 400 samples from 100 subjects) under variant expressions, disguises or illuminations.

***Index Terms*—** Sparse subspace clustering, dictionary learning, face clustering

1. INTRODUCTION

Multiple subspaces clustering has been one of the hot and intriguing topics in the widespread applications in computer vision and image processing over the decades. This framework has been among the most successful in real-world problems

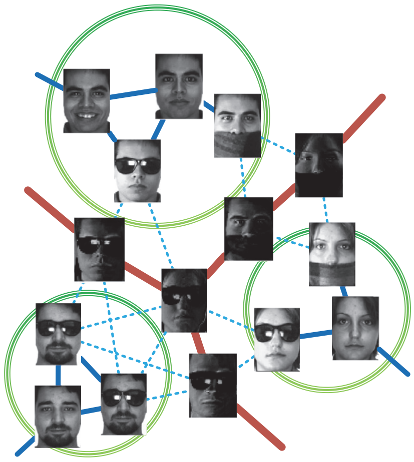

Fig. 1: Face clustering with the adaptive difference dictionary. The common features (connected with solid red lines) in the adaptive difference dictionary can help separating the face images near the intersections of subspaces. Thus the specific features (connected with solid blue lines) tend to gather in their own subspaces. The combination of the common features and specific features improves the robustness and accuracy of the proposed ESSC.

which is mainly due to the fact that data in the practical applications, such as face clustering and motion segmentation, are mostly drawn from multiple subspaces. For example, the image samples from the same subject in face clustering are the points from the corresponding subspace. In this paper, the image sample and data point are interchangeable. The subject and subspace are interchangeable. Inspired by this, many subspace algorithms such as Low-Rank Representation (LRR) [1, 2] and Sparse Subspace Clustering (SSC) [3, 4] are proposed to recover the discriminative structure of subspaces. LRR is a generalization of the recently established Robust Principal Component Analysis (RPCA) [5], extending the recovery of corrupted data from single subspace to multiple subspaces. Elhamifar *et al.* [4] demonstrated that, when each image in the intraclass subspace can be approximately represented as a linear combination of the other images, SSC can take advantage of multiple subspaces to ac-

The corresponding author is Shengjie Zhao. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2014CB340404 and in part by the National Science Foundation of China under Grant 61471267.

quire the excellent clustering results. These subspace clustering methods, such as LRR, SSC and many other applications [6, 7], can cluster face images from different subjects under variant illuminations with high accuracy. Both subspace clustering methods have theory basic and perform well in some real world problems. However, the intersections of subspaces are much more closed and indistinct in practice, so the performance of these methods degenerates severely. Using labeled data, the intraclass variant dictionary [8] can promotes the recognition of faces under variations. However, subspace clustering is unsupervised learning and can not use labeled data.

In this paper, we introduce the adaptive difference dictionary to make the subspace clustering method more generalized for complicated variations. Fig. 1 illustrates the basic idea of the proposed Enhanced Sparse Subspace Clustering (ESSC) for solving the face clustering problem under disguises. The adaptive differences play the role to separate the samples so that they can gather in their own subspaces. The principle of this dictionary is that any face with a kind of variations can be represented as a linear combination of the faces from the corresponding subject and the common facial features from the adaptive differences dictionary. Thus, by introducing the adaptive difference dictionary, the proposed ESSC can better separate the points lying in the intersections of subspaces. We adopt an Alternating Direction Method of Multipliers (ADMM) approach for solving the proposed ℓ_1 norm sparse optimization program. Thereafter a spectral clustering is employed to obtain clusters of face images.

The favorable results on the AR Face datebase [9] show that using adaptive difference dictionary can considerably improve the sparse representation based subspace clustering accuracy. Furthermore, the experiment results show there are 2 valuable observations: (1) The introduction of the adaptive difference dictionary improves the accuracy for clustering face images with complicated variations such as disguises. (2) The adaptive difference dictionary using the Local Binary Pattern (LBP) [10] can enhances the scalability and generalization for clustering more subjects.

2. RELATED WORKS

In this section, we give a brief review of the subspace clustering methods and SCC.

2.1. Subspace clustering

For several decades, researchers have proposed many practical and efficient methods related to the subspace clustering problem [11]. Fischler and Bolles [12] proposed a robust statistical approach to fit a subspace called Random Sample Consensus (RANSAC). Without knowing the number of subspaces, RANSAC can deal with outliers directly. But the computational complexity of RANSAC soon becomes the bottleneck with the increasing of the dimensions of subspaces. Vidal *et al.* [13] proposed Generalized Principal Component Analysis (GPCA) based on algebraic geometry, which fits the data points using a polynomial whose gradient at a point gives the vector belonging to the subspace including that point. The main drawback of GPCA is too sensitive to deal with outliers or noises.

The SIM algorithm [14] is a simple but effective solution for independent subspaces clustering. However, the structure of subspaces are not disjoint and the data samples are corrupted in real world. So the performance of SIM degenerates. Robust Principal Component Analysis (RPCA) [5] is a recently established method to remove the sparse outliers and recover the latent low rank structure. But for the complicated structure of multiple subspace, RPCA converges very slow. LRR is a generalization of RPCA, extending the recovery of corrupted data from single subspace to multiple subspaces. LRR can be used in face clustering with variant illuminations, motion segmentation, saliency detections, etc.

2.2. Sparse subspace clustering

SSC is another subspace clustering algorithm based on sparse representation and the self-expressiveness property of data. SSC leads to better performance in dealing with the data points near the intersections of subspaces. Given K subjects and the k-th subject with N_k columns of M-dimensional image samples $\boldsymbol{Y}_k$, SSC intends to find a sparse and non-trivial representation of the images themselves by solving the following problem

$$\boldsymbol{y}_i = \boldsymbol{Y}\boldsymbol{c}_i, \quad \text{s.t.} \quad c_{ii} = 0, \tag{1}$$

where $\boldsymbol{Y} = [\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_K] = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N] \in \mathbb{R}^{M\times N}$ is the input matrix, $N = \sum_{k=1}^{K} N_k$ and $\boldsymbol{y}_i \in \mathbb{R}^M$ is the i-th data point vector. The constraint $c_{ii} = 0$ removes the trivial solution which represents a data point as a linear combination of itself. Extending the solution to include all data points, Eq. (1) becomes

$$\boldsymbol{Y} = \boldsymbol{Y}\boldsymbol{C}, \quad \text{s.t.} \quad \text{diag}(\boldsymbol{C}) = \boldsymbol{0}, \tag{2}$$

where $\boldsymbol{C} = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N] \in \mathbb{R}^{N\times N}$, and diag$(\boldsymbol{C})$ is a column vector whose elements are the diagonal entries of $\boldsymbol{C}$. The data matrix $\boldsymbol{Y}$ at the right-hand of Eq. (2) plays the role of self-expressive dictionary.

Thereafter, a similarity graph based on the correlative matrix $\boldsymbol{C}$, which contains the information of subspace and data clusters, is built for the following conventional spectral clustering [15] to get the final results. In practice, most face images are corrupted by sparse outliers which degenerates the performance of clustering. This issue can be addressed with introducing the auxiliary outliers matrix $\boldsymbol{E} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N] \in \mathbb{R}^{M\times N}$, where $\boldsymbol{e}_i$ represents the i-th point's entries of observed outlier pixels. Thus, we can rewrite Eq. (2) as

$$\boldsymbol{Y} = \boldsymbol{Y}\boldsymbol{C} + \boldsymbol{E}, \quad \text{s.t.} \quad \text{diag}(\boldsymbol{C}) = \boldsymbol{0}. \tag{3}$$

The solution $(\boldsymbol{C}, \boldsymbol{E})$ in ℓ_1-norm can be computed via the convex programming tools [16, 17], which involve the ADMM approach and its computational complexity increases with the growing of the dimensions of subspace.

Eq. (3) can be used to cluster face images under variant illuminations and sparse outlier pixels with high accuracy. However, the accuracy degrades drastically for complicated variations such as disguises. This is due to the fact that the latent structures of these multiple subspaces are too complicated to recover only using the specific features.

3. SUBSPACE CLUSTERING BY ADAPTIVE DIFFERENCE DICTIONARY

In this section, we introduce the proposed ESSC algorithm for clustering face images under variant expressions, illuminations or disguises using sparse representation techniques. We improve SSC for clustering more complicated face images via the adaptive difference dictionary. The proposed ESSC consists of three main steps: (1) construction of the adaptive difference dictionary; (2) solving for sparse representation via adaptive difference dictionary; (3) spectral clustering.

3.1. Construction of the adaptive difference dictionary

Due to the fact that the shapes, appearances and expressions of faces are highly correlated, we make the following assumption:

Assumption 1. *Nearly all faces can be approximately represented as a linear combination of the variant face from the corresponding subject and the facial differences (i.e., the common features caused by different expressions, occlusions or disguises) from the adaptive differences.*

Based on the Assumption 1, we construct the adaptive difference dictionary with the differences between the most similar pairs of images, to make the subspace clustering method more generalized for variant expressions, illuminations or disguises. Given the data matrix $\boldsymbol{Y} = [\boldsymbol{y}_1, \dots, \boldsymbol{y}_N] \in \mathbb{R}^{M \times N}$ with N columns of M-dimensional images.

Firstly, we define the similarity of data points with the correlative coefficient matrix $\boldsymbol{C} \in \mathbb{R}^{N \times N}$ in SSC Eq. (3). Then, we compute the sparsity concentration ratio(SCR) of each sample $\boldsymbol{y}_i$ as

$$\text{SCR}(\boldsymbol{c}_i) \triangleq \frac{\max(\boldsymbol{c}_i)}{\|\boldsymbol{c}_i\|}. \tag{4}$$

Any data point $\boldsymbol{y}_*$ with SCR> 0.1 is used to construct the adaptive difference dictionary $\boldsymbol{D}$ as

$$\begin{aligned} \boldsymbol{D} &\triangleq \{\boldsymbol{d}_* | \forall \text{SCR}(\boldsymbol{c}_*) > 0.1\} \in \mathbb{R}^{M \times N_d}, \\ \boldsymbol{d}_* &\triangleq \boldsymbol{y}_* - \boldsymbol{y}_{\max(\boldsymbol{c}_*)}, \end{aligned} \tag{5}$$

where N_d is the number of points with SCR> 0.1.

3.2. Sparse optimization program via the adaptive difference dictionary

After constructing the adaptive difference dictionary $\boldsymbol{D}$, we construct this model to enhance the generalization of the proposed ESSC as

$$\boldsymbol{Y} = \begin{bmatrix}\boldsymbol{Y}\boldsymbol{D}\end{bmatrix} \begin{bmatrix}\boldsymbol{C} \\ \boldsymbol{B}\end{bmatrix} + \boldsymbol{Z}, \quad \text{s.t.} \quad \text{diag}(\boldsymbol{C}) = \boldsymbol{0}, \tag{6}$$

where $\boldsymbol{Y} \in \mathbb{R}^{M \times N}$, $\boldsymbol{D} \in \mathbb{R}^{M \times N_d}$, $\boldsymbol{C} \in \mathbb{R}^{N \times N}$, $\boldsymbol{B} \in \mathbb{R}^{N_d \times N}$ is the correlative coefficient matrix of the difference dictionary and $\boldsymbol{Z} \in \mathbb{R}^{M \times N}$ is the Gaussian-noise matrix. The corresponding constrained optimization program is

$$\begin{aligned} &\min \quad \left\| \begin{bmatrix}\boldsymbol{C} \\ \boldsymbol{B}\end{bmatrix} \right\|_1 + \frac{\lambda_z}{2} \|\boldsymbol{Z}\|_F^2 \\ &\text{s.t. } \boldsymbol{Y} = \begin{bmatrix}\boldsymbol{Y}\boldsymbol{D}\end{bmatrix} \begin{bmatrix}\boldsymbol{C} \\ \boldsymbol{B}\end{bmatrix} + \boldsymbol{Z},\ \boldsymbol{C}^{\mathrm{T}}\boldsymbol{1} = \boldsymbol{1},\ \text{diag}(\boldsymbol{C}) = \boldsymbol{0}. \end{aligned} \tag{7}$$

The ℓ_1 norm in Eq. (7) enhances the sparsity of the columns of $\begin{bmatrix}\boldsymbol{C} \\ \boldsymbol{B}\end{bmatrix}$ and the Frobenius norm promotes having small Gaussian-noise entries in the columns of $\boldsymbol{Z}$. Here, the parameter $\lambda_z = 20/\mu_z$ balances the two terms in Eq. (7), with

$$\mu_z = \min_{1 \le i \le N} \max_{1 \le j \ne i \le N} \left\| \boldsymbol{y}_i^{\mathrm{T}} \boldsymbol{y}_j \right\|_1. \tag{8}$$

We use the ADMM approach to solve Eq. (7). Thereafter, the coefficient matrix $\boldsymbol{C}$ which encodes the specific features is passed to the following spectral clustering.

3.3. Spectral clustering

This step involves a conventional spectral clustering [15]. Firstly, The optimal sparse coefficient matrix $\boldsymbol{C}$ should be normalized as $\boldsymbol{c}_i = \boldsymbol{c}_i / \|\boldsymbol{c}_i\|_\infty$. This normalization regularizes data points with different norms. Next, the similarity matrix is chosen as $\boldsymbol{W} = |\boldsymbol{C}| + |\boldsymbol{C}|^{\mathrm{T}}$, which means the similarity between the i-th point and the j-th point is equal to the sum of the absolute values of their correlative coefficients, *i.e.*, $|c_{ij}| + |c_{ji}|$. Finally, this similarity matrix is used as the affinity matrix of spectral clustering to cluster data points.

3.4. Geometric interpretation

In this section, we provide a geometric interpretation that why introducing the adaptive difference dictionary can improve the robustness of the proposed clustering method for variant expressions or disguises.

Considering a collection of image samples taken from K independent subspaces $\{\mathcal{S}_k\}_{k=1}^K$. Let $\boldsymbol{Y}_k$ denote N_k samples in $\mathcal{S}_k$ and let $\boldsymbol{Y}_{-k}$ denote all samples not in $\mathcal{S}_k$. Thus $\boldsymbol{Y} = [\boldsymbol{Y}_k, \boldsymbol{Y}_{-k}]$. For a data point $\boldsymbol{y}$, recalling the relationship between the solution of the proposed ℓ_1-minimization

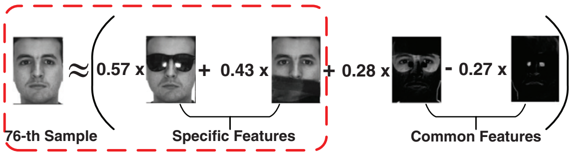

Fig. 2: The sparse correlative coefficients of the 76-th sample recovered by the proposed ESSC. This sample image can be constructed approximately by the 2 specific features and 2 common features.

program Eq. (7) and the symmetric convex polytopes constructed with the column vectors of $\boldsymbol{Y}_k$ and $\boldsymbol{Y}_{-k}$, we denote these symmetric convex polytopes as

$$\begin{aligned} \mathcal{P}_k &\triangleq \text{conv\_poly}(\pm \boldsymbol{y}_{1,k}, \ldots, \pm \boldsymbol{y}_{N_k,k}), \\ \mathcal{P}_{-k} &\triangleq \underset{\forall \boldsymbol{y}_i \notin \boldsymbol{Y}_k}{\text{conv\_poly}}(\pm \boldsymbol{y}_i, \ldots). \end{aligned} \tag{9}$$

Then the solution of Eq. (7) corresponds to the smallest scale coefficient $\alpha > 0$ for the scaled polytope $\alpha\mathcal{P}$ to reach $\boldsymbol{y}$. Thus the geometric interpretation of the condition to recover the sparse representation correctly is as follows.

> *The ℓ_1-minimization recovers the sparse representation correctly if and only if for any nonzero point $\boldsymbol{y} \in \mathcal{S}_k$ in the intersection of $\mathcal{S}_k$ and $\mathcal{S}_{-k}$, its shortest distance to $\mathcal{P}_k$ is shorter than the distance to $\mathcal{P}_{-k}$, i.e., the scaled polytope $\alpha\mathcal{P}_k$ reaches $\boldsymbol{y}$ with a smaller α compared to $\alpha\mathcal{P}_{-k}$.*

As illustrated in Fig. 3a, given 3 subspaces $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$, for a image sample $\boldsymbol{y} \in \mathcal{S}_1$, which is lying in the intersection of $\mathcal{S}_1$ and $\mathcal{S}_2 \oplus \mathcal{S}_3$, where $\oplus$ means the direct sum operation, the distance to $\mathcal{P}_1$ is shorter than to $\mathcal{P}_{-1}$, so the sparse representation recovers correctly. On the other hand, when the distribution of the samples in $\mathcal{S}_1$ is not uniform, such as in Fig. 3b, the subspace spanned by these samples is close to a line and orthogonal to the direction of $\boldsymbol{y}$. In this case, the distance to $\mathcal{P}_1$ is larger than to $\mathcal{P}_{-1}$, so the sparse representation recovers incorrectly.

To address this issue, we adopt the adaptive difference dictionary. The role of this dictionary is to filter out the common features from any sample, so that the discriminative specific features are preserved to improve the robustness of the proposed ESSC. As illustrated in Fig. 3c, the adaptive difference dictionary generates the common feature space $\mathcal{S}_\text{D}$ where any image sample can travel around to find the nearest polytope of the subspace correctly.

Finally, in order to illustrate how ESSC works, we conduct an experiment from 2 randomly selected subjects, each with 3 images under variant disguises, and 6 adaptive dictionary items. Fig. 2 shows the sparse correlative coefficients of the 76-th sample recovered by the proposed ESSC. This trial is to cluster face images under variant disguises with downsampled images in the size of 55×40. One can see from Fig. 2 that the 76-th sample with neutral face can almost be recovered by the top 2 specific features and the top 2 common features. The 2 specific features are from the samples of the corresponding subject (one with sunglasses and the other one with scarf) and the 2 common features are from the adaptive difference images. Thus this test sample will be clustered with the 2 highly correlative samples regardless of the other 2 difference dictionary items. That is because the adaptive difference dictionary only plays the role of outliers and they are not involved in the following spectral clustering procedure. Besides the disguises, similar results have been observed when clustering face images with different illuminations or expressions.

4. EXPERIMENTAL RESULTS

In this section, we conduct experiments on the AR Face database to demonstrate the efficiency of the proposed clustering method. The objective is to cluster face images according to their subjects. We introduce the dictionary to replace the simple linear combinations of face images in SSC with the robust combinations of the specific features (*i.e.*, the image from the corresponding subject) and the common features (*i.e.*, the features that can be shared across the subjects such as sunglasses), which play the roles of improving accuracy and generalization respectively.

Implementation. For fair comparisons, all the clustering methods compute the clustering accuracy in the same way that is used in SSC [4]. The parameter λ_z is computed in Eq. (8). Besides, we adopt the same ADMM approach to solve the constrained optimization programs of SSC, which ensures that the performance difference is only caused by the adoption of the adaptive difference dictionary.

4.1. Clustering variant face images

This experiment is to test the validation of the adaptive difference dictionary for the variant illuminations,expressions or disguises. We use the AR Face database which consists of over 4,000 color frontal-face images from 126 subjects [9]. For each subject, 26 images were taken from two independent sessions and each session includes 4 images under variant expressions, 3 neutral expression images under variant illuminations, 2 neutral expression images under variant disguises (*i.e.*, scarf or sunglasses) and 4 neutral expression images under variant illuminations and disguises. In this experiment, we randomly choose a subset of the database consisting of 50 male subjects and 50 female subjects, and then convert the color images into gray scale with the size of $M = 165 \times 120$. For feature selection, the original images in the size of $M = 165 \times 120$, the downsampled images in the size of $M = 55 \times 40$, the randomface (random projection) in the size of $M = 2,200$ and the uniform LBP feature in the size of $M = 5,192$ are taken into consideration. The LBP

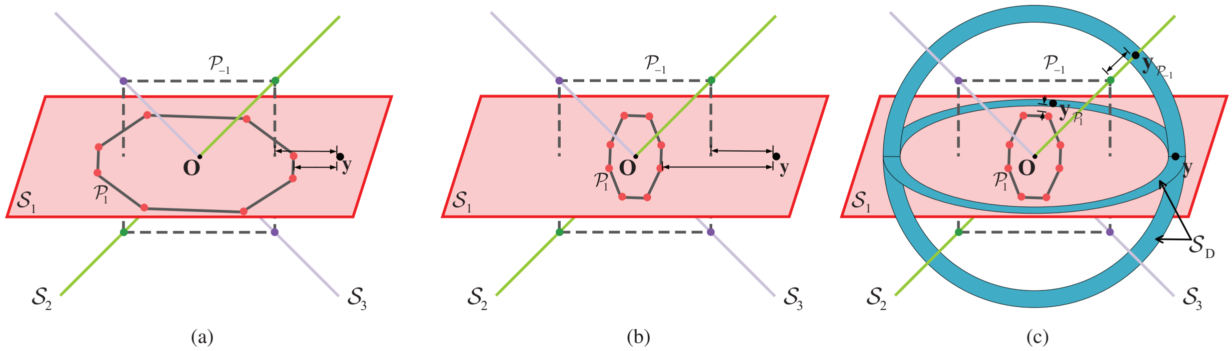

Fig. 3: The sparse representation for recovering an image sample $\boldsymbol{y} \in \mathcal{S}_1$ in the intersection of $\mathcal{S}_1$ and $\mathcal{S}_2 \oplus \mathcal{S}_3$. (a) The distance to $\mathcal{P}_1$ is shorter than to $\mathcal{P}_{-1}$, so the sparse representation recovers correctly. (b) The distribution of the samples in $\mathcal{S}_1$ is odd because the spanned subspace is close to a line. The distance to $\mathcal{P}_1$ is larger than to $\mathcal{P}_{-1}$, so the sparse representation recovers incorrectly. (c) The adaptive difference dictionary generates the common feature space $\mathcal{S}_{\mathrm{D}}$, where any image sample can travel around to find the nearest polytope of the subspace correctly.

feature in the size of $M = 5,192$ is extracted by concatenating the histograms of the 59 uniform patterns for the original image of $M = 165 \times 120$ whose cell size is 15×15 [10]. We also consider a "Roubst PCA + SSC" method which means removing occlusions and noises by Robust PCA and then applying SSC. Because there are no labelled data for clustering, Robust PCA is used simultaneously to all the image samples.

Images from the two sessions are test separately and then the results are averaged.

Table 1 enumerates the error rates of the different algorithms on the AR Database using different features for $K = 100$ subjects. The variations include expressions, illuminations or disguises. The clustering error for ESSC is the lowest in almost all cases which confirms the effectiveness of the adaptive difference dictionary. Especially for variant illuminations, ESSC using LBP feature achieves **0.33%** error rate.

The proposed ESSC using LBP features are suitable for the face representation and achieve the best clustering accuracy for All variations in the AR Face database.

4.2. Clustering scalability

In this experiment, we test the scalability of the proposed ESSC and compare its performance with SSC for different features. It is prohibitive to cluster all the combinations of $K \in \{2, 5, 10, 20, 40, 60, 80, 100\}$ subjects for each test. Therefore, we average the results over 40 trials for each test to get the convincing clustering accuracy. Different subjects are selected randomly every time.

Fig. 4 illustrates the clustering error rates for variant disguises on the AR database as a function of the number of subjects K. In each trial, all $N_k = 3$ image samples for each subject are involved. More subspaces lead to more indistinct intersections, so the accuracy drops as the number of subject

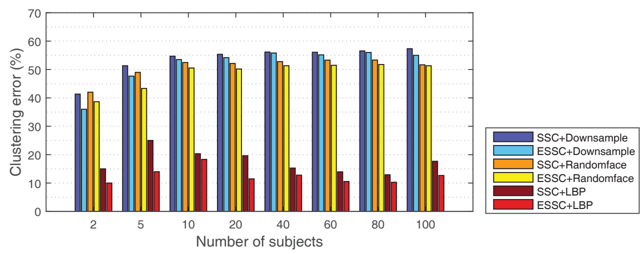

Fig. 4: Clustering error rates for variant disguises on the AR database as a function of the number of subjects.

increases using downsample or randomface. But the LBP features can enhance the scalability of ESSC up to 100 subjects with the error rate of **12.67%**, which is 5% less than SSC. This could be the fact that the LBP related features tighten up the spanned multiple subspaces.

5. DISCUSSIONS AND CONCLUSIONS

In this paper, we generalize the subspace clustering algorithm SSC for dealing with the complicated face images under variant expressions, illuminations or disguises. After introducing the adaptive difference dictionary, we extend the simple linear combinations of face images in SSC to the robust combinations of the specific features and the common features, which play the role of improving accuracy and generalization respectively. Furthermore, the LBP features tighten up the spanned multiple subspaces to make the proposed ESSC more scalable even for clustering 300 samples from 100 subjects. These robust improvements are validated by numerous experiments on publicly available face databases.

Table 1: Clustering Error Rates (%) of Different Algorithms on the AR Database Using Different Features for $K = 100$ Subjects

Variation Sample×Subject	Feature *(Dimension)*	Method			
		LRR	SSC	RPCA+SSC	ESSC
Expression	Downsample(55×40)	73.00	14.50	16.00	**13.00**
4×100	LBP(5192)	70.75	8.75	**4.25**	10.00
Illumination	Downsample(55×40)	65.67	31.00	**30.33**	31.00
3×100	LBP(5192)	67.67	6.00	6.00	**0.33**
Disguise	Downsample(55×40)	68.00	57.33	60.33	**55.00**
3×100	LBP(5192)	65.33	17.67	14.33	**12.67**

6. REFERENCES

[1] G. Liu, H. Xu, J. Tang, Q. Liu, and S. Yan, "A deterministic analysis for lrr," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 3, pp. 417–430, March 2016.

[2] Guangcan Liu and Shuicheng Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Proceedings of the 2011 International Conference on Computer Vision*. IEEE, 2011, pp. 1615–1622.

[3] Ehsan Elhamifar and René Vidal, "Sparse subspace clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 2790–2797.

[4] Ehsan Elhamifar and Rene Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.

[5] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," *Journal of the ACM*, vol. 58, no. 3, pp. 11:1–11:37, June 2011.

[6] X. Pei, Z. Lyu, C. Chen, and C. Chen, "Manifold adaptive label propagation for face clustering," *IEEE Transactions on Cybernetics*, vol. 45, no. 8, pp. 1681–1691, Aug 2015.

[7] L. Zhang, W. Zuo, and D. Zhang, "Lsdt: Latent sparse domain transfer learning for visual adaptation," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1177–1191, March 2016.

[8] W. Deng, J. Hu, and J. Guo, "Extended src: Undersampled face recognition via intraclass variant dictionary," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1864–1870, Sept 2012.

[9] A. M. MARTINEZ, "The ar face database," *CVC Technical Report*, vol. 24, 1998.

[10] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, Dec 2006.

[11] René Vidal, "Subspace clustering," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 52–68, 2011.

[12] Martin A Fischler and Robert C Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[13] Rene Vidal, Yi Ma, and Shankar Sastry, "Generalized principal component analysis (gpca)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1945–1959, 2005.

[14] João Paulo Costeira and Takeo Kanade, "A multibody factorization method for independently moving objects," *International Journal of Computer Vision*, vol. 29, no. 3, pp. 159–179, 1998.

[15] Andrew Y Ng, Michael I Jordan, Yair Weiss, et al., "On spectral clustering: Analysis and an algorithm," *Advances in Neural Information Processing Systems*, vol. 2, pp. 849–856, 2002.

[16] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[17] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale ℓ_1-regularized least squares," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, Dec 2007.