

Research on Face Recognition Algorithm Based on Robust 2DPCA

Haijun Kuang, Wanzhou Ye, Ze Zhu

Department of Mathematics, College of Science, Shanghai University, Shanghai, China

Email: sky1417168@163.com, wzhy@shu.edu.cn

How to cite this paper: Kuang, H.J., Ye, W.Z. and Zhu, Z. (2021) Research on Face Recognition Algorithm Based on Robust 2DPCA. *Advances in Pure Mathematics*, 11, 149-161.

<https://doi.org/10.4236/apm.2021.112010>

Received: January 13, 2021

Accepted: February 19, 2021

Published: February 22, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution-NonCommercial International License (CC BY-NC 4.0).

<http://creativecommons.org/licenses/by-nc/4.0/>



Open Access

Abstract

As a new dimension reduction method, the two-dimensional principal component (2DPCA) can be well applied in face recognition, but it is susceptible to outliers. Therefore, this paper proposes a new 2DPCA algorithm based on angel-2DPCA. To reduce the reconstruction error and maximize the variance simultaneously, we choose F norm as the measure and propose the Fp-2DPCA algorithm. Considering that the image has two dimensions, we offer the Fp-2DPCA algorithm based on bilateral. Experiments show that, compared with other algorithms, the Fp-2DPCA algorithm has a better dimensionality reduction effect and better robustness to outliers.

Keywords

2DPCA, Face Recognition, Dimension Reduction, F Norm

1. Introduction

Principal component analysis (PCA) [1] is extensively used in dimension reduction, pattern recognition, and computer vision; however, when a principal component analysis is applied to face image representation and recognition. It is necessary to transform each image in the form of a matrix into a one-dimensional vector column by column or row by row, which can't make full use of the spatial structure information of image pixels and their neighborhood. The concatenation of a two-dimensional matrix into a one-dimensional vector often leads to a too high vector space dimension. Because the covariance matrix size is too large and training samples are relatively small, it is difficult to calculate the covariance matrix accurately. To solve this problem, Yang *et al.* [2] proposed a two-dimensional principal component analysis method based on two-dimensional images rather than a one-dimensional vector. It has successfully been applied in computer vision and signal processing [3] [4].

PCA only works in the image's row direction for dimensionality reduction, and the compression ratio is meager. To solve this problem, some methods based on bilateral compression are proposed. Kong *et al.* [5] presented 2DPCA, which constructs two subspaces to encode the row vector and the column vector of the image matrix, respectively. Zhang *et al.* [6] proposed a (2D) 2DPCA with both row and column directions. Xu *et al.* [7] constructed two projection transformation matrices by defining two image covariance matrices. Kim *et al.* [8] proposed calculating the linear transformation matrix using two covariance matrices in 2DPCA.

PCA and 2DPCA are both based on the sum of the least square F-norm, equivalent to the least square loss or the square L2 norm. As we all know, the least direct loss is not robust because the edge data points can easily make the solution deviate from the expected answer. Compared with square F-norm, the L1 norm is more vital to outliers. Kwak [9] used the L1 model to measure variance and developed PCA-l1. Then, to better use the spatial structure embedded in the image, 2DPCA-l1 is proposed [10]. One disadvantage of the L1 norm is that it has no rotation invariance, emphasizing learning algorithms [11]. Rotation invariance helps avoid performance degradation. On this basis, due to the rotation invariance of F-norm, some methods based on F-norm are proposed. For the F-norm-based form, minimizing the reconstruction error is not equal to maximizing the variance. Therefore, the above way does not consider the relationship between reconstruction error and disagreement.

In recent years, to overcome this defect, Gao *et al.* [12] proposed an angel-2DPCA, which takes F-norm as the distance measure and obtains the optimal projection matrix by minimizing the ratio of reconstruction error variance. Angel-2DPCA [12] [13] is not only robust to outliers but also rotation invariant. The experimental results of reference show that angel-2DPCA has high efficiency. However, this method only reduces the column direction dimension and limits the degree of F norm to 1. In this case, we extend angel-2DPCA to Fp-2DPCA. We can prove that the new algorithm can converge to the optimal local solution. Besides, we propose a bilateral Fp-2DPCA to solve the problem of bilateral dimensionality reduction.

The rest of this paper is organized as follows. The second section introduces the theory of 2DPCA, including 2DPCA and 2DPCA-L1. The third section proposes the Fp-2DPCA algorithm and the bilateral Fp-2DPCA algorithm and proves the algorithm's convergence. We also discuss the rotation invariance of the algorithm. In Section 4, we perform numerical experiments to evaluate the performance of our algorithm. Finally, the conclusion of this paper is given in the fifth section.

2. Related Theories of 2DPCA

2.1. 2DPCA

2DPCA is the result of expanding to two-dimensional space based on PCA. Still,

its basic idea is the same as PCA: to maximize the variance sum of the original data projected to the main components to maintaining the maximum amount of information of the original data as far as possible [14]. For image data, let $A = \{A_i \in \mathbb{R}^{m \times n}, i = 1, 2, \dots, M\}$ be the training data set, where M is the number of training samples, m and n are the dimensions of a row and column pixels, respectively, and the feature matrix $W = [w_1, w_2, \dots, w_d] \in \mathbb{R}^{n \times d}$, where d is the number of main projection vectors after transformation, then the objective function of 2DPCA is expressed as follows:

$$\max_{W^T W = I_k} \text{tr} \left(\sum_{i=1}^M W^T (A_i)^T A_i W \right) = \max_{W^T W = I_k} \sum_{i=1}^M \|A_i W\|_F^2 \quad (1)$$

$\text{tr}(\bullet)$ is the trace of the matrix; if there is a n dimensional matrix A , then the trace of the matrix A is equal to the sum of the eigenvalues of A , that is, the sum of the main diagonal elements of the matrix A . Since the F-norm is used in (1), the following equation is satisfied:

$$\sum_{i=1}^M \|E_i\|_F^2 + \sum_{i=1}^M \|A_i W\|_F^2 = \sum_{i=1}^M \|A_i\|_F^2 \quad (2)$$

So (1) is equivalent to:

$$\min_{W^T W = I} \sum_{i=1}^M \|E_i\|_F^2 \quad (3)$$

where $I_k \in \mathbb{R}^{d \times d}$ is a k dimensional identity matrix, $\|\cdot\|_F$ is the F norm of the matrix, $E_i = A_i - A_i W W^T$. Objective functions Equation (1) and Equation (3) show that 2DPCA mainly considers the reconstruction error or variance contribution of image data.

As shown in the objective function Equation (1), the square F norm is used as the distance measure. Still, the square F norm is not robust because the edge's observation value will quickly make the solution deviate from the expected answer. There is 2DPCA based on L_1 norm to solve this problem, which can reduce this influence to a certain extent.

2.2. 2DPCA-L1

The objective function of 2DPCA-L1 is as follows:

$$\max_{W^T W = I} \sum_{i=1}^M \|A_i W\|_{L_1}^2 \quad (4)$$

$\|\cdot\|_{L_1}$ is the L_1 norm of the matrix. Compared with the traditional 2DPCA, 2DPCA-L1 is more robust to the data with outliers, but it also has some defects. Firstly, it does not satisfy the rotation invariance:

$$\sum_{i=1}^M \|E_i\|_{L_1} + \sum_{i=1}^M \|A_i W\|_{L_1} \neq \sum_{i=1}^M \|A_i\|_{L_1} \quad (5)$$

Obviously, the solution of Equation (4) is not the solution of Equation (6); that is to say, it does not consider the reconstruction error. As a result, their robustness has not been greatly improved. It still thinks that every image has the

same contribution. The outliers or noises make the samples have sparse distribution, which also affects the robustness of the model. Secondly, it is very difficult to solve the objective function Equation (6). In order to solve these problems, a new robust 2DPCA model is proposed in the third part.

$$\min_{W^T W=I} \sum_{i=1}^M \|E_i\|_{L_1}^2 \tag{6}$$

3. Fp-2DPCA

3.1. Objective Function

It can be seen from the above analysis that the square F norm exaggerates the role of some data (mainly noise) in solving the 2DPCA model. This reduces the robustness of 2DPCA to noise. Therefore, to overcome the limitations of the above methods, we should adopt an appropriate distance measure, which can reduce the influence of outliers in the objective function and characterize the geometric structure of the objective function. F norm and square f norms have the same position in characterizing data dispersion and geometric design in the normative sense. The main difference between them is that, compared with the square F norm, the f norm can make the influence difference of different data smaller. Therefore, if F -norm is selected as the distance measure in 2DPCA, it will have the following two advantages:

- 1) It can capture the geometry structure well and has rotation invariance.
- 2) It can reduce the role of outliers in solving the optimal projection direction.
- 3) It helps to enhance the part of some adjacent data points with different labels.

Since the relationship between variance and reconstruction error is nonlinear, the maximum variation does not guarantee the minimum reconstruction error. According to the above analysis, we propose a new dimension reduction method, namely Fp-2DPCA. Fp-2DPCA uses F -norm to represent the low dimensional representation and reconstruction errors and integrates them into the criterion function. Specifically, our goal is to find the projection direction, minimize the angle between the projection directions, and reconstruct each data's error. The objective function of Fp-2DPCA is as follows.

$$\min_{W^T W=I_k} \sum_{i=1}^M \frac{\|E_i\|_F^p}{\|A_i W\|_F^p} \quad (0 < p < 2) \tag{7}$$

3.2. Algorithm

3.2.1. Unilateral Fp-2DPCA

Through the simple algebraic operation, we can get the following results:

$$\begin{aligned} \sum_{i=1}^M \frac{\|E_i\|_F^p}{\|A_i W\|_F^p} &= \sum_{i=1}^M \frac{\|E_i\|_F^2 \|E_i\|_F^{p-2}}{\|A_i W\|_F^p} = \sum_{i=1}^M \|E_i\|_F^2 * d_i = \sum_{i=1}^M \|A_i - A_i W W^T\|_F^2 * d_i \\ &= \sum_{i=1}^M \text{tr}(A_i^T A_i - W^T A_i^T A_i W) * d_i = \sum_{i=1}^M \|\text{tr}(G) - \text{tr}(W^T G W)\| \end{aligned} \tag{8}$$

Among them,

$$G = \sum_{i=1}^M \text{tr}(A_i^T A_i d_i), \quad d_i = \frac{\|E_i\|_F^{p-2}}{\|A_i W\|_F^p} = \frac{\|A_i - A_i W W^T\|_F^{p-2}}{\|A_i W\|_F^p} \tag{9}$$

According to Equation (8) and Equation (9), the objective function Equation (7) becomes:

$$\text{tr}(G) - \text{tr}(W^T G W) \tag{10}$$

Several theorems are introduced before solving the objective function Equation (10).

Lemma 1 Cauchy Schwarz inequality: for all sequences of real numbers a_i and b_i , we have

$$\left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right) \geq \left(\sum_{i=1}^n a_i b_i\right)^2 \tag{11}$$

Equality holds if and only if $a_i = k b_i$ for a non-zero constant $k \in \mathbb{R}$.

Theorem 1 For matrices P, Q of the same order as any two, we can get:

$$\text{tr}(P^T Q) \leq \|P\|_F \|Q\|_F \tag{12}$$

If and only if $P = lQ$, the equal sign holds, and l is any real number.

Proof of Theorem 1:

According to the definition of matrix trace, we can get the following results

$$\text{tr}(P^T Q) = (\text{vec}(P))^T \text{vec}(Q) \tag{13}$$

According to Cauchy Schwarz inequality (Equation (11)), there is a

$$\begin{aligned} (\text{vec}(P))^T \text{vec}(Q) &\leq \|\text{vec}(P)\|_2 \|\text{vec}(Q)\|_2 \\ &= \|\text{vec}(P)\|_F \|\text{vec}(Q)\|_F \end{aligned} \tag{14}$$

$\text{vec}(\cdot)$ is the vectorization of the matrix; that is, let $A_{m \times n} = (a_1, a_2, \dots, a_n)$ define the vector (Equation (15)) of $mn \times 1$, which is the vector that arranges the matrix A in column vectors.

$$\text{vec}(A) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \tag{15}$$

Therefore, according to Equation (13) and Equation (14), it can be obtained that:

$$\text{tr}(P^T Q) \leq \|P\|_F \|Q\|_F \tag{16}$$

And if and only if $P = lQ$ (l is any real number), the equal sign holds.

Theorem 2 Let the SVD of $H \in R^{m \times n}$ be decomposed into $H = U \Sigma V^T$, where $U^T U = V^T V = I_k$, $\Sigma \in R^{k \times k}$ is a nonsingular diagonal matrix, and its diagonal element $\lambda_j (j = 1, \dots, k)$ is the singular value of H , $k = \text{rank}(H)$. Then $W = UV^T$ is the solution of:

$$\max_{W^T W = I_k} \text{tr}(W^T H) \tag{17}$$

The proof of theorem 2:

According to the SVD decomposition of H , we can get:

$$\begin{aligned} \text{tr}(W^T H) &= \text{tr}(W^T U \Sigma V^T) \\ &= \text{tr}(U \Sigma^{1/2} \Sigma^{1/2} V^T W^T) \end{aligned} \tag{18}$$

According to theorem 1, we can get:

$$\begin{aligned} \text{tr}(W^T H) &\leq \|U \Sigma^{1/2}\|_F \| \Sigma^{1/2} V^T W^T \|_F^T \\ &= \| \Sigma^{1/2} \|_F \| \Sigma^{1/2} \|_F \end{aligned} \tag{19}$$

The equation holds if and only if:

$$\Sigma^{1/2} U^T = \Sigma^{1/2} V^T W^T \tag{20}$$

holds, so the solution is:

$$W = UV^T \tag{21}$$

We consider how to solve the objective function Equation (10), where there are unknown variables d_i related to V . Therefore, it has no closed-form solution. We can develop an algorithm to alternately update V (fixed d_i) and d_i (fixed V). Specifically, we have two steps to solve the objective function Equation (10). First, update V while revising d_i . In this case, the objective function (10) is constant. Therefore, the objective function Equation (10) becomes:

$$\max_{W^T W = I_k} \text{tr}(W^T H) \tag{22}$$

where $H = GW$ and G are the weighted covariance matrix of the image data.

Let SVD of H be decomposed into $H = U \Sigma V^T$, where $H = U \Sigma V^T$ and $\Sigma \in R^{k \times k}$ are nonsingular diagonal matrices, and $\lambda_j (j = 1, \dots, k)$ is the singular value of H . According to theorem 2, the optimal solution of the objective function Equation (22) is as follows:

$$W = UV^T \tag{23}$$

Secondly, d_i is calculated with the updated V . Algorithm CC lists the pseudo-code to solve the objective function (10) namely Fp-2DPCA algorithm (Table 1).

3.2.2. Bilateral Fp-2DPCA

As discussed in Section 3, compared with the traditional 2DPCA, angel-2DPCA is more robust to outliers. However, angel-2DPCA only reduces dimensions in the row direction of the image. In other words, angel-2DPCA learns the projection matrix through a set of training images, which only reflects the information between image rows and does not consider the information embedded in image columns. Therefore, it needs more dimensions to represent images, so it needs more storage space to store large-scale data sets. Inspired by reference [15], we use projection matrix R on the right and L on the left to overcome the shortcomings of angel-2DPCA.

Table 1. Algorithm 1 Fp-2DPCA.

Algorithm 1 Fp-2DPCA

Input: $A_i \in R^{m \times n}$ ($i = 1, \dots, N$), $k, \sum_{i=1}^M A_i = 0, J(W^{(t)}) = \sum_{i=1}^M \frac{\|E_i\|_F^p}{\|A_i W^{(t)}\|_F^p}$

Initialize: $W^{(t)} \in R^{m \times k}, (W^{(t)})^T W^{(t)} = I_k, t = 1, \delta = |J(W^{(t)}) - J(W^{(t-1)})|$

While: $\delta \geq \varepsilon$

1. Calculate $d_i^{(t)}$ for each A_i according to Equation (9),
Calculate $W: W = UV^T$ according to the conclusion of Equation (23),
2. Calculate $H^{(t)}$ according to Equation (9) and Equation (23),
3. SVD decomposition of $H^{(t)}: H^{(t)} = U\Sigma V^T$,
4. Calculate $W: W = UV^T$ according to the conclusion of Equation (23),
5. Update δ ,
6. Update $t: t = t + 1$.

Output: $W^{t+1} \in R^{m \times k}$

Specifically, firstly, the dimension of training sample $A_i \in R^{m \times n}$ ($i = 1, 2, \dots, N$) is reduced to get the right projection matrix $R \in R^{n \times r}$. The image A_i is projected onto R to reach $Y_i = A_i R \in R^{m \times r}$, called the suitable feature matrix of A_i . Then, we use Fp-2DPCA to reduce the training sample $Y_i^T \in R^{r \times m}$ ($i = 1, 2, \dots, N$) dimension and map it to the feature matrix $B_i^T = Y_i^T L$ with size $r \times l$. Through the above two processes, the training sample A_i is projected to a smaller feature matrix B_i :

$$B_i = L^T A_i R, i = 1, 2, \dots, N \tag{24}$$

We call R right projection matrix, L left the projection matrix, and the corresponding algorithm is called bilateral Fp-2DPCA. Because $l \times r$ is much smaller than $m \times n$, bilateral Fp-2DPCA can use fewer dimensions to represent the input image. Experimental results show that, compared with 2DPCA, (2D) 2DPCA, and angle-2DPCA, bilateral Fp-2DPCA can achieve higher performance with fewer dimensions. Algorithm 2 (Table 2) gives the specific algorithm steps of bilateral Fp-2DPCA.

3.3. Theoretical Analysis

3.3.1. Convergence Analysis

Theorem 3 in the iteration of algorithm 1, we can get:

$$\sum_{i=1}^M \|A_i W^{(t+1)}\|_F \geq \sum_{i=1}^M \|A_i W^{(t)}\|_F \tag{25}$$

Proof: in the $t + 1$ iteration, according to the fourth step of algorithm 1, the following inequality can be obtained:

$$\sum_{i=1}^M \frac{tr\left(\left(W^{(t+1)}\right)^T A_i^T A_i W^{(t+1)}\right)}{\|A_i W^{(t+1)}\|_F} \geq \sum_{i=1}^M \frac{tr\left(\left(W^{(t)}\right)^T A_i^T A_i W^{(t)}\right)}{\|A_i W^{(t)}\|_F} \tag{26}$$

Table 2. Algorithm 2 Bilateral Fp-2DPCA.

Algorithm 2 Bilateral Fp-2DPCA

Input: $A_i \in \mathbb{R}^{m \times n}$ ($i = 1, \dots, N$), $\sum_{i=1}^M A_i = 0$

1. Taking A_i ($i = 1, \dots, N$) as the input of algorithm 1, the right projection matrix $R \in \mathbb{R}^{n \times r}$ is obtained, and the characteristic matrix $Y_i = A_i R$ of A_i is calculated;
2. Taking Y_i^T ($i = 1, \dots, N$) as the input of algorithm 1, the left projection matrix $L \in \mathbb{R}^{m \times d}$ is obtained, and the characteristic matrix $B_i^T = Y_i^T L$ of Y_i^T is calculated;

Output: Output: output left projection matrix $L \in \mathbb{R}^{m \times d}$, right projection matrix $R \in \mathbb{R}^{n \times r}$, and characteristic matrix $B_i = L^T A_i R \in \mathbb{R}^{d \times r}$, $i = 1, 2, \dots, N$.

Then, we get:

$$\sum_{i=1}^M \frac{\text{tr}\left(\left(W^{(t+1)}\right)^T A_i^T A_i W^{(t)}\right)}{\left\|A_i W^{(t)}\right\|_F} \geq \sum_{i=1}^M \left\|A_i W^{(t)}\right\|_F \tag{27}$$

For each i ($i = 1, \dots, M$), we can get:

$$\begin{aligned} \text{tr}\left(\left(W^{(t+1)}\right)^T A_i^T A_i W^{(t)}\right) &= \text{tr}\left(\left(A_i W^{(t+1)}\right)^T A_i W^{(t)}\right) \\ &= \left(\text{vec}\left(A_i W^{(t+1)}\right)\right)^T \text{vec}\left(A_i W^{(t)}\right) \end{aligned} \tag{28}$$

According to Cauchy Schwarz inequality, we can get:

$$\begin{aligned} \left(\text{vec}\left(A_i W^{(t+1)}\right)\right)^T \text{vec}\left(A_i W^{(t)}\right) &\leq \left\|\text{vec}\left(A_i W^{(t+1)}\right)\right\|_2 \left\|\text{vec}\left(A_i W^{(t)}\right)\right\|_2 \\ &= \left\|A_i W^{(t+1)}\right\|_F \left\|A_i W^{(t)}\right\|_F \end{aligned} \tag{29}$$

According to Equation (28) and Equation (29), we can get:

$$\sum_{i=1}^M \frac{\text{tr}\left(\left(W^{(t+1)}\right)^T A_i^T A_i W^{(t)}\right)}{\left\|A_i W^{(t)}\right\|_F} \leq \frac{\left\|A_i W^{(t+1)}\right\|_F \left\|A_i W^{(t)}\right\|_F}{\left\|A_i W^{(t)}\right\|_F} \tag{30}$$

According to Equation (27) and Equation (30), we can get:

$$\sum_{i=1}^M \left\|A_i W^{(t+1)}\right\|_F \geq \sum_{i=1}^M \left\|A_i W^{(t)}\right\|_F \tag{31}$$

Theorem 4 in the iteration of algorithm 1, we can get:

$$\sum_{i=1}^M \left\|A_i - A_i W^{(t+1)} \left(W^{(t+1)}\right)^T\right\|_F \leq \sum_{i=1}^M \left\|A_i - A_i W^{(t)} \left(W^{(t)}\right)^T\right\|_F \tag{32}$$

Prove: after calculation, get:

$$\begin{aligned} \sum_{i=1}^M \left\|A_i - A_i W^{(t+1)} \left(W^{(t+1)}\right)^T\right\|_F^2 &= \sum_{i=1}^M \text{tr}\left(A_i^T A_i\right) - \text{tr}\left(\left(W^{(t+1)}\right)^T A_i^T A_i W^{(t+1)}\right) \\ &= \sum_{i=1}^M \text{tr}\left(A_i^T A_i\right) - \left\|A_i W^{(t+1)}\right\|_F^2 \end{aligned} \tag{33}$$

According to theorem 1, there are:

$$\sum_{i=1}^M \|A_i W^{(t+1)}\|_F \geq \sum_{i=1}^M \|A_i W^{(t)}\|_F \tag{34}$$

So, Equation (34) can get:

$$\sum_{i=1}^M \|A_i - A_i W^{(t+1)} (W^{(t+1)})^T\|_F \leq \sum_{i=1}^M \|A_i - A_i W^{(t)} (W^{(t)})^T\|_F \tag{35}$$

Theorem 5 From Theorem 3 and theorem 4, we can get:

$$\sum_{i=1}^M \frac{\|E_i\|_F^p}{\|A_i W\|_F^p} \Big|_{W=W^{(t+1)}} \leq \sum_{i=1}^M \frac{\|E_i\|_F^p}{\|A_i W\|_F^p} \Big|_{W=W^{(t)}} \tag{36}$$

Proof: By Theorem 3 and theorem 4, and $0 < p < 2$, there are

$$\sum_{i=1}^M \|A_i W^{(t+1)}\|_F^p \geq \sum_{i=1}^M \|A_i W^{(t)}\|_F^p \tag{37}$$

and

$$\sum_{i=1}^M \|A_i - A_i W^{(t+1)} (W^{(t+1)})^T\|_F^p \leq \sum_{i=1}^M \|A_i - A_i W^{(t)} (W^{(t)})^T\|_F^p \tag{38}$$

So, it's easy to get:

$$\sum_{i=1}^M \frac{\|E_i\|_F^p}{\|A_i W\|_F^p} \Big|_{W=W^{(t+1)}} \leq \sum_{i=1}^M \frac{\|E_i\|_F^p}{\|A_i W\|_F^p} \Big|_{W=W^{(t)}} \tag{39}$$

According to the conclusion of Theorem 5, algorithm 1 continuously reduces the function value of an objective function Equation (7) in iteration, so W will continue to approach the optimal solution. Finally, algorithm 1 will converge to the optimal local resolution of the objective function Equation (7). Algorithm 2 is based on Algorithm 1, so algorithm 2 must link to the optimal local solution.

3.3.2. Analysis of Rotation Invariance

In this part, we mainly show that Fp-2DPCA has good rotation invariance. Rotation invariance means that the low dimensional representation remains unchanged under the rotation transformation of the sample space.

Theorem 6 The solution of Fp-2DPCA is rotationally invariant.

Proof: given any orthogonal matrix Γ ($\Gamma^T \Gamma = I$), for each step of algorithm 1 to get the solution W , there is

$$\begin{aligned} \frac{\|E_i\|_F^p}{\|A_i W\|_F^p} &= \frac{\|A_i - A_i W W^T\|_F^p}{\|A_i W\|_F^p} = \frac{\|A_i - Z_i W^T\|_F^p}{\|Z_i\|_F^p} = \frac{\|(A_i - Z_i W^T) \Gamma^T \Gamma\|_F^p}{\|Z_i\|_F^p} \\ &= \frac{\left(\sqrt{\sum_{j=1}^m \|(A_i(j, :) \Gamma^T - Z_i(j, :) W^T \Gamma^T)\|_2^2} \right)^p}{\|Z_i\|_F^p} \\ &= \frac{\left(\sqrt{\sum_{j=1}^m \|A_i(j, :) \Gamma^T - Z_i(j, :) W^T \Gamma^T\|_2^2} \right)^p}{\|Z_i\|_F^p} \end{aligned} \tag{40}$$

$$= \frac{\left(\sqrt{\sum_{j=1}^m \|\tilde{A}_i(j, \cdot) - Z_i(j, \cdot) \tilde{W}^T\|_2^2} \right)^p}{\|Z_i\|_F^p} = \frac{\|\tilde{A}_i - Z_i \tilde{W}^T\|_F^p}{\|Z_i\|_F^p}$$

where $\tilde{W} = \Gamma W$, $\tilde{A}_i = A_i \Gamma$, so $\tilde{A}_i \tilde{W} = A_i \Gamma^T \Gamma W = A_i W$. Equation (40) shows that if W is the solution of the objective function, \tilde{W} is the solution of the objective function under the orthogonal matrix Γ transformation.

Besides, compared with the existing 2DPCA method based on L_1 norm, this method considers the reconstruction error directly and synthesizes the variance of low dimensional data in the criterion function. Also, it has strong robustness to outliers and is related to the covariance matrix of the image.

4. Numerical Experiment

In this part, we use four most advanced algorithms, namely 2DPCA [2], 2DPCAL1 [10], angel-2DPCA [12] and (2D)²PCA [6] to compare unilateral Fp-2DPCA and bilateral Fp-2DPCA. Our experiment uses two famous face datasets, the ORL database [16], and the Extended Yale Face B database [17].

Because the feature matrices of different algorithms have different dimensions, we use the same feature size for all methods for a fair comparison. For example, if the column dimension of one-sided dimensionality reduction methods (2DPCA, 2DPCAL1, angel-2DPCA) is r . The row and column dimension reduction results of two-sided dimensionality reduction methods (2D) 2DPCA and Fp-2DPCA) are l' and r' respectively to make $l' \times r' \approx m \times r$. For simplification, set $l' = r'$. For all algorithms, 1-Nearest neighbor classification (1-NN) is used for variety.

ORL face database consists of 400 frontal images collected from different lighting conditions, with ten shots per person. In this database, each print is adjusted to 112×92 pixels. We randomly selected seven pictures for each person and put the noise in the range of 0 - 255 in the chosen images. The noise location is random, and the ratio of noise pixels to image pixels is 0.05 - 0.15. **Figure 1** shows some photos of this database and corresponding pictures with increased noise. In our experiment, we randomly selected seven images for training, and the remaining three ideas were tested

Figure 2 shows the reconstructed images of 2DPCA, L1-2DPCA, (2D) 2PCA, angel-2DPCA, Fp-2DPCA ($p = 1$) and Fp-2DPCA ($p = 0.5$) from left to right. The reconstructed image of 2DPCA is still very blurred, L1-2DPCA is slightly better, and angel-2DPCA and Fp-2DPCA are significantly better than the previous methods.

If the dimension reduction is too low, the reconstructed image will be difficult to recognize, so we choose to reduce columns' dimension to 30. For (2D) 2DPCA, the dimensions of rows and columns are reduced to 50 ($50 \times 50 < 112 \times 30$). Our algorithm selects rows and columns to reduce to 30 ($30 \times 30 < 50 \times 50$) to compare the effect. In **Table 3**, we list the average classification accuracy (Acc) and standard deviation (std) obtained by the algorithm on the original dataset

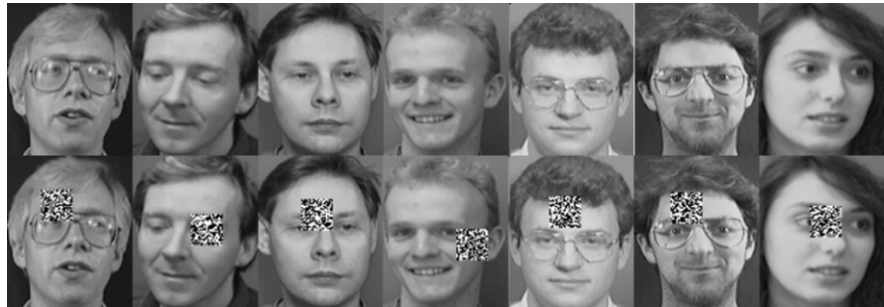


Figure 1. Some noise images in ORL dataset.



Figure 2. Comparison of reconstructed images under various methods.

Table 3. Recognition accuracy.

Algorithm	dim	Acc \pm std% (clean)	Acc \pm std% (noised)	Diff%
2DPCA	112×30	66.66 ± 1.18	62.50 ± 1.25	4.16
2DPCA-L1	112×30	67.50 ± 1.14	64.10 ± 1.52	3.40
(2D) 2PCA	50×50	73.33 ± 1.09	70.19 ± 1.50	3.14
Angel-2DPCA	112×30	75.83 ± 1.01	73.22 ± 1.29	2.61
B-Fp-2DPCA ($p = 1$)	30×30	79.16 ± 1.12	77.45 ± 1.49	1.71
B-Fp-2DPCA ($p = 0.5$)	30×30	81.66 ± 1.04	79.98 ± 1.35	1.68

(clean) and noisy dataset (noisy). The last column shows the difference of precision means between the clean data set and noise data set.

In **Table 3**, the method based on angel-2DPCA is better than 2DPCA and 2DPCA-L1 in both clean data sets and noisy data sets. It may be due to the difference that angel-2DPCA is reduced from square F-norm to F-norm, robust to outliers. Besides, the last column's values show that the performance of all algorithms is degraded by noise. However, the angle type algorithm has a small decrement. It again supports the robustness of the angle model. On the other hand, among all the comparison algorithms, Fp-2DPCA is the best. It has the highest accuracy and the smallest difference.

5. Conclusion

In this paper, we propose a new 2DPCA model, which considers the reconstruction error and considers the maximum variance, and adopts the f norm, which has good robustness to outliers. We increase the parameter p to make the model have more choices. The experiment shows that the effect of P-value 0.5 is better. We also extend it to the bilateral projection model and propose the bilateral

FP-2DPCA. The new Fp-2DPCA reduces the dimension of the original image matrix from both row and column. Experimental results on face data sets show that the proposed method can achieve higher performance with fewer measurements.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Wold, S., Esbensen, K.H. and Geladi, P. (1987) Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, **2**, 37-52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)
- [2] Yang, J., Zhang, D., Frangi, A.F. and Yang, J.-Y. (2004) Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **26**, 131-137. <https://doi.org/10.1109/TPAMI.2004.1261097>
- [3] Jeong, Y.W. and Kim, H.S. (2009) New Speaker Adaptation Method Using 2-D PCA. *IEEE Signal Processing Letters*, **17**, 193-196. <https://doi.org/10.1109/LSP.2009.2036696>
- [4] Wang, D. and Lu, H. (2012) Object Tracking via 2DPCA and ℓ_1 -Regularization. *IEEE Signal Processing Letters*, **19**, 711-714. <https://doi.org/10.1109/LSP.2012.2215320>
- [5] Kong H., Li, X. and Wang, L. (2005) Generalized 2D Principal Component Analysis. 2005 *IEEE International Joint Conference on Neural Networks*, Montreal, 31 July-4 August 2005, 108-112. <https://doi.org/10.1109/IJCNN.2005.1555814>
- [6] Zhang, D. and Zhou, Z.H. (2005) (2D)²PCA: Two-Directional Two-Dimensional PCA for Efficient Face Representation and Recognition. *Neurocomputing*, **69**, 224-231. <https://doi.org/10.1016/j.neucom.2005.06.004>
- [7] Xu A., Jin X. and Jiang Y (2006) Complete Two-Dimensional PCA for Face Recognition. *18th International Conference on Pattern Recognition*, Hong Kong, 20-24 August 2006, 481-484. <https://doi.org/10.1109/ICPR.2006.395>
- [8] Kim, Y.G., Song, Y.J., Chang, U.D., Kim, D.-W., Yun, T.-S. and Ahn, J.-H. (2008) Face Recognition Using a Fusion Method Based on Bidirectional 2DPCA. *Applied Mathematics and Computation*, **205**, 601-607. <https://doi.org/10.1016/j.amc.2008.05.032>
- [9] Kwak, N. (2008) Principal Component Analysis Based on L1-Norm Maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **30**, 1672-1680. <https://doi.org/10.1109/TPAMI.2008.114>
- [10] Li, X., Pang, Y. and Yuan, Y. (2010) L1-Norm-Based 2DPCA. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 1170-1175. <https://doi.org/10.1109/TSMCB.2009.2035629>
- [11] Ng, A.Y. (2004) Feature Selection, L1 vs. L2 Regularization, and Rotational Invariance. *Proceedings of the 21st International Conference on Machine Learning*, Banff, June 2004, 78. <https://doi.org/10.1145/1015330.1015435>
- [12] Gao, Q., Ma, L. and Liu, Y. (2018) Angle 2DPCA: A New Formulation for 2DPCA. *IEEE Transactions on Systems, Man, and Cybernetics*, **48**, 1672-1678.

-
- <https://doi.org/10.1109/TCYB.2017.2712740>
- [13] Wang, Y. and Li, Q. (2019) Robust 2DPCA with F-Norm Minimization. *IEEE Access*, **7**, 68083-68090. <https://doi.org/10.1109/ACCESS.2019.2918702>
- [14] Min, L. and Song, L. (2012) Face Recognition Based on PCA and 2DPCA with Single Image Sample. *Proceedings of the 2012 9th Web Information Systems and Applications Conference*, Haikou, 16-18 November 2012, 111-114. <https://doi.org/10.1109/WISA.2012.20>
- [15] Xu, C.M., Jiang, H.B. and Yu, J.J. (2008) Robust Two-Dimensional Principal Component Analysis. *Proceedings of the Chinese Control Conference*, Kunming, 16-18 July 2008, 452-455. <https://doi.org/10.1109/CHICC.2008.4605066>
- [16] Samaria, F.S. and Harter, A.C. (1994) Parameterisation of a Stochastic Model for Human Face Identification. *Proceedings of the Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*, Sarasota, 5-7 December 1994, 138-142. <https://doi.org/10.1109/ACV.1994.341300>
- [17] Georghiades, A.S., Belhumeur, P.N. and Kriegman, D.J. (2001) From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**, 643-660. <https://doi.org/10.1109/34.927464>