Face Recognition with L1-norm Subspaces

Federica Maritato[1], Ying Liu[2], Stefania Colonnese[1], and Dimitris A. Pados[2]

[1]Department of Telecommunication Engineering, La Sapienza University of Rome, Rome, 00185 Italy

[2]Department of Electrical Engineering, The State University of New York at Buffalo, Buffalo, NY 14260 USA

ABSTRACT

We consider the problem of representing individual faces by maximum L_1-norm projection subspaces calculated from available face-image ensembles. In contrast to conventional L_2-norm subspaces, L_1-norm subspaces are seen to offer significant robustness to image variations, disturbances, and rank selection. Face recognition becomes then the problem of associating a new unknown face image to the "closest," in some sense, L_1 subspace in the database. In this work, we also introduce the concept of adaptively allocating the available number of principal components to different face image classes, subject to a given total number/budget of principal components. Experimental studies included in this paper illustrate and support the theoretical developments.

Keywords: Classification, dimensionality reduction, eigenfaces, face recognition, L_1 principle component analysis, L_2 principle component analysis, subspace signal processing.

1. INTRODUCTION

Face recognition has been a task of growing importance in the past decade with a wide range of commercial and law enforcement applications. Among the different face recognition techniques available, subspace learning based algorithms have attracted significant interest. The foundation of face recognition by subspace learning has been the early (L_2-norm) principal component analysis (PCA) work, in which a linear projection matrix is learned by maximizing the data variance in the projection subspace.[1] The columns of the computed projection matrix are the so-called principal components, or features, or eigenfaces. Nevertheless, the features extracted by L_2-PCA can be easily affected by outliers in the training data. To address this problem, Kwak[2] proposed to compute the principal components by L_1-norm maximization, solved approximately by an iterative algorithm. Markopoulos et al.[3] developed for the first time in the literature an optimal exponential-time algorithm and an optimal polynomial-time algorithm to solve the L_1-norm maximization problem exactly. A fast suboptimal algorithm was also later developed.[4] The developed[3,4] L_1-PCA algorithm was successfully applied to face recognition tasks with outlier corrupted training face images.[5]

While it is shown[5] that L_1-PCA outperforms L_2-PCA for face recognition with corrupted data, the computed "common" L_1 subspace does not exploit the prior knowledge about the class labels of the training samples. In this work, we consider the computation of "individual" L_1 subspaces, in which an L_1 subspace is computed for each class of face images using the training face images from that particular class. Then, classification is performed by a nearest subspace (NS) criterion, which assigns a class label to the unknown testing sample according to the subspace that most closely represents the testing sample. To control the total number of calculated principal components over all samples, we extend the developed "individual" L_1-subspace procedure to an adaptive version, in which a fixed budget of principal components is allocated across different classes of face images according to within-class variances. This scheme adaptively determines the number of principal components required by different classes for feature extraction and is demonstrated to be more effective than the "fixed" L_1-principal-components allocation scheme.

The rest of the paper is organized as follows. In Section 2, we briefly review related works on L_2 and L_1 principal-component-analysis-based face recognition schemes. In Section 3, we introduce the proposed "individual" L_1-subspace calculation method for face recognition and its adaptive extension. Section 4 presents experimental studies on three face data sets that strongly support our algorithmic developments. Finally, we draw some conclusions in Section 5.

2. BACKGROUND

2.1 L_2-subspace Face Recognition

In common L_2-PCA formulation for face recognition,[1] the N training face images of pixel size $m \times n$ are vectorized as $\mathbf{x}_i \in \mathbb{R}^D$, $D = mn$, $i = 1, ..., N$. Consider the data matrix that contains all training images $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N] \in \mathbb{R}^{D \times N}$. L_2-PCA aims at finding a projection matrix $\mathbf{R} \in \mathbb{R}^{D \times K}$, $K \leq D$, that minimizes the sum of the element-wise squared error between the original matrix and its rank-K representation,

$$(\mathbf{R}_{L_2}, \mathbf{S}_{L_2}) = \arg \min_{\substack{\mathbf{R} \in \mathbb{R}^{D \times K}, \mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}_K \\ \mathbf{S} \in \mathbb{R}^{N \times K}}} \|\mathbf{X} - \mathbf{R}\mathbf{S}^{\mathrm{T}}\|_2, \tag{1}$$

which is equivalent to solving

$$\mathbf{R}_{L_2} = \arg \max_{\substack{\mathbf{R} \in \mathbb{R}^{D \times K} \\ \mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}_K}} \|\mathbf{X}^{\mathrm{T}}\mathbf{R}\|_2. \tag{2}$$

The well known solution to problem (2) is given by the K singular vectors of $\mathbf{X}$ that correspond to the maximum K singular values. The face recognition problem is then tackled by projecting an unknown face image $\mathbf{x}$ onto $\mathbf{R}_{L_2}$ to obtain it L_2-subspace representation $\mathbf{y} = \mathbf{R}^T\mathbf{x} \in \mathbb{R}^K$, followed by nearest neighbor (NN) classification in the L_2-subspace.

2.2 L_1-subspace Face Recognition

If we consider the L_1-norm instead of the L_2-norm in (1), (2), the two problems are not equivalent anymore. If we focus on the projection maximization problem, we seek matrix

$$\mathbf{R}_{L_1} = \arg \max_{\substack{\mathbf{R} \in \mathbb{R}^{D \times K} \\ \mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}_K}} \|\mathbf{X}^{\mathrm{T}}\mathbf{R}\|_1. \tag{3}$$

Jointly in D, N, (3) is an NP-hard problem[3,6]. This is not true for fixed D, however.[3] The first ever optimal algorithm with complexity polynomial in D has been developed.[3] In fact, if we consider the solution for a single principal component we have

$$\mathbf{r}_{L_1} = \arg \max_{\mathbf{r} \in \mathbb{R}^D, \|\mathbf{r}\|_2 = 1} \left\|\mathbf{X}^{\mathrm{T}}\mathbf{r}\right\|_1 \tag{4}$$

and[3]

$$\mathbf{r}_{L_1} = \frac{\mathbf{X}\mathbf{b}_{\mathrm{opt}}}{\|\mathbf{X}\mathbf{b}_{\mathrm{opt}}\|_2} \tag{5}$$

where $\mathbf{b}_{\mathrm{opt}}$ is a binary vector obtained by

$$\mathbf{b}_{\mathrm{opt}} = \arg \max_{\mathbf{b} \in \{\pm 1\}^N} \mathbf{b}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{b}. \tag{6}$$

A fast algorithm is developed[4] for solving (5), in which (6) is solved approximately by iteratively flipping the bit in $\mathbf{b}$ which most negatively contributes to the L_2 projection energy. In the face recognition algorithm of Johnson and Savakis,[5] the fast algorithm[4] for computing $\mathbf{r}_{L_1}$ and the multiple L_1-PCA algorithm[2] are combined such that K L_1-principal components are calculated in a greedy conditional optimization manner. In particular, the first L_1-principal component $\mathbf{r}_1$ is calculated using the fast method[4] and then the contribution of $\mathbf{r}_1$ is removed form each data sample,

$$\mathbf{x}_i^{(update)} = \mathbf{x}_i - \mathbf{r}_1\mathbf{r}_1^{\mathrm{T}}\mathbf{x}_i \tag{7}$$

where $i = 1, 2, ..., N$. The same procedure is continued until K L_1 principal components $\mathbf{r}_1, \mathbf{r}_2, .., \mathbf{r}_K$ are found.

3. PROPOSED METHOD

In this section, we first propose and describe "individual" L_1-subspace face representation per person. Then, we deal with the extension to "adaptive individual" L_1-subspace representation subject to a total number of principal components across individuals.

3.1 Individual L_1-subspace Face Representation

Instead of training a "common" L_1-subspace for robust face recognition, we propose an "individual" L_1-subspace face representation algorithm, in which an "individual" subspace is trained/calculated for each class of face images using all training samples from the corresponding class. We consider a total number of C classes. Without loss of generality, each class has N training samples that can be organized in matrix form as $\mathbf{X}^{(j)} = \left[\mathbf{x}_1^{(j)}, \mathbf{x}_2^{(j)}, ..., \mathbf{x}_N^{(j)}\right] \in \mathbb{R}^{D\times N}$ where $j = 1, 2, ..., C$ is the class index and each column is a vectorized training image of class j that has D pixels. To compute the L_1-subspace of $\mathbf{X}^{(j)}$, we first subtract the sample-mean $\boldsymbol{\mu}^{(j)} = \frac{1}{N}\mathbf{X}^{(j)}\mathbf{1}_N$ from each column of $\mathbf{X}^{(j)}$ so that the training samples are zero-centered. Then, a rank-K L_1-subspace representation of class j is computed as follows. We calculate the first L_1-principal component using the fast algorithm,[4] i.e. we find

$$\mathbf{q}_1^{(j)} = \arg \max_{\mathbf{q}\in\mathbb{R}^D, \|\mathbf{q}\|_2=1} \left\|\mathbf{q}^T\mathbf{X}^{(j)}\right\|_1. \tag{8}$$

Then, the contribution of $\mathbf{q}_1^{(j)}$ is removed from the entire training ensemble as follows

$$\mathbf{X}^{(j)} = \mathbf{X}^{(j)} - \mathbf{q}_1^{(j)}\mathbf{q}_1^{\mathrm{T}(j)}\mathbf{X}^{(j)}. \tag{9}$$

The updated training ensemble $\mathbf{X}^{(j)}$ is used to calculate the next principal component and the representation continues until the desired number of components K is reached. The above greedy "individual" L_1 subspace is calculated for each class $j = 1, 2, ..., C$ to obtain $\mathbf{Q}_{L_1}^{(j)} = \left[\mathbf{q}_1^j, \mathbf{q}_2^j, ..., \mathbf{q}_K^j\right]$.

Next, for classification we adopt a nearest-subspace (NS) approach. For each test face image $\mathbf{x}_t$, we subtract the mean of the *j-th* class $\boldsymbol{\mu}^{(j)}$ from $\mathbf{x}_t$. Then, the zero-centered test data point is projected onto the *j-th* L_1 subspace $\mathbf{Q}_{L_1}^{(j)}$ and the reconstruction error using the *j-th* class L_1-subspace is calculated as $\left\|(\mathbf{x}_t - \mu^j) - \mathbf{Q}_{L_1}^{(j)}\mathbf{Q}_{L_1}^{\mathrm{T}(j)}(\mathbf{x}_t - \mu^j)\right\|_2$. The procedure is performed for each calculated "individual" L_1-subspace and the classification criterion is

$$\widehat{j} = \arg \min_{1\leq j\leq C} \left\|(\mathbf{x_t} - \mu^j) - \mathbf{Q}_{L_1}^{(j)}\mathbf{Q}_{L_1}^{\mathrm{T}(j)}(\mathbf{x_t} - \mu^j)\right\|_2 \tag{10}$$

where $\widehat{j}$ is the determined class of $\mathbf{x}_t$.

3.2 Extension to Adaptive Individual L_1-subspace Calculation

In Section 3.1, the "individual" L_1-subspaces are computed with the same rank K (number of L_1-principal components). In this section, we consider the problem of allocating a total number of KC principal components to different classes allowing the computed L_1-subspace of different classes to have -possibly- different rank. Intuitively, the within-class sample variance indicates the rank value required to represent the subspace of the particular face image class. Therefore, we propose to utilize the following method to allocate the total number of KC principal components across the C classes. We partition the per-class rank allocation value into two parts, $K_{fix}^{(j)}$ and $K_{var}^{(j)}$ where $K_{fix}^{(j)}$ stands for the fixed guaranteed number of principal components allocated to class j and $K_{var}^{(j)}$ stands for the variable number of principal components allocated to class j. The following condition should of course be satisfied

$$\sum_{j=1}^{C}\left(K_{fix}^{(j)} + K_{var}^{(j)}\right) = KC. \tag{11}$$

In our algorithm, $K_{fix}^{(j)}$, $j = 1, ..., C$ are fixed and known and $K_{var}^{(j)}$, $j = 1, ..., C$ are determined by the within-class variance defined as

$$var^{(j)} = \mathrm{tr}(\mathbf{Cov}^{(j)}) \tag{12}$$

where

$$\mathbf{Cov}^{(j)} = \frac{1}{N}\left[\sum_{i=1}^{N}(\mathbf{X}_i^{(j)} - \boldsymbol{\mu}^{(j)})(\mathbf{X}_i^{(j)} - \boldsymbol{\mu}^{(j)})^{\mathrm{T}}\right] \tag{13}$$

is the covariance matrix for the *j-th* class. The variable number of principal components for class j is calculated by

$$K_{var}^{(j)} = \left(KC - \sum_{j=1}^{C} K_{fix}^{(j)} \right) \frac{var^{(j)}}{\sum_{j=1}^{C} var^{(j)}}. \quad (14)$$

4. EXPERIMENTAL RESULTS

In this section, we experiment with three different databases, Extended Yale Face database,[7] ORL database,[8] and Aberdeen database,[9] to illustrate and evaluate our theoretical developments.

4.1 Extended Yale Face Database

The Extended Yale Face database[1] has $C = 8$ classes. Each class has 25 images of size 50×50 pixels ($D = 2500$).

We carried out 50 independent experiments. In each experiment, $N = 8$ images per class are randomly selected for training and the remaining 17 images per class are used for testing. Both the training and testing data sets are partially corrupted by "salt and pepper" noise patches, as seen in Fig. 1 for example. The size of the noise patches is randomly chosen from the options $\{15 \times 15, 20 \times 20, 25 \times 25, 30 \times 30\}$. For class 1 and 2, 10% of the images are corrupted. The corruption percentage is set at 30% for class 3 and 4, 50% for class 5 and 6, and 70% for class 7 and 8.

For "common" subspaces, the number of L_1 (or L_2) principal components calculated varies from 1 to 20. For "individual" subspaces (fixed-rank or adaptive-rank), the fixed budget of L_1 (or L_2) principal components varies from 8 to 48.

Fig. 2 shows the recognition error rates for different schemes in comparison. We observe that "common" L_1-PCA performs better than "common" L_2-PCA due to its resistance to outliers (noise patches). The same conclusion can be drawn for the "individual" subspaces. In particular, the adaptive-rank L_1 subspace approach achieves a lower recognition error rate than its fixed-rank counterpart.

4.2 ORL Database

The ORL Face database has $C = 8$ classes. Each class has 10 images of size 50×50 pixels ($D = 2500$).

We carried out 50 independent experiments. In each experiment, $N = 7$ images per class are randomly selected for training and the remaining 3 images per class are used for testing. Both the training and testing data sets are partially corrupted by "salt and pepper" noise patches, as in the example of Fig. 3. The size of the noise patches is randomly chosen from the set $\{25 \times 25, 30 \times 30, 35 \times 35, 40 \times 40\}$. For class 1 and 2, 30% of the images are corrupted. This corruption percentage is set at 40% for class 3 and 4, 60% for class 5 and 6, and 80% for class 7 and 8.

For "common" subspaces, the number of L_1 (or L_2) principal components calculated varies from 1 to 20. For "individual" subspaces (fixed-rank or adaptive-rank), the fixed budget of L_1 (or L_2) principal components varies from 8 to 40.

Fig. 4 shows the recognition error rates for different schemes in comparison. We observe that "common" L_1-PCA performs better than "common" L_2-PCA due to its resistance to outliers (noise patches). The same conclusion can be drawn for the "individual" subspaces. In particular, the adaptive-rank L_1 subspace achieves a lower recognition error rate than its fixed-rank counterpart.

[1]The cropped images are used in the experiment.

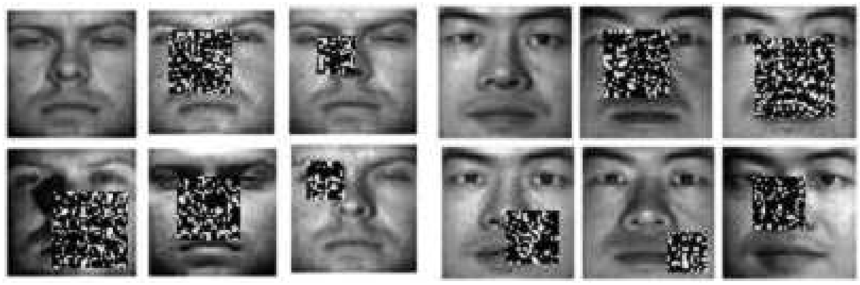

Figure 1. Two different subjects from the Extended Yale Face database and the applied noise patches.

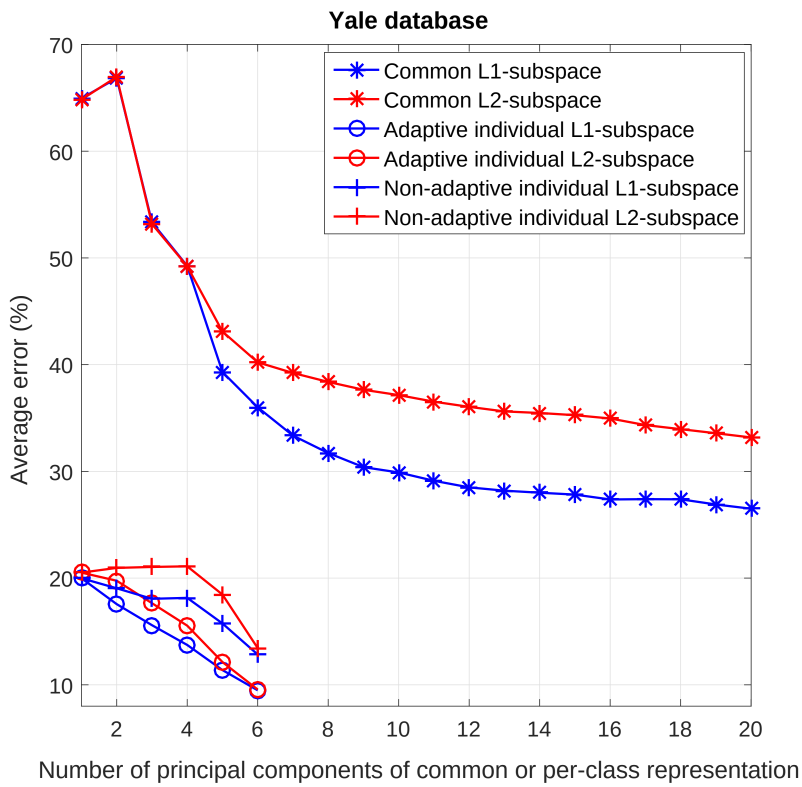

Figure 2. Comparison of common L_1 (L_2)-subspace, non-adaptive individual L_1 (L_2)-subspace, and adaptive individual L_1 (L_2)-subspace classification on the Extended Yale Face database.

4.3 Aberdeen Database

The Aberdeen Face database has $C = 8$ classes. Each class has 18 images of size 50×50 pixels ($D = 2500$).

We carried out 50 independent experiments. In each experiment, $N = 8$ images per class are randomly selected for training and the remaining 10 images per class are used for testing. Both the training and testing data sets are partially corrupted by "salt and pepper" noise patches, as in the example of Fig. 5. The size of the noise patches is randomly chosen from the set $\{15 \times 15, 20 \times 20, 25 \times 25, 30 \times 30\}$. For class 1 and 2, 10% of the images are corrupted. The corruption percentage is set at 30% for class 3 and 4, 50% for class 5 and 6, and 70% for class 7 and 8.

For "common" subspaces, the number of L_1 (or L_2) principal components calculated varies from 1 to 20. For "individual" subspaces (fixed-rank or adaptive-rank), the fixed budget of L_1 (or L_2) principal components varies from 8 to 48.

Fig. 6 shows the recognition error rates for different schemes in comparison. We observe that "common"

Figure 3. Two different subjects from the ORL database and the applied noise patches.

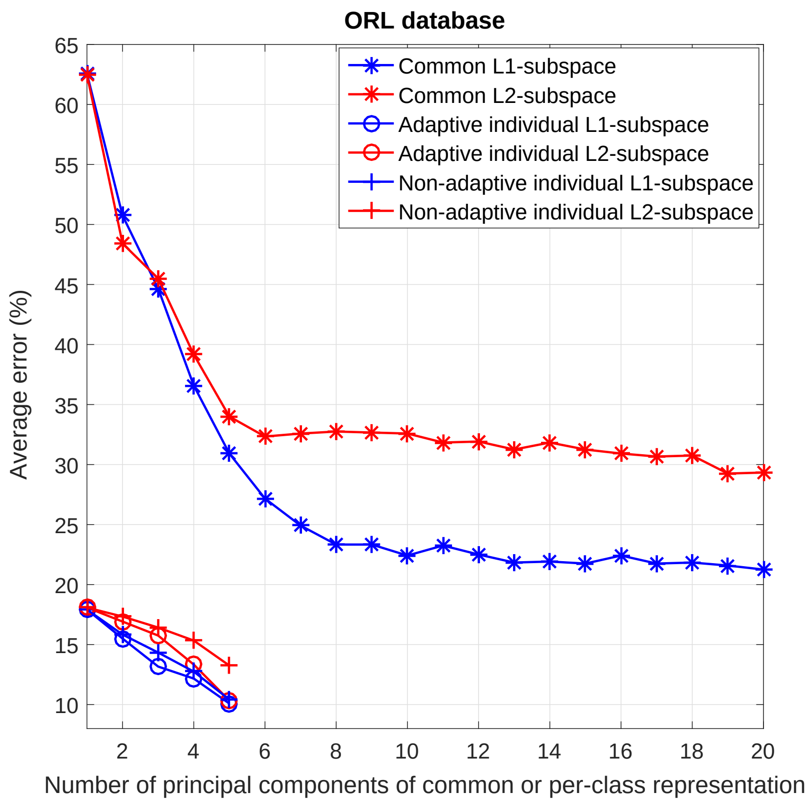

Figure 4. Comparison of common L_1 (L_2)-subspace, non-adaptive individual L_1 (L_2)-subspace, and adaptive individual L_1 (L_2)-subspace classification on the ORL database.

L_1-PCA performs better than "common" L_2-PCA due to its resistance to outliers (noise patches). The same conclusion can be drawn for the "individual" subspaces. In particular, the adaptive-rank L_1 subspace achieves a lower recognition error rate than its fixed-rank counterpart.

5. CONCLUSION

In this paper, we proposed to represent individual faces by maximum L_1-norm projection subspaces calculated from available face-image ensembles. We considered adaptive rank formulation of the "individual" subspace, in which we allocate principal components based on the variance value of each class. We demonstrated the superiority of L_1-PCA versus L_2-PCA and the benefits of optimized use of principal components, saving in components where a small number of them is sufficient and adding more principal components where necessary under a class variance criterion.

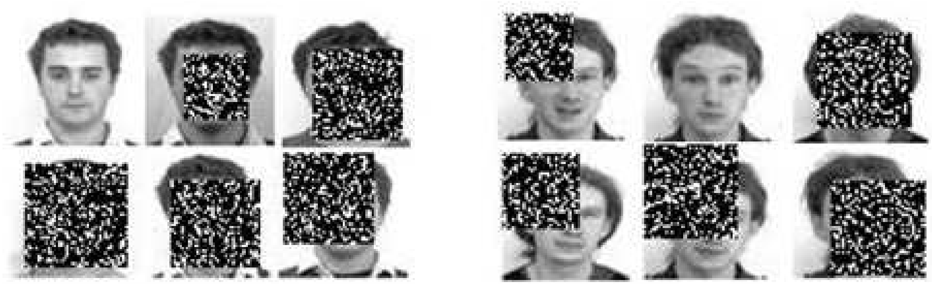

Figure 5. Two different subjects from the Aberdeen database and the applied noise patches.

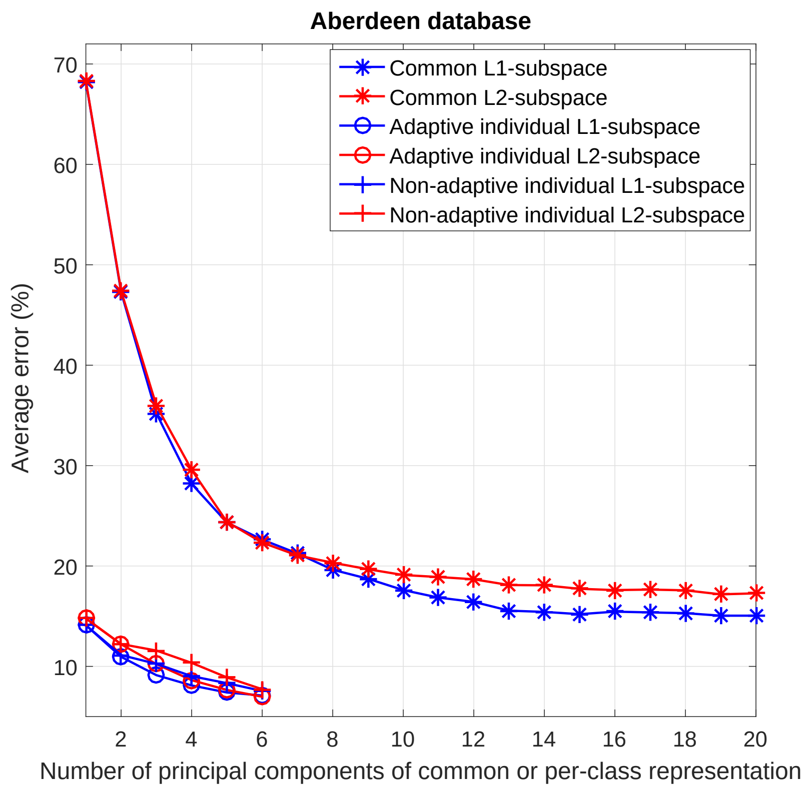

Figure 6. Comparison of common L_1 (L_2)-subspace, non-adaptive individual L_1 (L_2)-subspace, and adaptive individual L_1 (L_2)-subspace classification on the Aberdeen database.

REFERENCES

1. M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Maui, HI, June 1991, pp. 586-591.
2. N. Kwak, "Principal component analysis based on L1-norm maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, pp. 1672-1680, Sep. 2008.
3. P. P. Markopoulos, G. Karystinos, and D. A. Pados, "Optimal algorithms for L_1-subspace signal processing," *IEEE Trans. Signal Process.*, vol. 62, pp. 5046-5058, Oct. 2014.
4. S. Kundu, P. P. Markopoulos, and D. A. Pados, "Fast computation of the L1-principal component of real-valued data," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process. (ICASSP)*, Florence, Italy, May 2014, pp. 8028-8032.
5. M. Johnson and A. Savakis, "Fast L1-eigenfaces for robust face recognition," in *Proc. IEEE Western NY Image and Signal Process. Workshop*, Rochester, NY, Nov. 2014, pp. 1-5.

6. M. McCoy and J. A. Tropp, “Two proposals for robust PCA using semidefinite programming,” *Electron. J. Statist.*, vol. 5, pp. 1123-1160, Jun. 2011.
7. A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, “From few to many: Illumination cone models for face recognition under variable lighting and pose,” *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 23, pp. 643-660, June 2001.
8. F. S. Samaria, A. C. Harter, “Parameterisation of a stochastic model for human face identification,” in *Proc. IEEE Workshop Applications of Computer Vision*, Sarasota, FL, Dec. 1994, pp. 138-142.
9. [Online]. Available: `http://pics.stir.ac.uk/2D_face_sets.htm.`