L1-Norm-Based 2DPCA

Xuelong Li, *Senior Member, IEEE*, Yanwei Pang, *Senior Member, IEEE*, and Yuan Yuan, *Senior Member, IEEE*

***Abstract*—In this paper, we first present a simple but effective L1-norm-based two-dimensional principal component analysis (2DPCA). Traditional L2-norm-based least squares criterion is sensitive to outliers, while the newly proposed L1-norm 2DPCA is robust. Experimental results demonstrate its advantages.**

***Index Terms*—L1 norm, outlier, subspace, two-dimensional principal component analysis (2DPCA).**

I. Introduction

IMAGES are intrinsically two-dimensional (2-D) signals in which spatial relationship is very important for kinds of analysis and/or recognition tasks. Many feature extraction techniques have been developed to preprocess an image by scanning it into a long vector. These image-as-vector methods work in many cases; unfortunately, as the underlying spatial (structural) information is destroyed, the extraction procedure is normally not optimal for the extraction of the most representative or discriminating features. These image-as-vector methods include principal component analysis (PCA) [22], linear discriminant analysis (LDA) [5], [7], [13], locality preserving projections (LPP) [3], etc. Among them, PCA is considered to be capable of extracting the most expressive features, while LDA is employed to extract the most discriminating ones. LPP is linear approximation to nonlinear manifold learning techniques.

To further exploit the spatial (structural) information carried by images, many image-as-matrix (or tensor [18]) methods have been developed. In contrast to the image-as-vector methods, the image-as-matrix ones treat an image as a two-order tensor, and their objective functions are expressed as functions of the image matrix instead of the high-dimensional image vector. Representative image-as-matrix methods are 2-D PCA (2DPCA) [8], [21], binary 2DPCA [14], 2-D LDA [23], general tensor discriminant analysis [17], discriminant locally linear embedding [11], correlation tensor analysis [4], multilinear PCA (MPCA) [12], and rank-one projections [19]. All these methods have contributed much to both academic research and industrial applications.

However, both the aforementioned image-as-vector and image-as-matrix techniques are based on the L2- or Frobenius-norm distance (error). It has been well known that L2-norm-based techniques are not robust in the sense that outlying measurements can arbitrarily skew the solution from the desired solution. It is also well recognized that L1 norm is more robust to outliers [1], [20]. Therefore, several methods have been developed to adopt the L1-norm metric for measuring the error between the original data point and its reconstruction [2]. L1-norm PCA (L1-PCA) [6] and R1-PCA [2] find the optimal basis vectors through linear or quadratic programming. L1-norm-based PCA methods are computationally expensive. PCA-L1 [9] is a fast and robust L1-norm based method. PCA-L1 converts the L1-norm based variance in feature space into a direct sum of inner product of the basis vector and signed training points. The basis vector is updated by the sum of re-signed training points. As a result, this convergence procedure is fast.

Although robust against outliers, PCA-L1 and L1-PCA, as well as traditional PCA (we name it as PCA-L2 hereinafter), do not fully exploit the spatial redundancies of images. In PCA-L1, each image is preprocessed by stacking the pixels into a high-dimensional vector, and thus, the natural spatial relationship is destroyed.

This paper generalizes PCA-L1 to 2DPCA-L1. 2DPCA-L1 is an L1-norm version of 2DPCA. For notational clarity, we call traditional 2DPCA [21] as 2DPCA-L2. 2DPCA-L1 has three major advantages: 1) It is robust to outliers because of utilizing L1-norm-based metric; 2) much less spatial information is lost because of the higher order image representation; and 3) intensive eigendecomposition is avoided, and the iteration process of 2DPCA-L2 is very simple.

The rest of this paper is organized as follows. Section II introduces 2DPCA. PCA-L1 is given in Section III. 2DPCA-L1 is described in Section IV. Section V reports the experimental results. Section VI concludes this paper.

II. 2DPCA

2DPCA [21] is a well-known image-as-matrix subspace learning algorithm. Compared with that in PCA, more spatial information is utilized in 2DPCA, which benefits classification and reconstruction.

Let $\mathbf{X}_i, i = 1, \ldots, N$, denote N training image matrices. The size of the matrix $\mathbf{X}_i$ is $h \times w$. 2DPCA aims to find a projection matrix $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_d] \in \mathbb{R}^{w \times d}$ that minimizes the mean-square reconstruction error. The image-as-matrix representation

Manuscript received July 13, 2008; revised December 18, 2008; accepted April 8, 2009. Date of publication January 15, 2010; date of current version July 16, 2010. This work was supported in part by the National Natural Science Foundation of China under Grants 60605005 and 60975001, by the 100 Talents Program of the Chinese Academy of Sciences, by the Open Project Program of the State Key Laboratory of CAD&CG under Grant A0902, Zhejiang University, and by the Open Project Program of National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. This paper was recommended by Associate Editor F. Hoffmann.

X. Li is with the State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China (e-mail: xuelong_li@opt.ac.cn).

Y. Pang is with the School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: pyw@tju.edu.cn).

Y. Yuan is with the School of Engineering and Applied Science, Aston University, Birmingham B4 7ET, U.K. (e-mail: y.yuan1@aston.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCB.2009.2035629

reduces the learning process of 2DPCA to an eigendecomposition problem of a small covariance matrix [21]

$$\mathbf{C} = \sum_{i}^{N} (\mathbf{X}_i - \overline{\mathbf{X}})^T (\mathbf{X}_i - \overline{\mathbf{X}}) \in \mathbb{R}^{w \times w} \tag{1}$$

where $\overline{\mathbf{X}}$ is the mean image matrix

$$\overline{\mathbf{X}} = \frac{1}{N} \sum_{i}^{N} \mathbf{X}_i. \tag{2}$$

The covariance matrix in 2DPCA is of size $w \times w$, while it is $d \times d$ in image-as-vector-based PCA, where $d = h \times w$ and h and w are the image height and width, respectively. Thus, 2DPCA is efficient in computation.

III. L1-Norm-Based PCA

The proposed 2DPCA-L1 method is a 2-D generalization of PCA-L1 [9]. Among many L1-norm-based PCA [2], [6], [9], PCA-L1 is deemed to be the most simple and efficient one.

Let $\mathbf{u} \in \mathbb{R}^d$ be the principal vector to be obtained. The training data are given by $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$, with $\mathbf{x}_i \in \mathbb{R}^{D \times 1}$ being a data vector and $D = h \times w$ being the image size. In the following, the subspace dimension is set to be one to simplify the procedure. The objective of PCA-L1 is to maximize the L1-norm variance $f(\mathbf{u})$ in feature space

$$f(\mathbf{u}) = \|\mathbf{u}^T \mathbf{X}\|_1 = \sum_{i=1}^{N} |\mathbf{u}^T \mathbf{x}_i| \tag{3}$$

subject to

$$\|\mathbf{u}\|_2 = 1 \tag{4}$$

where $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the L1 and L2 norms, respectively.

One of the novelties of PCA-L1 is to convert the sum of the absolute values in (3) into an ordinary sum. This is done by introducing a polarity function p_i

$$p_i = \begin{cases} 1, & \text{if } \mathbf{u}^T \mathbf{x}_i \geq 0 \\ -1, & \text{if } \mathbf{u}^T \mathbf{x}_i < 0. \end{cases} \tag{5}$$

Armed with this polarity function, (3) can be written as

$$f(\mathbf{u}) = \sum_{i=1}^{N} |\mathbf{u}^T \mathbf{x}_i| = \sum_{i=1}^{N} p_i \mathbf{u}^T \mathbf{x}_i. \tag{6}$$

The maximization process is performed in iterative manner

$$\mathbf{u}(t) = \frac{\sum_{i=1}^{N} p_i(t-1)\mathbf{x}_i}{\left\| \sum_{i=1}^{N} p_i(t-1)\mathbf{x}_i \right\|_2} \tag{7}$$

where the polarity function at time $t-1$ is chosen so that

$$p_i(t-1) = \begin{cases} 1, & \mathbf{u}(t-1)^T \mathbf{x}_i \geq 0 \\ -1, & \mathbf{u}(t-1)^T \mathbf{x}_i < 0. \end{cases} \tag{8}$$

The solution of PCA-L1 is guaranteed to reach a local maximum. Refer to [9] for the proof. It has been shown that PCA-L1 converges faster than R1-PCA [2].

As can be seen from (7), the computation of PCA-L1 is as simple as a weighted sum of the training samples and the weight is either 1 or -1.

IV. L1-Norm-Based 2DPCA

In this section, we extend PCA-L1 to 2DPCA-L1. Suppose that $\mathbf{u} \in \mathbb{R}^{w \times 1}$ is the first principal vector to be found. Because $\mathbf{u}$ is the convergence result of iterations, we use $\mathbf{u}(t)$ to denote the result of iteration t. Then, the corresponding feature $\mathbf{y}_i \in \mathbb{R}^{h \times 1}$ of image $\mathbf{x}_i \in \mathbb{R}^{h \times w}$ is given by

$$\mathbf{y}_i = \mathbf{x}_i \mathbf{u} = \begin{bmatrix} \mathbf{x}_{i1} \\ \mathbf{x}_{i2} \\ \vdots \\ \mathbf{x}_{ih} \end{bmatrix} \mathbf{u} \tag{9}$$

where $\mathbf{x}_{ij} \in \mathbb{R}^{1 \times w}$ is the jth row of $\mathbf{x}_i$.

The proposed method is aimed at maximizing the L1-norm variance in low-dimensional feature space

$$f(\mathbf{u}) = \sum_{i=1}^{N} |\mathbf{y}_i| = \sum_{i=1}^{N} \sum_{j=1}^{h} |\mathbf{x}_{ij} \mathbf{u}| \tag{10}$$

subject to

$$\|\mathbf{u}\|_2 = 1. \tag{11}$$

To remove the absolute-value operation in (1), we define the polarity function $p_{ij}(t)$ as

$$p_{ij}(t) = \begin{cases} 1, & \text{if } |\mathbf{x}_{ij}\mathbf{u}(t)| > 0 \\ -1, & \text{if } |\mathbf{x}_{ij}\mathbf{u}(t)| \leq 0. \end{cases} \tag{12}$$

Substituting (12) into (10), it follows that

$$f\left(\mathbf{u}(t)\right) = \sum_{i=1}^{N} \sum_{j=1}^{h} |\mathbf{x}_{ij}\mathbf{u}(t)| = \sum_{i=1}^{N} \sum_{j=1}^{h} p_{ij}(t) \mathbf{x}_{ij} \mathbf{u}(t). \tag{13}$$

To maximize (13), we let $\mathbf{u}(t+1)$ to be updated by

$$\mathbf{u}(t+1) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{h} p_{ij}(t) \mathbf{x}_{ij}^T}{\left\| \sum_{i=1}^{N} \sum_{j=1}^{h} p_{ij}(t) \mathbf{x}_{ij}^T \right\|_2}. \tag{14}$$

After $\mathbf{u}(t+1)$ is updated, $p_{ij}(t+1)$ should be recomputed according to (12).

The convergence of iteration equation (14) can be justified as follows. Because $p_{ij}(t+1)$ is designed for $\mathbf{u}(t+1)$, thus, it can guarantee that

$$p_{ij}(t+1)\mathbf{x}_{ij}\mathbf{u}(t+1) \geq 0 \tag{15}$$

but for $p_{ij}(t)$, it may hold that

$$p_{ij}(t)\mathbf{x}_{ij}\mathbf{u}(t+1) < 0. \tag{16}$$

It then follows that

$$\sum_{i=1}^{N}\sum_{j=1}^{h}|\mathbf{x}_{ij}\mathbf{u}(t+1)| = \sum_{i=1}^{N}\sum_{j=1}^{h}[p_{ij}(t+1)\mathbf{x}_{ij}\mathbf{u}(t+1)] \geq \sum_{i=1}^{N}\sum_{j=1}^{h}p_{ij}(t)\mathbf{x}_{ij}\mathbf{u}(t+1). \quad (17)$$

Investigating the right-hand side of (17), we have

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{j=1}^{h}p_{ij}(t)\mathbf{x}_{ij}\mathbf{u}(t+1)\\
&= \left[\sum_{i=1}^{N}\sum_{j=1}^{h}p_{ij}(t)\mathbf{x}_{ij}\right]\mathbf{u}(t+1)\\
&= \left[\sum_{i=1}^{N}\sum_{j=1}^{h}p_{ij}(t)\mathbf{x}_{ij}\right]\left[\frac{\sum_{i=1}^{N}\sum_{j=1}^{h}p_{ij}(t)\mathbf{x}_{ij}^T}{\left\|\sum_{i=1}^{N}\sum_{j=1}^{h}p_{ij}(t)\mathbf{x}_{ij}^T\right\|_2}\right]\\
&\geq \left[\sum_{i=1}^{N}\sum_{j=1}^{h}p_{ij}(t)\mathbf{x}_{ij}\right]\left[\frac{\sum_{i=1}^{N}\sum_{j=1}^{h}p_{ij}(t-1)\mathbf{x}_{ij}^T}{\left\|\sum_{i=1}^{N}\sum_{j=1}^{h}p_{ij}(t-1)\mathbf{x}_{ij}^T\right\|_2}\right]\\
&= \left[\sum_{i=1}^{N}\sum_{j=1}^{h}p_{ij}(t)\mathbf{x}_{ij}\right]\mathbf{u}(t)\\
&= \sum_{i=1}^{N}\sum_{j=1}^{h}|\mathbf{x}_{ij}\mathbf{u}(t)|. \quad (18)
\end{aligned}$$

Combining (17) and (18) yields

$$\begin{aligned}
f(\mathbf{u}(t+1)) &= \sum_{i=1}^{N}\sum_{j=1}^{h}|\mathbf{x}_{ij}\mathbf{u}(t+1)| \geq \sum_{i=1}^{N}\sum_{j=1}^{h}|\mathbf{x}_{ij}\mathbf{u}(t)|\\
&= f(\mathbf{u}(t)). \quad (19)
\end{aligned}$$

■

Equation (19) implies that each step of updating $\mathbf{u}$ increases the value of the objective function, and thus, the value of $\mathbf{u}$ moves toward the optimal solution. Our experimental results show that the algorithm converges in about a dozen of iterations. The iteration stops when it satisfies that $f(\mathbf{u}(t+1)) = f(\mathbf{u}(t))$.

Equations (12) and (14) are used to compute the first principal vector $\mathbf{u}_1$. To compute $\mathbf{u}_k$, with $k > 1$, the training data have to be updated

$$\mathbf{x}_{ij}^k = \mathbf{x}_{ij}^{k-1} - \mathbf{x}_{ij}^{k-1}\mathbf{u}_{k-1}\mathbf{u}_{k-1}^T, \qquad i=1,\ldots,N, \quad j=1,\ldots,w \quad (20)$$

with $\mathbf{x}_{ij}^0 = \mathbf{x}_{ij}$.

In the following, it can be proven that $\mathbf{u}_{k-1}$ is orthogonal to $\mathbf{u}_k$. Because $\mathbf{u}_k$ is a linear combination of $\mathbf{x}_{ij}^{k-1}$ [see (14)], we need to examine the value of $\mathbf{x}_i^k\mathbf{u}_{k-1}$. Because

$$\begin{aligned}
\mathbf{x}_i^k\mathbf{u}_{k-1} &= \mathbf{x}_i^{k-1}\mathbf{u}_{k-1} - \mathbf{x}_i^{k-1}\mathbf{u}_{k-1}\mathbf{u}_{k-1}^T\mathbf{u}_{k-1}\\
&= \mathbf{x}_i^{k-1}\mathbf{u}_{k-1} - \mathbf{x}_i^{k-1}\mathbf{u}_{k-1}\\
&= 0 \quad (21)
\end{aligned}$$

so it holds that $\mathbf{u}_k^T\mathbf{u}_{k-1} = 0$.

V. Experimental Results

As an unsupervised subspace analysis algorithm, the newly proposed 2DPCA-L1 does not utilize the information of class labels. Therefore, to make the comparison fair, only unsupervised subspace analysis algorithms were used in the comparisons. These algorithms are 2DPCA-L2 [21], MPCA [12], PCA-L2, and PCA-L1 [9]. 2DPCA-L1, 2DPCA-L2, and MPCA are image-as-matrix-based methods, while PCA-L2 and PCA-L1 are image-as-vector-based methods. Note that PCA-L2 here is, in fact, the classical PCA.

2DPCA-L1 and other existing subspace analysis algorithms can be used for data compression, object recognition, image retrieval, and other applications. However, in this paper, we mainly focus on their performance of data compression. As in [9], we compare the performances of different methods in terms of reconstruction error for inliers when outliers are present in the training data.

Three popular databases, FERET face database [15], Yale face database [25] and ETH-80 image database [10], were used in our experiments. Let $(\mathbf{x}_1,\ldots,\mathbf{x}_{N_I})$ be the N_I inlying images, $\overline{\mathbf{x}}$ be the mean image, and $\mathbf{U}$ be the projection matrix obtained by 2DPCA-L1, 2DPCA-L2, or MPCA. Then, the average reconstruction error is

$$e = \frac{1}{N_I}\sum_{i=1}^{N_I}\left\|\mathbf{x}_i - \left((\mathbf{x}_i - \overline{\mathbf{x}})\mathbf{U}\mathbf{U}^T + \overline{\mathbf{x}}\right)\right\|_F \quad (22)$$

where $\|\cdot\|_F$ stands for the Frobenius norm. Similarly, the average reconstruction errors of two image-as-vector methods, namely, PCA-L2 and PCA-L1, can also be defined.

In the experiments carried out upon the FERET database, the face images marked with two-character strings, i.e., "ba," "bj," "be," "bk," "bf," "bd," and "bg," were used. Thus, the entire data set includes 1379 images of 197 different subjects, with 7 images per subject. All these images were aligned according to the center points of the eyes. The images are of size 64 by 64. Some sample images are shown in Fig. 1. The 721 images of 103 subjects were used for training, while the remaining 658 images of 94 subjects were used for testing. Our purpose is to investigate the representation capacity of the subspace learned from the training set when the training set is spoiled by outliers. About 20% of images in the training set were used to generate the outliers. Each outlying image was formed by adding rectangle noise [9]. Fig. 2 shows some samples of the outliers (outlying images). The sizes of the rectangles range from 15×10 to 60×60. Within each rectangle, the pixel value is randomly set to be either 0 or 255.

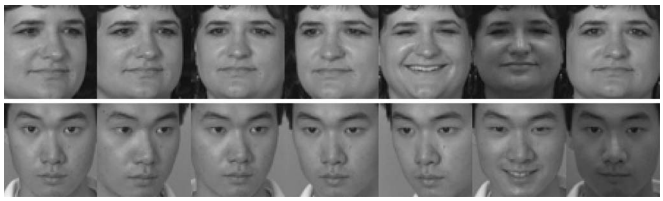


Fig. 1. Sample images of the normalized FERET face database.

Fig. 2. Some generalized outlying face images of the FERET database.

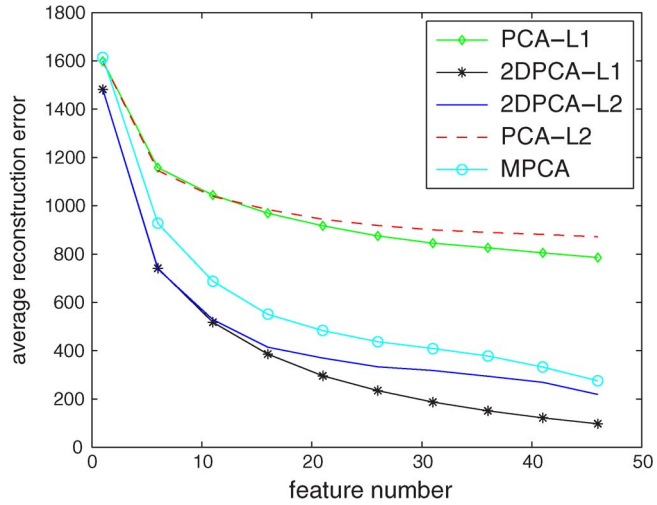


Fig. 3. Average reconstruction error versus feature number on the FERET database.

Fig. 3 shows the average reconstruction errors of the following five algorithms: PCA-L2, PCA-L1, 2DPCA-L2, MPCA, and 2DPCA-L1, respectively. It can be found that the performance of 2DPCA-L1 is better than that of 2DPCA-L2 when the feature number is larger than ten. Both 2DPCA-L1 and 2DPCA-L2 gave better performance than PCA, PCA-L1, and MPCA. It is well recognized that MPCA was proven to be powerful for gait recognition and other biometric-recognition applications. While when there are outliers in the training set, the solution of MPCA is also biased by the outliers.

Fig. 4 shows the reconstructed images of different methods when the feature number is 30. The first column of Fig. 4 gives two original images of the FERET database. The second, third, and fourth columns are the corresponding images reconstructed by the algorithms of 2DPCA-L1, 2DPCA-L2, and MPCA. It can be found that there is little visual difference between the original image and the image reconstructed by the proposed 2DPCA-L1. However, the obvious degradation of the results of 2DPCA-L2 and MPCA can also be seen.

The proposed 2DPCA-L1 algorithm converges in about 27 iterations for a single basis vector. The training procedure of the newly proposed method upon the FERET database took 5 min on a Pentium 1.80-GHz processor with 1.0 GB of RAM with a MATLAB implementation, while it took 3 s for traditional 2DPCA.

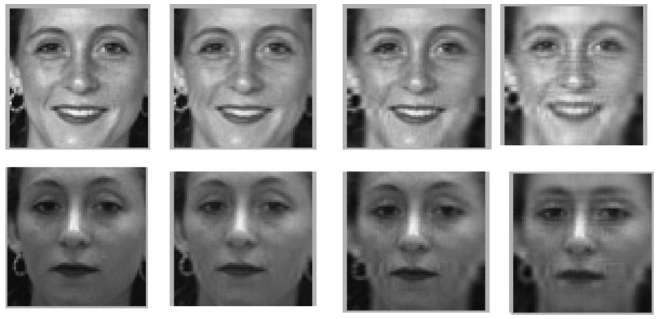


Fig. 4. Original images (in the FERET database) and reconstructed images. The first column shows the original images. The second, third, and fourth columns are the images reconstructed by 2DPCA-L1, 2DPCA-L2, and MPCA, respectively.

Fig. 5. Sample images of the Yale face database.

The second experiment was conducted on the Yale face database [25]. The experimental configuration is almost the same as that in [9]. However, due to the random sampling operation, the experimental data are not exactly identical to that in [9]. Therefore, the experimental results of PCA-L1 and PCA-L2 are not perfectly identical to those reported in [9].

The Yale face database contains 15 individuals, with 11 images for each individual. The images were captured under different illumination and expression conditions. Sample images are shown in Fig. 5. The image size is 100×80. The images were aligned and cropped in the same manner as that of the first experiment. In our experiment, the grayscale images were not preprocessed, except that 20% of images were used to generate outliers by adding rectangle noise. The data set contains 132 inlying face images and 33 outlying face images. The formation of rectangle noise is the same as that in the aforementioned FERET experiment, except that the maximum size of the rectangles of the former is 100×80.

Five methods (i.e., PCA-L2, PCA-L1, 2DPCA-L2, GLRAM [24], and 2DPCA-L1) were applied on the Yale data set that includes both the inlying and outlying images. MPCA is equivalent to GLRAM, with the exception of the initialization procedure and termination criterion [12]. The inlying images were reconstructed, and the average reconstruction error is shown in Fig. 6. The figure tells that PCA-L1 outperforms PCA-L2 when more than 12 features are used and that 2DPCA-L1 outperforms 2DPCA-L2 when the feature number is larger than 20. When the feature number is less than 12, the reconstruction error of PCA-L1 is slightly lower than that of PCA-L2. As to 2DPCA-based methods, when the feature number is less than 20, the performances of 2DPCA-L1 and 2DPCA-L2 are identical. This

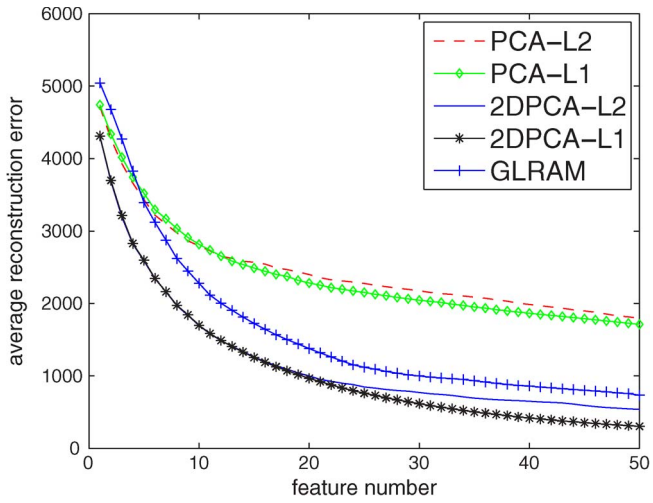


Fig. 6. Average reconstruction error versus feature number on the Yale database.

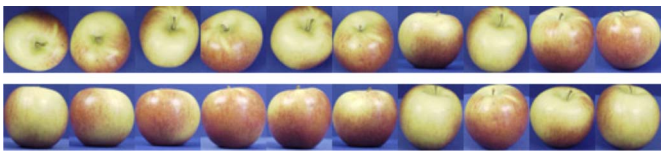


Fig. 7. Example of the apple images in the ETH-80 image set. Note that, in our experiments, the images are converted into grayscale images, and thus, color information is discarded.

demonstrates the superiority of 2DPCA-L1 and the importance of introducing L1 norm into 2DPCA.

In the third experiment, a small subset of the ETH-80 image set was used. The object concerned here was an apple. Fig. 7 shows some of the apple images. The image size is 128×128. There are 41 apple images in the data set. To introduce server outliers, 8 images among the 41 images were randomly chosen and then added rectangle noise. The sizes of the rectangles range from 35×30 to 128×128. Note that only grayscale information was utilized. No other preprocessing step was applied.

The average reconstruction errors of 2DPCA-L2 and 2DPCA-L1 are shown in Fig. 8. It can be observed that, when more than six features are used, 2DPCA-L1 is significantly superior to 2DPCA-L2. Outlying data degrade 2DPCA-L2 much more than they degrade 2DPDA-L1.

It is found that, for different databases, there are different thresholds for the number of features, starting from which the proposed 2DPCA-L1 clearly shows the advantage over traditional 2DPCA-L2. The thresholds are related to the nature of not only the outliers but also the intrinsic variations of each database themselves. On the other hand, it can also be observed that the superiority of 2DPCA-L1 over 2DPCA-L2 is not obvious when the feature number (subspace dimension) is small. A possible explanation of this phenomenon is as follows. In fact, outliers are similar to noise in the sense that they both have a lot of high-frequency components. Generally speaking, high-frequency components can only be captured by minor basis vectors/matrices instead of principal ones. Therefore, the

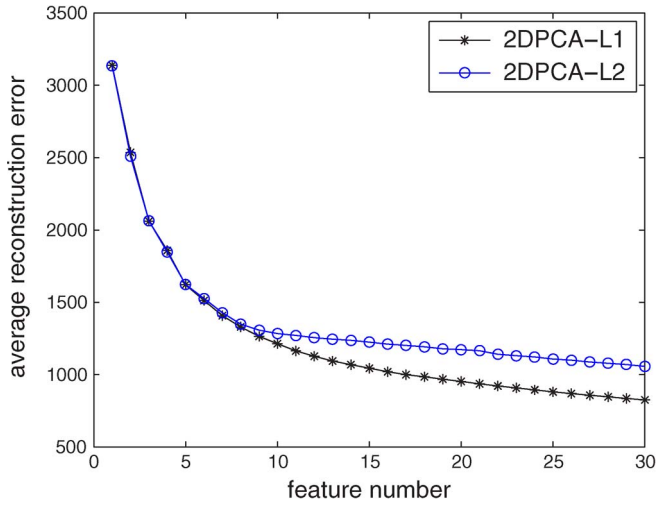


Fig. 8. Average reconstruction error versus feature number on a subset of the ETH-80 image set.

proposed algorithm exhibits its superiority when the subspace dimension is larger than some value.

VI. Conclusion and Future Works

In this paper, an L1-norm-based 2-D principal component (termed as 2DPCA-L1) has been put forward, which avoids computation of the eigendecomposition process and its iteration step is easy to be performed. 2DPCA-L1 not only makes good use of spatial (structural) information but is also robust to outliers. These merits come from its employment of L1 norm and image-as-matrix representation form. If only few features are needed, its performance is comparable with that of 2DPCA-L2. Both sets of experiments have demonstrated the efficacy of the proposed 2DPCA-L1. The robust tensor analysis algorithm is being studied by making use of L1 norm.

Appendix

Proof of the inequality in (18)

$$\left[\sum_{i=1}^{N}\sum_{j=1}^{h} p_{ij}(t)\mathbf{x}_{ij}\right]\left[\frac{\sum_{i=1}^{N}\sum_{j=1}^{h} p_{ij}(t)\mathbf{x}_{ij}^T}{\left\|\sum_{i=1}^{N}\sum_{j=1}^{h} p_{ij}(t)\mathbf{x}_{ij}^T\right\|_2}\right]$$

$$= \frac{\left\|\sum_{i=1}^{N}\sum_{j=1}^{h} p_{ij}(t)\mathbf{x}_{ij}^T\right\|_2^2}{\left\|\sum_{i=1}^{N}\sum_{j=1}^{h} p_{ij}(t)\mathbf{x}_{ij}^T\right\|_2}$$

$$= \left\|\sum_{i=1}^{N}\sum_{j=1}^{h} p_{ij}(t)\mathbf{x}_{ij}^T\right\|_2 \geq \left\|\sum_{i=1}^{N}\sum_{j=1}^{h} p_{ij}(t)\mathbf{x}_{ij}^T\right\|_2$$

$$\times \cos\left(\sum_{i=1}^{N}\sum_{j=1}^{h} p_{ij}(t)\mathbf{x}_{ij}^T, \sum_{i=1}^{N}\sum_{j=1}^{h} p_{ij}(t-1)\mathbf{x}_{ij}^T\right)$$

$$= \left\| \sum_{i=1}^{N} \sum_{j=1}^{h} p_{ij}(t) \mathbf{x}_{ij}^T \right\|_2 \times \frac{\left(\sum_{i=1}^{N} \sum_{j=1}^{h} p_{ij}(t) \mathbf{x}_{ij}^T \right)^T \left(\sum_{i=1}^{N} \sum_{j=1}^{h} p_{ij}(t-1) \mathbf{x}_{ij}^T \right)}{\left\| \sum_{i=1}^{N} \sum_{j=1}^{h} p_{ij}(t) \mathbf{x}_{ij}^T \right\|_2 \times \left\| \sum_{i=1}^{N} \sum_{j=1}^{h} p_{ij}(t-1) \mathbf{x}_{ij}^T \right\|_2}$$

$$= \left[\sum_{i=1}^{N} \sum_{j=1}^{h} p_{ij}(t) \mathbf{x}_{ij} \right] \left[\frac{\sum_{i=1}^{N} \sum_{j=1}^{h} p_{ij}(t-1) \mathbf{x}_{ij}^T}{\left\| \sum_{i=1}^{N} \sum_{j=1}^{h} p_{ij}(t-1) \mathbf{x}_{ij}^T \right\|_2} \right].$$

ACKNOWLEDGMENT

The authors would like to thank the Editor and all the four anonymous reviewers for their constructive comments on the earlier versions of this paper.

REFERENCES

[1] M. J. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *Int. J. Comput. Vis.*, vol. 19, no. 1, pp. 57–91, Jul. 1996.

[2] C. Ding, D. Zhou, X. He, and H. Zha, "R1-PCA: Rotational invariant L1-norm principal component analysis for robust subspace factorization," in *Proc. Int. Conf. Mach. Learn.*, 2006, pp. 281–288.

[3] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

[4] Y. Fu and T. S. Huang, "Image classification using correlation tensor analysis," *IEEE Trans. Image Process.*, vol. 17, no. 2, pp. 226–234, Feb. 2008.

[5] J. Huang, P. C. Yuen, W. Chen, and J. Lai, "Choosing parameters of kernel subspace LDA for recognition of face images under pose and illumination variations," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 4, pp. 847–862, Aug. 2007.

[6] Q. Ke and T. Kanade, "Robust L1 norm factorization in the presence of outliers and missing data by alternative convex programming," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, 2005, pp. 739–746.

[7] H. Kong, L. Wang, E. K. Teoh, J. Wang, and R. Venkateswarlu, "A framework of 2D Fisher discriminant analysis: Application to face recognition with small number of training samples," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, 2005, vol. 2, pp. 1083–1088.

[8] H. Kong, L. Wang, E. K. Teoh, J. Wang, and R. Venkateswarlu, "Generalized 2D principal component analysis for face image representation and recognition," *Neural Netw.*, vol. 18, no. 5/6, pp. 585–594, Jun. 2005.

[9] N. Kwak, "Principal component analysis based on L1-norm maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1672–1680, Sep. 2008.

[10] B. Leibe and B. Schiele, "Analyzing appearance and contour based methods for object categorization," in *Proc. Int. Conf. Comput. Vis. Pattern Recog.*, 2003, p. II-409-15.

[11] X. Li, S. Lin, S. Yan, and D. Xu, "Discriminant locally linear embedding with high-order tensor data," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 2, pp. 342–352, Apr. 2008.

[12] H. Lu, K. Plataniotis, and A. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 18–39, Jan. 2008.

[13] J. Lu, K. Plataniotis, A. Venetsanopoulos, and S. Li, "Ensemble-based discriminant learning with boosting for face recognition," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 166–178, Jan. 2006.

[14] Y. Pang, D. Tao, Y. Yuan, and X. Li, "Binary two-dimensional PCA," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 3, pp. 1176–1180, Aug. 2008.

[15] P. J. Phillips, H. Moon, S. A. Rivzi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.

[16] S. Shan, W. Zhang, Y. Su, X. Chen, and W. Gao, "Ensemble of piecewise FDA based on spatial histograms of local (Gabor) binary patterns for face recognition," in *Proc. Int. Conf. Pattern Recog.*, 2006, vol. 3, pp. 590–593.

[17] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.

[18] D. Tao, X. Li, X. Wu, W. Hu, and S. Maybank, "Supervised tensor learning," *Knowl. Inf. Syst.*, vol. 13, no. 1, pp. 1–42, Sep. 2007.

[19] D. Xu, S. Lin, S. Yan, and X. Tang, "Rank-one projections with adaptive margins for face recognition," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 5, pp. 1226–1236, Oct. 2007.

[20] L. Xu and A. L. Yuille, "Robust principal component analysis by self-organizing rules based on statistical physics approach," *IEEE Trans. Neural Netw.*, vol. 6, no. 1, pp. 131–143, Jan. 1995.

[21] J. Yang, D. Zhang, A. F. Frangi, and J. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.

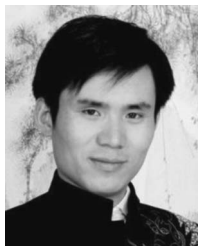
[22] J. Yang, D. Zhang, and J. Yang, "Constructing PCA baseline algorithms to reevaluate ICA-based face-recognition performance," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 4, pp. 1015–1021, Aug. 2007.

[23] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2004.

[24] J. Ye, "Generalized low rank approximation of matrices," *Mach. Learn.*, vol. 61, no. 1–3, pp. 167–191, Nov. 2005.

[25] Yale Face DB: Yale Univ.2007. [Online]. Available: cvc.yale.edu/projects/yalefaces/yalefaces.html

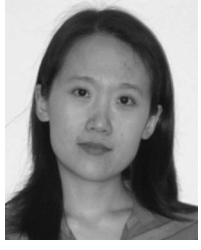
Xuelong Li (M'02–SM'07) is currently a Researcher (Full Processor) with the State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China.



Yanwei Pang (M'07–SM'09) received the Ph.D. degree in electronic engineering from the University of Science and Technology of China (USTC), Hefei, China.

He was a Postdoc at USTC. He is currently an Associate Professor with the School of Electronic Information Engineering, Tianjin University, Tianjin, China. His research areas include computer vision, Web search and mining, pattern recognition, and digital image processing, in which he has published about 50 scientific papers, including eight IEEE TRANSACTIONS papers.

Dr. Pang was a program committee member for more than ten international conferences. He was also a reviewer for more than ten international journals. He is a member of the editorial boards of the *International Journal of Computer Mathematics* (Taylor & Francis) and the *International Journal of Image and Graphics* (World Scientific).



Yuan Yuan (M'05–SM'09) received the B.Eng. degree from the University of Science and Technology of China, Hefei, China, and the Ph.D. degree from the University of Bath, Bath, U.K.

She is currently a Lecturer with the School of Engineering and Applied Science, Aston University, Birmingham, U.K. She has over 60 scientific publications in journals and conferences on visual information processing, compression, retrieval, etc.

Dr. Yuan is an Associate Editor of the *International Journal of Image and Graphics* (World Scientific) and an editorial board member of the *Journal of Multimedia* (Academy Publisher). She is also a reviewer for several IEEE TRANSACTIONS, other international journals, and conferences. She was a member of the program committees of many conferences.