Robust 2DPCA and Its Application

Qianqian Wang
Xidian University
Xi'an China
610887187@qq.com

Quanxue Gao
Xidian University
Xi'an China
xd_ste_pr@163.com

Abstract

Two-dimensional Principal Component Analysis (2DPCA) has been widely used for face image representation and recognition. However, 2DPCA, which is based on F-norm square, is sensitive to the presence of outliers. To enhance the robustness of 2DPCA model, we proposed a novel Robust 2DPCA objective function, called R-2DPCA. The criterion of R-2DPCA is maximizing the covariance of data in the projected subspace, while minimizing the reconstruction error of data. In addition, we use the efficient non-greedy optimization algorithms solving our objective function. Extensive experiments are done on the AR, CMU-PIE, Extended Yale B face image databases, and results illustrate that our method is more effective and robust than other robust 2DPCA algorithms, such as L1-2DPCA, L1-2DPCA-S, and N-2DPCA.

1. Introduction

In image recognition, especially face recognition, subspace learning, which is also called dimensionality reduction, has been widely used to extract features for image representation. Principal component analysis (PCA) [20], linear discriminant analysis (LDA) [1], locality preserving projection (LPP) [16] and neighborhood preserving embedding (NPE) [6] are four of the most representative methods, where PCA is used to extract the most expressive features, LDA is considered to be capable of extracting the most discriminating features. Different from PCA and LDA, which characterize the global geometric structure, LPP and NPE, which are respectively the linear approximation of Laplacian embedding (LE) [2] and locally linear embedding (LLE) [18] respectively, well characterize the local geometric structure.

However appling PCA into image representation and recognition, we need transform each image, which is usually represented as a matrix, into 1D image vector by concatenating all rows/columns. So, these methods cannot well exploit the spatial structure information that is embedded in pixels and their neighbors of image and important for image representation and recognition. To handle this problem, many two-dimensional subspace learning methods or tensor methods have been developed. The representative two-dimensional methods include two-dimensional PCA (2DPCA) [25], multi-linear PCA [11], two-dimensional LDA (2DLDA) [26]. Although the motivations of the aforementioned methods are different, they can be unified within the graph embedding framework [24] and measure the similarity between images using Euclidean distance square. It is commonly known that Euclidean distance square techniques are not robust in the sense the outlying measurements can arbitrarily skew the solution from the desired solution. Thus, these methods are not robust in the presence of outliers.

Recently, $\ell1$-norm based subspace learning technique has become an active topic in dimensionality reduction to improve the robustness to outliers. For example, Ke and Kanade [7] proposed L1-PCA which uses $\ell1$-norm to measure the reconstruction error in the objective function. Kwak [8] used $\ell1$-norm to measure the covariance and proposed PCA-L1 with greedy algorithm. Motivated by $\ell1$-norm based PCA, some $\ell1$-norm based subspace learning algorithms have been developed, such as LDA-L1, which replaces Euclidean distance square with $\ell1$-norm in the objective function of LDA [28]. Inspired by these works, researchers brought $\ell1$-norm based theory to two-dimensional methods, such as 2DPCA-L1 with greedy algorithm [10], L1-2DLDA [9]. and $\ell1$-norm based tensor subspace learning [17]. To achieve a better performance of $\ell1$-norm based method, Wang *et al*. [21] proposed 2DPCA-L1 with sparsity(2DPCAL1-S). It is a sparse variant of 2DPCA-L1 for unsupervised learning. However, the aforementioned methods utilize greedy optimization algorithm to solve $\ell1$-norm problems, which is computationally expensive. It consumes more time than 2DPCA due to that each principal vector is obtained via iteration operation. non-greedy algorithm was proposed to increase efficiency of optimization algorithm. For example, Nie *et al*. [15] proposed a non-greedy iterative to solve PCA-L1. Wang *et al*. [22] proposed 2DPCA-L1 with non-greedy algorithm. Although, all those methods are more robust to outliers compared with Euclidean distance square based methods, on the condition of ℓ1-norm, minimizing the reconstruction error is not equal to maximizing the covariance under the projected subspace, which means these methods do not consider the relationship between reconstruction error and low dimensional representation. Hence their performance is not improved a lot.

Moreover, ℓ1-norm is not invariant to rotations [3], which is a fundamental property of Euclidean space with ℓ2-norm. It has been emphasized in the context of learning algorithms [13]. Based on this content, Ding *et al*. [3] proposed the rotationally invariant ℓ1-norm for feature extraction and developed R1-PCA, which measures the similarity among nearby data by ℓ2,1-norm. Inspired by this, many joint ℓ2,1-norm regularization methods were proposed for robust feature selection in image processing, image annotation and machine learning community. For example, Nie *et al*. [14] proposed Joint ℓ2,1-norms robust feature selection. Wong *et al*. [23] proposed ℓ2,1-norm based tensor feature selection for image analysis. Gui *et al*. [5] enjoyed joint feature extraction and selection by exploiting ℓ2,1-norm on the projection matrix to subspace learning. These methods all use ℓ2,1-norm as a regularization rather than distance metric learning in the objective function and cannot guarantee the robustness of the objective function to outliers. Furthermore, they all need to transform 2D image into a vector by concatenating all rows of image. So, these methods cannot well exploit the spatial structure information of data that is important for image recognition [25, 27].

In this paper, to handle the aforementioned problems,we first propose a new Robust 2DPCA criterion, which explicitly considers the relationship between reconstruction error and covariance of data in the projected subspace, and we use the efficient non-greedy optimization algorithms solving our objective function. The main contributions of our approach are summarized as follows:

First, our proposed algorithm not only is robust to outliers but also has rotational invariance, which has been emphasized in the context of learning algorithms.

Second, our proposed non-greedy algorithm has a local optimal solution and can best maximize the objective function value.

Third, we explicitly consider the reconstruction error in our objective function, which is real goal of 2DPCA. Experimental results on real databases show that our method always performs more robust than other robust 2DPCA algorithms.

The remainder of this paper is organized as follows. Section 2 reviews 2DPCA and L1-2DPCA. Objective function and algorithm of our Robust 2DPCA are introduced in Section 3. Section 4 reports experimental results. Finally, the conclusions are drawn in Section 5.

2. 2DPCA and L1-2DPCA

Given a set of N sample images $\{\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_N\}$, where $\mathbf{A}_i \in \mathbf{R}^{m\times n}(i = 1, 2, ..., N)$ denote the i th training images. Without loss of generality, we assume the data set are centralized, i.e., $\sum_{i=1}^N \mathbf{A}_i = 0$. 2DPCA [25] aims to seek the projection matrix $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_d] \in \mathbf{R}^{n\times d}$ which can minimize the reconstruction error, i.e.,

$$\mathbf{V} = \underset{\mathbf{V}^T\mathbf{V}=\mathbf{I}_d}{arg\min} \sum_{i=1}^{N} \left\| \mathbf{A}_i - \mathbf{A}_i\mathbf{V}\mathbf{V}^T \right\|_F^2 \tag{1}$$

where $\|\cdot\|_F^2$ denotes the square Frobenius norm of matrix and is the sum of square ℓ2-norm of row/column vectors of matrix. It is easy to see that, the objective function (1) is totally equivalent to the following objective function, due to the fact $\sum_{i=1}^N \left\|\mathbf{A}_i - \mathbf{A}_i\mathbf{V}\mathbf{V}^T\right\|_F^2 + \sum_{i=1}^N \|\mathbf{A}_i\mathbf{V}\|_F^2 = \sum_{i=1}^N \|\mathbf{A}_i\|_F^2$

$$\mathbf{V} = \underset{\mathbf{V}^T\mathbf{V}=\mathbf{I}_d}{arg\max} \sum_{i=1}^{N} \|\mathbf{A}_i\mathbf{V}\|_F^2 \tag{2}$$

where $tr(\cdot)$ is the trace operator of a matrix. As $\sum_{i=1}^N \|\mathbf{A}_i\mathbf{V}\|_F^2 = tr\left(\sum_{i=1}^N \mathbf{V}^T\mathbf{A}_i^T\mathbf{A}_i\mathbf{V}\right)$, we let $\mathbf{S}_t = \sum_{i=1}^N \mathbf{A}_i^T\mathbf{A}_i$ denotes the covariance matrix. The solution of function (1) and (2) can be obtained by finding orthogonal eigenvectors of $\mathbf{S}_t$ corresponding to the first d largest eigenvalues.

Because 2DPCA characterizes the geometric structure of data by F-norm distance square, which is sensitive to noise and outliers. Thus the optimal projection matrix of the objective function (1) and (2) is not robust in the sense that outlying measurements can skew the solution from the desired solution. To address this problem, ℓ1-norm based 2DPCA was proposed. It aims to find projection matrix by solving the following objective function [10, 22].

$$\mathbf{V} = \underset{\mathbf{V}^T\mathbf{V}=\mathbf{I}_d}{arg\max} \sum_{i=1}^{N} \|\mathbf{A}_i\mathbf{V}\|_{L_1} \tag{3}$$

where $\|\cdot\|_{L_1}$ denotes the ℓ1-norm of a matrix, which is defined as $\|\mathbf{D}\|_{L_1} = \sum_{i=1}^m \sum_{j=1}^n |\mathbf{D}(i,j)|$. $\mathbf{D}(i,j)$ denotes the element of the i-th row j-th column of matrix $\mathbf{D}$. $\mathbf{A}_i(j,:)$ denotes the j-th row of $\mathbf{A}_i$.

ℓ1-norm based 2DPCA is more robust to outliers than 2DPCA [10, 22]. However, ℓ1-norm is not invariant to rotations, which is an important property of learning algorithms [3], i.e., given an arbitrary rotation matrix $\mathbf{\Gamma}$ ($\mathbf{\Gamma}\mathbf{\Gamma}^T = \mathbf{I}$), we have $\|\mathbf{\Gamma}\mathbf{A}_i(j,:)\mathbf{V}\|_1 \neq \|\mathbf{A}_i(j,:)\mathbf{V}\|_1$. Moreover, the objective function (3) does not explicitly consider the reconstruction error. ℓ1-norm based 2DPCA method is maximizing the covariance of data in the projected subspace, in terms of ℓ1-norm, minimizing the reconstruction error is not equivalent to maximizing the covariance of data in the projected subspace, i.e.$\sum_{i=1}^{N} \left\| \mathbf{A}_i - \mathbf{A}_i\mathbf{V}\mathbf{V}^T \right\|_{L_1} + \sum_{i=1}^{N} \left\| \mathbf{A}_i\mathbf{V} \right\|_{L_1} \neq \sum_{i=1}^{N} \left\| \mathbf{A}_i \right\|_{L_1}$.Thus, their robust performance is not improved a lot. To handle these problems, we propose a novel Robust 2DPCA model in section 3.

3. Robust 2DPCA

In this paper, different from existing robust 2DPCA methods, we propose a new objective function as below:

$$\mathbf{V}^* = \underset{\mathbf{V}^T\mathbf{V}=\mathbf{I}_d}{\arg\max} \sum_{i=1}^{N} tr \frac{\mathbf{V}^T{\mathbf{A}_i}^T\mathbf{A}_i\mathbf{V}}{\left\| \mathbf{A}_i - \mathbf{A}_i\mathbf{V}\mathbf{V}^T \right\|_F^2} \quad (4)$$

The objective function (4) is called Robust 2DPCA(R-2DPCA). We add reconstruction error as restraint in objective function, so it can further weaken the effect of outliers. Compared with L1-2DPCA, our method has rotational invariance due to $\left\| \mathbf{\Gamma}\mathbf{A}_i\mathbf{V} \right\|_F^2 = \left\| \mathbf{A}_i\mathbf{V} \right\|_F^2$. However, directly solving this problem is difficult, thus we use non-greedy iterative to solve the objective function. Now we consider how to solve the optimal projection matrix $\mathbf{V}$ of the objective function (4).

Algorithm 1: R-2DPCA

Input: $\mathbf{A}_i \in \mathbf{R}^{m\times n}$($i = 1, \cdots, N$), d , where $\mathbf{A}$ is centralized, $\gamma = 0.001$.

Initialize $\mathbf{V}^{(t)} \in \mathbf{R}^{n\times d}$ which satisfies $\mathbf{V}^T\mathbf{V} = \mathbf{I}$, $t = 1$.

while not converge **do**

1.For all training samples, calculate $\boldsymbol{d}^{(t)}$. According to Eq. (6), the i-th element ${\boldsymbol{d}_i}^{(t)} = \dfrac{1}{\left\| \mathbf{A}_i - \mathbf{A}_i\mathbf{V}\mathbf{V}^T \right\|_F^2 + \gamma}$.

2.Calculate $\mathbf{H}^{(t)}$ according to Eq. (5): $\mathbf{H}^{(t)} = \sum\limits_{i=1}^{N} {\mathbf{A}_i}^T{\boldsymbol{d}_i}^{(t)}\mathbf{A}_i\mathbf{V}$.

3.Calculate the singular value decomposition (SVD) of matrix $\mathbf{H}^{(t)}$ by $\mathbf{H} = \mathbf{U}\sum\mathbf{Z}^T$

4.Solve $\mathbf{V}^{(t+1)} = \underset{\mathbf{V}^T\mathbf{V}=\mathbf{I}_d}{\arg\max}\ tr((\mathbf{V}^{(t+1)})^T\mathbf{H}^{(t)})$ by Eq. (11). i.e. $\mathbf{V}^{(t+1)} = \mathbf{U}\left(\mathbf{I}_{n\times d}\right)\mathbf{Z}^T$.

5.Update $t \leftarrow t + 1$.

end while

Output:$\mathbf{V}^{(t+1)} \in \mathbf{R}^{n\times d}$

We simplify the function (4) as follows

$$\begin{aligned} &\sum_{i=1}^{N} tr \frac{\mathbf{V}^T{\mathbf{A}_i}^T\mathbf{A}_i\mathbf{V}}{\left\| \mathbf{A}_i - \mathbf{A}_i\mathbf{V}\mathbf{V}^T \right\|_F^2} \\ &= \sum_{i=1}^{N} tr(\mathbf{V}^T{\mathbf{A}_i}^T \frac{1}{\left\| \mathbf{A}_i - \mathbf{A}_i\mathbf{V}\mathbf{V}^T \right\|_F^2} \mathbf{A}_i\mathbf{V}) \\ &= tr(\mathbf{V}^T \sum_{i=1}^{N} ({\mathbf{A}_i}^T\boldsymbol{d}_i\mathbf{A}_i)\mathbf{V}) \\ &= tr(\mathbf{V}^T\mathbf{H}) \end{aligned} \quad (5)$$

where $\boldsymbol{d}_i = \dfrac{1}{\left\| \mathbf{A}_i - \mathbf{A}_i\mathbf{V}\mathbf{V}^T \right\|_F^2}$, and $\mathbf{H} = \sum\limits_{i=1}^{N} ({\mathbf{A}_i}^T\boldsymbol{d}_i\mathbf{A}_i)\mathbf{V}$. in order to avoid infinite large, i.e., ∞, d_i is defined as

$$\boldsymbol{d}_i = \frac{1}{\left\| \mathbf{A}_i - \mathbf{A}_i\mathbf{V}\mathbf{V}^T \right\|_F^2 + \gamma} \quad (6)$$

So the objective function (4) can be converted to solve the following objective function.

$$\mathbf{V}^* = \underset{\mathbf{V}^T\mathbf{V}=\mathbf{I}_d}{\arg\max}\ tr\left(\mathbf{V}^T\mathbf{H}\right) \quad (7)$$

Then, we consider how to solve the new objective function (7). Denote the singular value decomposition (SVD) of matrix $\mathbf{H}$ by

$$svd\left(\mathbf{H}\right) = \mathbf{U}\sum\mathbf{Z}^T \quad (8)$$

where $\mathbf{U}$ and $\mathbf{Z}$ are full rank and orthogonal matrices, i.e., $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}_n$, $\mathbf{Z}\mathbf{Z}^T = \mathbf{Z}^T\mathbf{Z} = \mathbf{I}_d$. $\sum$ is a diagonal matrix whose elements on diagonal are singular values λ_j of matrix $\mathbf{H}$.

Substituting Eq. (8) into Eq. (7), we have

$$\begin{aligned} tr\left(\mathbf{V}^T\mathbf{H}\right) &= tr\left(\mathbf{V}^T\mathbf{U}\sum\mathbf{Z}^T\right) \\ &= tr\left(\sum\mathbf{Z}^T\mathbf{V}^T\mathbf{U}\right) \\ &= tr\left(\sum\mathbf{M}\right) \\ &= \sum_{j=1}^{d} \lambda_j\mathbf{M}(j,j) \end{aligned} \quad (9)$$

where $\mathbf{M} = \mathbf{Z}^T\mathbf{V}^T\mathbf{U}$, $\mathbf{M}(j,j)$ denotes the element of the jth row jth column of $\mathbf{M}$.

Substituting Eq. (9) into Eq. (7), the objective function (7) finally becomes

$$\mathbf{V}^* = \underset{\mathbf{V}^T\mathbf{V}=\mathbf{I}_d}{\arg\max} \sum_{j=1}^{d} \lambda_j\mathbf{M}(j,j) \quad (10)$$

In Equation (10), $\mathbf{M}$ is a row orthogonal matrix due to the fact $\mathbf{M}\mathbf{M}^T = \mathbf{Z}^T\mathbf{V}^T\mathbf{U}\mathbf{U}^T\mathbf{V}\mathbf{Z} = \mathbf{I}$. So, $\mathbf{M}(j,j) \leq 1$. It means that $\sum_{j=1}^{d} \lambda_j\mathbf{M}(j,j) \leq \sum_{j=1}^{d} \lambda_j$, and the equality holds when $\mathbf{M}(j,j) = 1$. That is to say, when $\mathbf{M} = \mathbf{I} \in \mathbf{R}^{d\times n}$, the objective function (10) attains the maximum value. According to $\mathbf{M} = \mathbf{Z}^T\mathbf{V}^T\mathbf{U}$ and $\mathbf{M} = \mathbf{I} \in \mathbf{R}^{d\times n}$, the optimal projection matrix $\mathbf{V}$ can be calculated by

$$\mathbf{V}^* = \mathbf{U}\left(\mathbf{I}_{n\times d}\right)\mathbf{Z}^T \quad (11)$$

Note that both $\mathbf{U}$ and $\mathbf{Z}$ are related with $\mathbf{H}$, which depends on $\mathbf{Z}$ and is also an unknown vector. So, we cannot obtain closed-form solution for the objection function (7). We herein propose an iterative algorithm, which is described in *Algorithm* 1, to solve the optimal projection matrix $\mathbf{V}$ of the objective function (7), i.e., Robust 2DPCA.

To be specific, we first initialize matrix $\mathbf{V}$, which satisfies $\mathbf{V}^T\mathbf{V} = \mathbf{I}$, and calculate $\mathbf{H}$ by current $\mathbf{V}$ and $\boldsymbol{d}_i$. After that, matrix $\mathbf{V}$ is updated with the updated $\mathbf{H}$.

4. Experimental results

In this section, we validate our method in three face databases (Extended Yale B, CMU PIE and AR) and compare it with 2DPCA [25], L1-2DPCA [10], 2DPCAL1-S [21] and N-2DPCA [27]. In our experiments, we use 1-nearest neighbor (1NN) for classification. We set the number of projection vectors as 25 in Extended Yale B and CMU PIE databases, 30 in AR database.

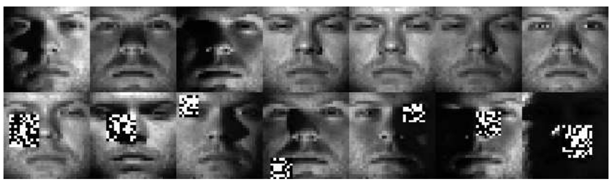

Figure 1: Some samples of one person in the Extended Yale B database. The second row is noised samples.

The Extended Yale B database [4] consists of 2144 frontal-face pictures of 38 individuals with different illuminations. There are 64 pictures for each person except 60 for 11th and 13th, 59 for 12th, 62 for 15th and 63 for 14th, 16th and 17th. In the experiments, each image was normalized to 32×32. 14 images of each individual were randomly selected and noised by black and white dotes with random distribution. The location of noise is random and ratio of the pixels of noise to number of image pixels is intervenient 0.05 to 0.15. Figure 1 shows some samples of one person in the Extend Yale B database. We randomly select 32 images, which include 7 noisy images, per person for training, and the remaining images for testing. 2DPCA, L1-2DPCA, 2DPCAL1-S, N-2DPCA and our method are used to extract features, respectively. We repeat this process 10 times.

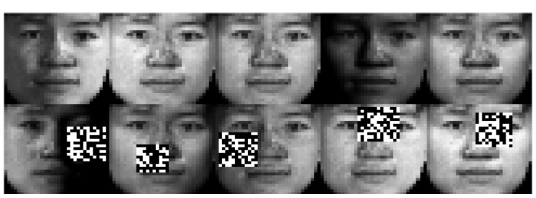

Figure 2: Some samples of one person in the CMU PIE database. The second row is noised samples.

The CMU PIE database [19] consists of 2856 frontal-face images of 68 individuals with different illuminations. In the experiments, each image was normalized to 32×32 pixels, we randomly selected 10 images and added the same noise as that in the CMU PIE database. Figures 2 shows some images of one person in the CMU PIE database. We randomly select 21 images, which include 16 images without noise, per person for training and the remaining images for testing. 2DPCA, L1-2DPCA , 2DPCAL1-S and N-2DPCA and our method are used to extract features, respectively. We repeat this process 10 times.

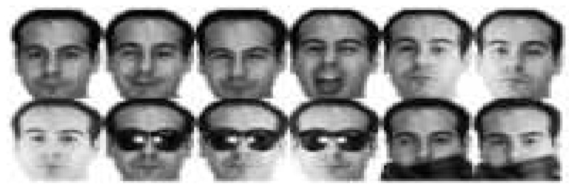

Figure 3: Some samples of one person in the AR database.

The AR database [12] contains over 4000 color face image of 126 people, including frontal views of faces with different facial expressions, lighting conditions and occlusions such as glasses and scarf. The pictures of 120 individuals were taken in two sessions. Each section contains 13 color images, which include 6 images with occlusions and 7 full facial images with different facial expressions and lighting conditions. We manually cropped the face portion of the image and then normalized it to 50×40 pixels. Figure 3 shows sample images of one person in the AR database. In the experiments, all of the pictures with occlusions are considered as noisy samples. For each person, we randomly select 13 images per person for training and the remaining images for testing, and then repeat this process 10 times.

Table 1: The optimal average classification accuracy (%) and the corresponding standard deviation of each method on three databases.

Methods	Experiments		
	Extended Yale B	AR	CMU PIE
2DPCA	59.92±0.42	80.40±0.88	85.39±0.73
L1-2DPCA	60.33±0.38	80.39±0.88	85.71±0.77
2DPCAL1-S	60.37±0.54	80.38±0.85	85.91±0.69
N-2DPCA	59.99±0.57	80.37±0.91	85.39±0.73
Ours	**62.85±0.75**	**89.17±0.73**	**90.03±0.95**

Table 1 lists the average recognition accuracy of each method and standard deviation (Std) on the Extended Yale B, AR, and CMU PIE databases. Table 2 lists the average running time of each method and standard deviation (Std) on the Extended Yale B, AR, and CMU PIE databases. Figures 4-6 show the top classification accuracy of each method under 10 experiments on three databases. Figures 7-9 plot the classification accuracy curve versus the number projection vectors on three databases, respectively. Figure 10 presents the convergence curve of our method on three databases.

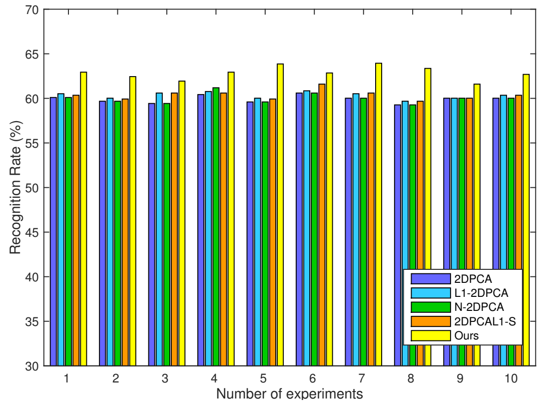

Figure 4: Average classification accuracy of five approaches under ten experiments on the Extended Yale B database.

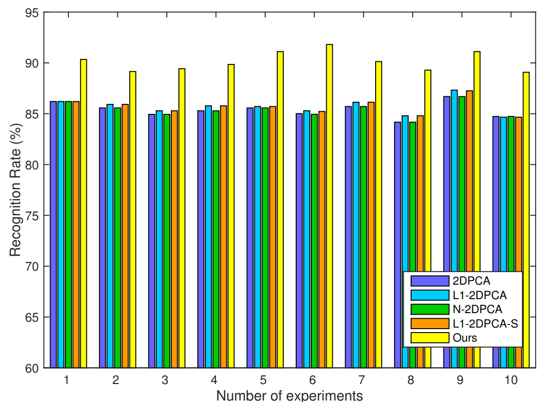

Figure 5: Average classification accuracy of five approaches under ten experiments on the CMU PIE database.

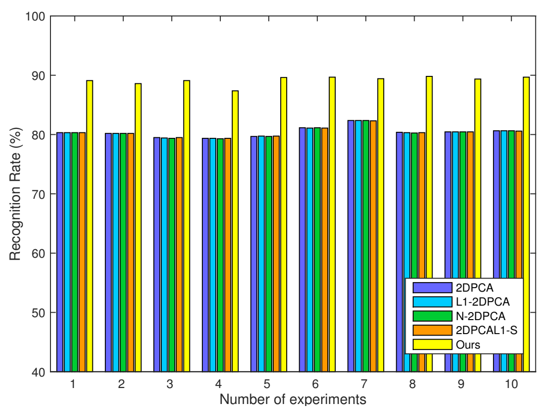

Figure 6: Average classification accuracy of five approaches under ten experiments on the AR database.

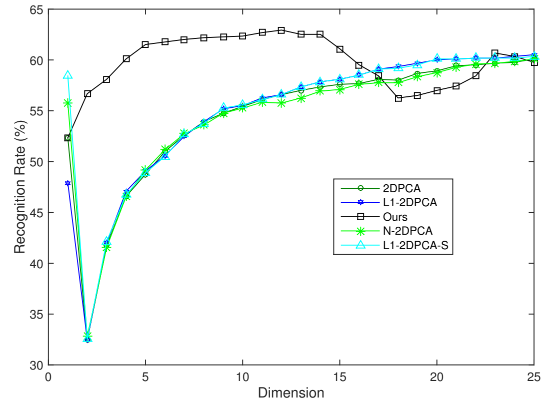

Figure 7: Average classification accuracy versus the number of projection vectors on the Extended Yale B database.

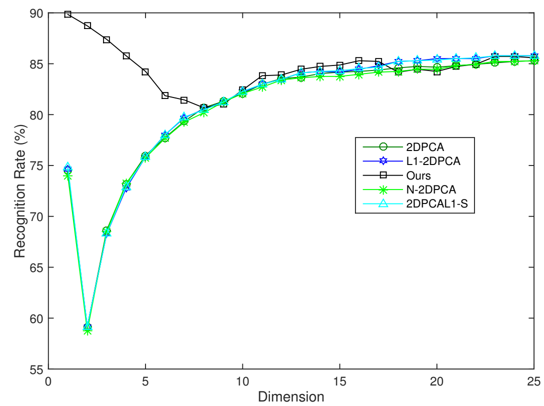

Figure 8: Average classification accuracy versus the number of projection vectors on the CMU PIE database.

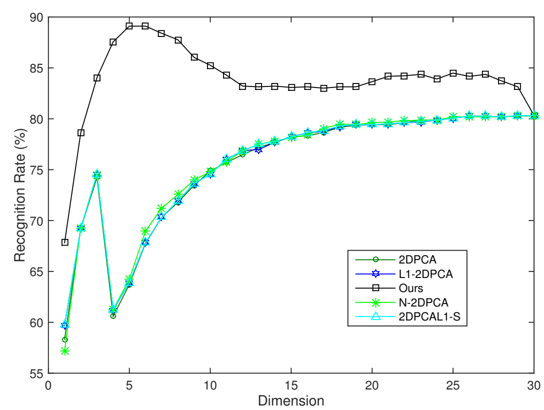

Figure 9: Average classification accuracy versus the number of projection vectors on the AR database.

Comparing the aforementioned experimental results, we have several interesting observations as follows: (1)2DPCA is inferior to the other three approaches. This is probably because that 2DPCA is sensitive to the small variation due to the illumination, pose and occlusion. It results in unstable representation for images. While L1-2DPCA, 2DPCAL1-S, N-2DPCA and our method are robust to these variations in solving the optimal projection matrix. (2)Our method is superior to the other three approaches when using Euclidean distance metrics in classification phase. This is probably because that our approach consider the relationship between reconstruction error and covariance. Another reason may be that $\ell 1$-norm based 2DPCA uses greedy strategy to solve the projection matrix, which cannot best optimize the objective function. (3)Table 3 and Figure 10 shows that our proposed algorithm is fast and convergent.

Table 2: The average running time and the corresponding standard deviation of each method on three databases.

Methods	Experiments		
	Extended Yale B	AR	CMU PIE
2DPCA	0.01±0.00	0.04±0.00	0.01±0.00
L1-2DPCA	7.07±0.73	11.30±1.37	9.20±1.227
2DPCAL1-S	6.66±0.48	11.73±0.79	8.33±0.71
N-2DPCA	12.69±1.38	24.36±5.16	15.41±3.15
Ours	**2.72±0.39**	**5.81±1.43**	**2.63±0.42**

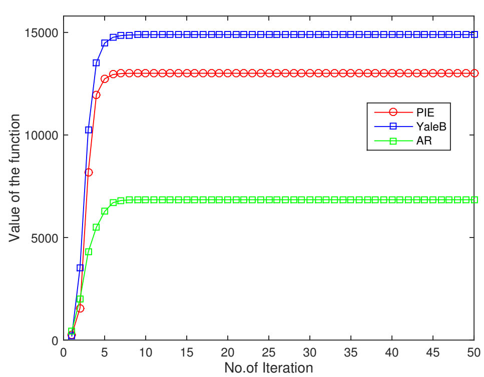

Figure 10: Convergence curve of our method on three databases.

5. Conclusions

This paper presents a robust dimensionality reduction and feature extraction method in image domain, namely R-2DPCA. Our proposed model is robust to outliers because it explicitly considers the relationship between reconstruction error and covariance, and weakens the effect of large distance. Moreover it has rotational invariance. We provide a fast and convergent algorithm to solve our method. Experimental results on the several face image databases show the effectiveness and advantages of the proposed method.

References

[1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[2] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

[3] C. Ding, D. Zhou, X. He, and H. Zha. R 1-pca: rotational invariant l 1-norm principal component analysis for robust subspace factorization. In *International Conference on Machine learning*, pages 281–288, Pittsburgh, Pennsylvania, 2006. ACM.

[4] 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 Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[5] J. Gui, D. Tao, Z. Sun, Y. Luo, X. You, and Y. Y. Tang. Group sparse multiview patch alignment framework with view consistency for image classification. *IEEE Transactions on Image Processing*, 23(7):3126–3137, 2014.

[6] X. He, D. Cai, S. Yan, and H.-J. Zhang. Neighborhood preserving embedding. In *IEEE International Conference on Computer Vision*, volume 2, pages 1208–1213, Beijing, China, 2005. IEEE.

[7] Q. Ke and T. Kanade. Robust l1 norm factorization in the presence of outliers and missing data by alternative convex programming. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 739–746, San Diego, CA, 2005. IEEE.

[8] N. Kwak. Principal component analysis based on l1-norm maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(9):1672–1680, 2008.

[9] C.-N. Li, Y.-H. Shao, and N.-Y. Deng. Robust l1-norm two-dimensional linear discriminant analysis. *Neural Networks*, 65:92–104, 2015.

[10] X. Li, Y. Pang, and Y. Yuan. L1-norm-based 2dpca. *IEEE Transactions on systems, man, and cybernetics. Part B, Cybernetics*, 40(4):1170–1175, 2010.

[11] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. Mpca: Multilinear principal component analysis of tensor objects. *IEEE Transactions on Neural Networks*, 19(1):18–39, 2008.

[12] A. M. Martinez. The ar face database. *CVC Technical Report*, 24, 1998.

[13] A. Y. Ng. Feature selection, l 1 vs. l 2 regularization, and rotational invariance. In *International Conference on Machine Learning*, page 78, Alberta, Canada, 2004. ACM.

[14] F. Nie, H. Huang, X. Cai, and C. H. Ding. Efficient and robust feature selection via joint $\ell 2$, 1-norms minimization. In *Advances in neural information processing systems*, pages 1813–1821, Vancouver, Canada, 2010.

[15] F. Nie, H. Huang, C. Ding, D. Luo, and H. Wang. Robust principal component analysis with non-greedy l1-norm maximization. In *International Joint Conference on Artificial Intelligence*, volume 22, page 1433, Barcelona, 2011. Citeseer.

[16] X. Niyogi. Locality preserving projections. In *Neural information processing systems*, volume 16, page 153, Vancouver, Canada, 2004.

[17] Y. Pang, X. Li, and Y. Yuan. Robust tensor analysis with l1-norm. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(2):172–178, 2010.

[18] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[19] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 46–51, Washington, DC, 2002. IEEE.

[20] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991.

[21] H. Wang and J. Wang. 2dpca with l1-norm for simultaneously robust and sparse modelling. *Neural Networks*, 46:190–198, 2013.

[22] R. Wang, F. Nie, X. Yang, F. Gao, and M. Yao. Robust 2dpca with non-greedy-norm maximization for image analysis. *IEEE Transactions on Cybernetics*, 45(5):1108–1112, 2015.

[23] W. K. Wong, Z. Lai, Y. Xu, J. Wen, and C. P. Ho. Joint tensor feature analysis for visual object recognition. *IEEE Transactions on Cybernetics*, 45(11):2425–2436, 2015.

[24] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin. Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):40–51, 2007.

[25] J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang. Two-dimensional pca: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, 26(1):131–137, 2004.

[26] J. Yang, D. Zhang, X. Yong, and J.-y. Yang. Two-dimensional discriminant transform for face recognition. *Pattern recognition*, 38(7):1125–1129, 2005.

[27] F. Zhang, J. Yang, J. Qian, and Y. Xu. Nuclear norm-based 2-dpca for extracting features from images. *IEEE Transactions on Neural Networks and Learning Systems*, 26(10):2247–2260, 2015.

[28] F. Zhong and J. Zhang. Linear discriminant analysis based on l1-norm maximization. *IEEE Transactions on Image Processing*, 22(8):3018–3027, 2013.