IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 24, 2019, accepted May 15, 2019, date of publication May 24, 2019, date of current version June 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2918702

Robust 2DPCA With *F*-Norm Minimization

YONG WANG[1] AND QIN LI[2]
[1]The 10th Research Institute of China Electronic Group Corporation, Chengdu 610036, China
[2]School of Software Engineering, Shenzhen Institute of Information Technology, Shenzhen 710071, China

Corresponding author: Qin Li (liqin@sziit.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61773302.

ABSTRACT While feature extraction based on two-dimensional principal component analysis (2DPCA) is widely used in image recognition, such a method usually fails to handle the noise and outliers, because adopted F-norm square actually exaggerates the effect of outliers. To tackle the aforementioned problem, we present a novel algorithm called Area-2DPCA, which uses F-norm to characterize the variance and reconstruction error. By doing so, the project directions, which minimize the summation of the area between projection directions and reconstruct error of each data, can be found. Moreover, the Area-2DPCA sets different weighted coefficients to each residual error. To find the solution of our model, a non-greedy algorithm, which has a closed form solution in each step, is presented. The extensive experimental results demonstrate the superiority of our proposed model, compared with the state-of-the-art.

INDEX TERMS Two-dimensional principal component analysis, robust feature extraction, $\ell_{2,1}$-norm.

I. INTRODUCTION

Finding an effective image representation is a fundamental issue in the areas of image representation and pattern recognition [1]–[4]. Principal component analysis (PCA) [5] and two-dimensional principal component analysis (2DPCA) [6] are two most typical unsupervised methods for image representation and classification. Basically, PCA and 2DPCA employ squared Euclidean distance, which is equivalent to squared F-norm of matrix or vector, to characterize the reconstruction error or variance in the objective function. As is well known, squared Euclidean distance are not immune to the noise and outliers due to the fact that outlying measurements can arbitrarily skew the solution from the desired solution [7]–[12].

To address the above-mentioned problem, many robust PCA methods have been developed in recent years, in which the techniques based on nuclear-norm and L1-norm are two representatives [13]–[16]. Nuclear-norm based technique can well attain the clean data from the noised data if data only include a fraction of the corrupted entries and get good performance [13], [17], but it is incapable of handling new data directly. To tackle this problem, Bao *et al*. proposed an inductive robust PCA (IRPCA) [18], which aims to learn a square projection matrix P with low-rank structure and directly gets the clean data from the testing data by linear transformation. However, it fails to obtain the low-dimensional representation for high-dimensional data.

Different from IRPCA and nuclear-norm PCA, L1-norm based PCA technique can get robust projection directions for dimensionality reduction, and many related methods have been developed for image classification especially for face recognition. L1-norm based PCA methods can be roughly divided into two formulations: reconstruction error based formulation and covariance based formulation. Reconstruction error based formulation minimizes L1-norm reconstruction error to get optimal solution. The most representative one is L1-PCA [19]. L1-PCA is robust to outliers and performs well, but it is hard to solve. Covariance based formulation maximizes L1-norm variance to obtain the optimal projection directions. For example, Kwak [20] sought the projection directions by maximizing L1-norm variance with a greedy algorithm. The proposed method is called PCA-L1 that does not guarantee the maximization of the objective function. To handle this disadvantages, Nie *et al.* attained the optimal solution of PCA-L1 by non-greedy algorithm [21]. Lu *et al.* integrated correlation of data into PCA-L1 and obtained the impressive experimental results [22].

To take advantage of the spatial structure information of image, many robust approaches of dimensionality reduction based on image matrix have been presented. 2DPCA-L1 is one of the most representative methods and can be considered as the extension work of 2DPCA. It is usually solved by different algorithms such as greedy strategy [23] and non-greedy

The associate editor coordinating the review of this manuscript and approving it for publication was Huanqiang Zeng.

strategy [24]. Wang and Wang [25] added a sparse limitation to 2DPCA-L1 and proposed 2DPCA-L1-S. Inspired by 2DPCA-L1 and nuclear-norm, Zhang *et al.* [17] used nuclear-norm to characterize the spatial structure information of the reconstruction error and developed N-2DPCA (nuclear-based 2DPCA). However, the aforementioned matrix based methods still have the following disadvantages. First, it is unclear whether they have rotational invariance that is important in image analysis and pattern recognition [26]–[29]. Second, they consider each data with the same contribution. Third, they ignore the relationship between the variance and the reconstruction error.

To tackle the above-mentioned disadvantages, we present a novel robust formula for 2DPCA, namely Area-2DPCA, which uses F-norm to jointly characterize the low-dimensional representation and reconstruction error, and eventually integrates them into one criterion function. To solve Area-2DPCA, an efficient iterative algorithm is presented. Experimental results demonstrate the effectiveness of our model. The highlights of our model are summarized as follows:

- In our model, we use F-norm to characterize the variation of the low-dimensional representation and corresponding reconstruction error, which is robust to outliers. Apart from it, it has rotational invariance [7], [30], [31];
- Our model considers the relationship between the variance and the reconstruction error, which can adaptively assign different coefficients to each reconstruction error.
- The solution is relevant to the weighted covariance matrix, which can depict the data structure well.

II. PROBLEM FORMULATION

2DPCA is a two-dimensional extension of PCA. Its basic idea is the same as PCA, which preserves the maximum information of original image data by maximizing the sum of the variance after projection of the original image data on the principal components. Suppose $A = \{A_i \in R^{m\times n}\ (i = 1, 2, \cdots, M)\}$ is a set of training samples with M picture data, where m and n are the rows and columns of the picture data, respectively. The linear transformation matrix is $W = [w_1, w_2, \cdots, w_d] \in R^{n\times d}$, where d denotes the number of transformed projection vectors. Without loss of generality, we assume the training images are centralized. Thus, the objective function of 2DPCA is [6]

$$\max_{W^TW=I_d} tr(\sum_{i=1}^{M} W^T(A_i)^T A_i W) = \max_{W^TW=I} \sum_{i=1}^{M} \|A_iW\|_F^2 \quad (1)$$

which is equivalent to the following model (2).

$$\arg\min_{W^TW=I_d} \sum_{i=1}^{M} \|E_i\|_F^2 \quad (2)$$

where $I_d \in R^{d\times d}$ is an identity matrix. $\|\cdot\|_F$ denotes the Frobenius norm (F-norm) of a matrix, $E_i = A_i - A_iWW^T$.

The objective functions (1) and (2) show that 2DPCA implicitly considers each image picture or reconstruction error with the same contribution. This affects the robustness of 2DPCA to noise and outliers. To deal with this disadvantage, many robust 2DPCA methods with L1-norm as distance metric have been developed. One of the most representative objective functions is [23]

$$\max_{W^TW=I_d} \sum_{i=1}^{M} \|A_iW\|_{L_1} \quad (3)$$

where $\|\cdot\|_{L_1}$ denotes the L1-norm of a matrix that can be defined as $\|Y\|_{L1} = \sum_{j=1}^{n} \|y_j\|_1$, y_j is the j-th column of matrix Y.

Compared with the traditional 2DPCA technique, the model (3) may reduce the impact of outliers by L1-norm. However, **it has several shortcomings**. Firstly, solution of the model (3) is irrelevant to the scatter matrix that well characterizes the data structure [27]; Secondly, the model (3) does not guarantee the minimization of the total reconstruction error of data, which is the true goal of 2DPCA, due to the fact $\sum_{i=1}^{M} \|E_i\|_{L_1} + \sum_{i=1}^{M} \|A_iW\|_{L_1} \neq \sum_{i=1}^{M} \|A_i\|_{L_1}$. It means that solution of the model (3) is not the solution of the model (4).

$$\min_{W^TW=I_d} \sum_{i=1}^{M} \|E_i\|_{L_1} \quad (4)$$

Thirdly, it does not consider the relationship between the variance in the projected subspace and corresponding reconstruction error under the L1-norm distance metric. Fourth, it still considers each picture with the same contribution. This also affects the robustness of the model due to the fact that outliers or noise make samples have sparse distribution. To handle the aforementioned disadvantages, Area-2DPCA is presented in Section 3.

III. AREA-2DPCA

A. MOTIVATION AND OBJECTIVE FUNCTION

It can be concluded from the above analysis that squared F-norm exaggerates the role of some data points, which significantly deviate from the clean data points, in solving the model of 2DPCA. This degrades the robustness of 2DPCA to outliers. Thus, to tackle the aforementioned limitation, we should employ a suitable distance metric that not only reduces the impact of outliers in the objective function but also characterizes the geometric structure. In a normative sense, *F-norm* and *squared F-norm* have the same role in characterizing both the scatter of data and geometric structure. The main difference between them is that *F-norm* can make the difference of the effect of different data points tend to become smaller, compared with *squared F-norm*. Thus, if we select F-norm as distance metric in 2DPCA, it will have the following two advantages. First, it can well capture geometric structure and have rotational invariance. Second, it can degrade the role of outliers in solving the optimal projection directions; Third, it helps enhance the role of some neighbors' data points having the different labels. This helps

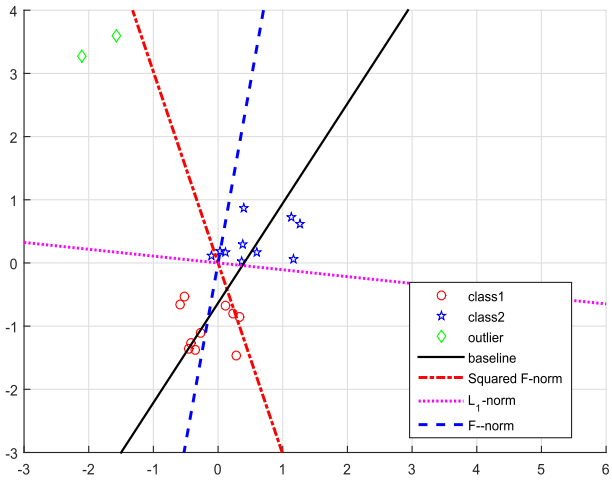

FIGURE 1. Data points and optimal projection directions of 2DPCA with different distance metrics.

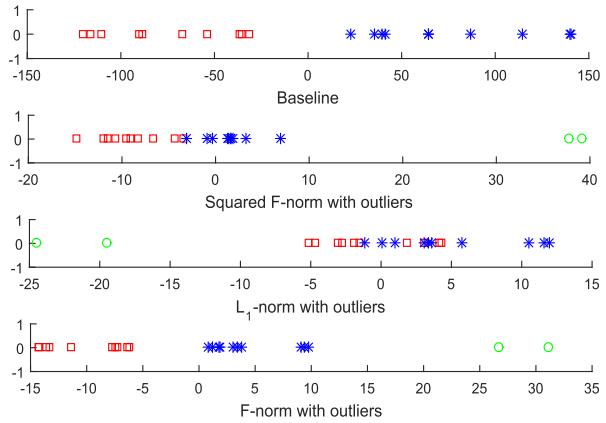

FIGURE 2. Projections of 2DPCA with different distance metrics.

encode the discriminant information of data. To illustrate the above-mentioned advantages, we use Matlab to get some data points, which are marked by different shapes, and two outliers as in [14]. Data with the same shape belong to the same class and each class has 10 true data points. The optimal direction of traditional 2DPCA is obtained under the 20 true data points and considered as the baseline (See Fig. 1). We also plot the optimal projections of 2DPCA with different distance metrics such as squared F-norm, L1-norm and F-norm under the 20 true data points and 2 outliers in Fig. 1. Fig. 2 shows the corresponding projected data of the aforementioned methods. Note that, we do not use the label information in solving the optimal directions of 2DPCA with different distance metrics. From Fig. 1 and Fig. 2, we have that, (1) the optimal projection of F-norm is closer to baseline, compared to squared F-norm and ℓ_1-norm. This shows that F-norm can improve the robustness of algorithm. (2) The low-dimensional representation of F-norm are separable, compared squared F-norm and L1-norm. It illustrates that F-norm well encodes discriminative information.

Moreover, the relationship between variance and corresponding reconstruction error is nonlinear due to the fact $\|E_i\|_F + \|A_iW\|_F \neq \|A_i\|_F$. Thus, maximization of the total variation of data does not guarantee the minimization of reconstruction error. According to the aforementioned analysis, we propose a novel method, namely Area 2DPCA for dimensionality reduction. Area 2DPCA uses F-norm to characterize the low-dimensional representation and reconstruction error, and then integrates them into the criterion function. Specifically, the goal of Area 2DPCA is to find the projection directions which minimize the summation of area between projection directions and reconstruct error of each data (See Fig. 3). The objective function of Area-2DPCA is

$$\min_{W^TW=I} \sum_{i=1}^{M} \left\|A_i - A_iWW^T\right\|_F \|A_iW\|_F$$

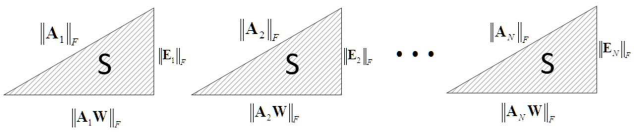

FIGURE 3. Reconstruction error vs. projection.

As can be seen in the aforementioned model, when $\|A_iW\|_F$ becomes zero, the above model is still minimum. In this case, the reconstruction error is large. This contradicts the purpose of PCA. In order to avoid this trivial solution, we add a constant γ for it. Thus, we rewrite the objective function of Area-2DPCA as follows:

$$\min_{W^TW=I} \sum_{i=1}^{M} \left\|A_i - A_iWW^T\right\|_F (\|A_iW\|_F + \gamma) \tag{5}$$

B. ALGORITHM

To solve the model (5), we first introduce the following related theorems [32]:

Theorem 1: For the same order matrix X, Y, we have

$$tr(X^TY) \leq \|X\|_F\|Y\|_F \tag{6}$$

with equality if and only if X or Y is a multiple of the other.

Theorem 2: Suppose $U\Sigma V^T$ is the compact singular value decomposition (SVD) of $A \in R^{m\times n}$, then $W = UV^T$ is the solution of the model (7).

$$\arg\max_{W^TW=I_d} tr(W^TA) \tag{7}$$

where $V^TV = U^TU = I_d$, $\Sigma \in R^{d\times d}$ is a diagonal matrix and $\Sigma(j,j) = \lambda_j$ is jth singular value of A. $d = rank(A)$ is the rank of A.

Algorithm 1: Algorithm For The Model (5)

Input: Given training data $A_i \in R^{m\times n}$, $(i = 1, \cdots, M)$, ε=1$e-8$.

Initialize: $W^{(1)} \in R^{n\times d}$ which satisfies ${W^{(1)}}^T W^{(1)} = I_d$, $t = 1$. $\delta = \left|J(W^{(t)}) - J(W^{(t-1)})\right|$, where $J(W^{(t)}) = \sum_{i=1}^{M} \left\|A_i - A_i W W^T\right\|_F \|A_i W\|_F$.

while $\delta \geq \varepsilon$ **do**

1.Calculate $d_i^{(t)}$ according to Eq. (9)

2.Calculate $H^{(t)}$ by Eq. (11).

3.Calculate SVD of matrix $H^{(t)}$, i.e., $H^{(t)} = U^{(t)}\Sigma(V^{(t)})^T$.

4.Solve $W^{(t+1)} = \arg\max\ tr\left(W^T H^{(t)}\right)$, i.e., $W^{(t+1)} = U^{(t)}(V^{(t)})^T$.

5.Update δ.

6.Update t: $t \leftarrow t + 1$.

end while

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

To solve the model (5), by simple algebra, we have

$$\begin{aligned}
&\sum_{i=1}^{M} \left\|A_i - A_i W W^T\right\|_F (\|A_i W\|_F + \gamma) \\
&\quad = \sum_{i=1}^{M} \frac{\left\|A_i - A_i W W^T\right\|_F^2}{\left\|A_i - A_i W W^T\right\|_F} (\|A_i W\|_F + \gamma) \\
&\quad = \sum_{i=1}^{M} \frac{tr(A_i^T A_i) - tr(W^T A_i^T A_i W)}{\left\|A_i - A_i W W^T\right\|_F} (\|A_i W\|_F + \gamma) \qquad (8)
\end{aligned}$$

Combining Eq. (8), we can rewrite the model (5) as

$$\min_{W^T W = I} \sum_{i=1}^{M} (tr(A_i^T A_i) - tr(W^T A_i^T A_i W)) d_i \qquad (9)$$

where $d_i = \dfrac{(\|A_i W\|_F + \gamma)}{\left\|A_i - A_i W W^T\right\|_F}$.

From the formula (9), we can see that there are two unknown variables W and d_i which relates to W. Thus, it is difficult to directly attain the optimal solution of the model. If we know d_i in advance, then it is easy to solve the model (9). Inspired by this, we propose an algorithm for alternatively updating W (while fixing d_i) and d_i (while fixing W) until the objective function value converges. Specifically, the solution process is as follows:

First, update W while fixing $d_i^{(t)}$. Then, the first term in the model (9) is constant, and the model (9) becomes

$$\arg\max_{W^T W = I_d} \sum_{i=1}^{M} tr(W^T A_i^T d_i A_i W) = \arg\max_{W^T W = I_d} tr(W^T H) \qquad (10)$$

where

$$H = \sum_{i=1}^{M} {A_i}^T d_i A_i W \qquad (11)$$

Suppose $H = U\Sigma V^T$ is the SVD of matrix H, then,

$$W = UV^T \qquad (12)$$

is the optimal solution of the model (10) via theorem 2.

Second, calculate d_i with the updated W. Repeat the iterative process until the objective function value converges, and we can obtain the final projection matrix. After the above analysis, we summarize the solution of solving Area 2DPCA in algorithm 1.

C. ROTATIONAL INVARIANCE

Rotational invariance means that the projected data remain unchanged under the given rotational matrix Γ ($\Gamma^T\Gamma = I$). This property is usually emphasized in image analysis and pattern recognition. Thus, we have

Theorem 3: Solution of Area-2DPCA has the rotational invariance property.

Proof: Given an arbitrary rotation matrix $\Gamma(\Gamma^T\Gamma = I)$, and denote by $Z_i = A_i W$, which is low-dimensional representation of A_i, then, for each term in Eq. (5), we have

$$\begin{aligned}
&\left\|A_i - A_i W W^T\right\|_F \|A_i W\|_F \\
&\quad = \left\|(A_i - Z_i W^T)\right\|_F \|Z_i\| \\
&\quad = \left\|(A_i - Z_i W^T)\Gamma^T\Gamma\right\|_F \|Z_i\| \\
&\quad = \sqrt{\sum_{j=1}^{m} \left\|(A_i(j,:)\Gamma^T - Z_i(j,:)W^T\Gamma^T)\Gamma\right\|_2^2} \|Z_i\| \\
&\quad = \sqrt{\sum_{j=1}^{m} \left\|A_i(j,:)\Gamma^T - Z_i(j,:)W^T\Gamma^T\right\|_2^2} \|Z_i\| \\
&\quad = \sqrt{\sum_{j=1}^{m} \left\|\widehat{A}_i(j,:) - Z_i(j,:)\widehat{W}^T\right\|_2^2} \|Z_i\|_F \\
&\quad = \left\|(\widehat{A}_i - Z_i\widehat{W}^T)\right\|_F \|Z_i\|_F \qquad (13)
\end{aligned}$$

where $\widehat{W} = \Gamma W$, $A_i(j,:)$, $\widehat{A}_i(j,:)$, and $Z_i(j,:)$ denote the j-th row of matrices A_i, $\widehat{A}_i$, Z_i, respectively.

Eq. (13) indicates that, suppose W^* is the solution of the model (5), then $\widehat{W}$ must be the solution of the model (5) under the rotational matrix Γ. Then,

$$\widehat{A}_i\widehat{W} = A_i\Gamma^T\Gamma W = A_i W = Z_i \qquad (14)$$

Eq. (14) shows that the projected data $Z_i = A_i W$ remains unchanged under a rotational matrix Γ.

From the above analysis, we can draw that the features extracted by subspace learning method should remain unchanged during the rotation transformation because the data distribution under the rotation transformation of the sample space remains unchanged. It can help prevent the performance of subspace learning technology from degrading. Moreover, as is known to all, the projection directions of our model(5) depend on the matrix H which relates to the weighted covariance matrix of data. Finally, our model assigns different coefficients, which relate to the variance of data, to each image. Thus, our model directly considers

the relationship between reconstruction error and variance of data.

IV. EXPERIMENTS

To verify the feasibility and effectiveness of our proposed algorithm, we validate the proposed algorithm in six famous databases including ORL, COIL20, AR, CMU PIE, LFWCrop and Extended Yale B, and compare it with 2DPCA-L1-S [25], 2DPCA-L1 [23], N-2DPCA [17] and 2DPCA [6]. In ORL, COIL20 and LFWCrop databases, each features are normalized into [0, 1]. In each database, we randomly select 20% images and randomly place a 1/4 size of occlusion (white black dots and image object respectively) in the selected images. Furthermore, we use the reconstruction error, which is calculated by Eq. (15), to measure the quality of the aforementioned dimensionality reduction methods.

$$error = \frac{1}{n}\sum_{i=1}^{n}\left\|x_i^{clean} - WW^T x_i^{clean}\right\|_2 \tag{15}$$

where n denotes the number of training data, W denotes the learned projection matrix composed of multiple projection directions. $x_i{}^{clean}$ denotes the ith clean data. In the following experiments, we set γ as 10^{-3}.

A. EXPERIMENTS ON THE IMAGE DATABASES

The AR database [33] has at least 4,000 images that are from 126 people, and includes 70 males and 56 Females. Images in this database have different facial expressions, lighting conditions and different picture occlusions. Most of the people in this database were collected in two periods, one week apart, with 13 images of 120 people collected in each period. The 120 people include 65 males and 55 females and The 13 images contain 6 images of light changes, 3 images of wearing scarves and 3 pictures of wearing sunglasses. We manually cut the face images, and then normalized each image to be 50×40 pixels [6]. The images of one person are shown in Fig. 4. In the experiment, we scrambled and rearranged all the data, and then randomly selected 13 pictures per person for training data and the remaining images were used as test data. We repeated this process 10 times randomly

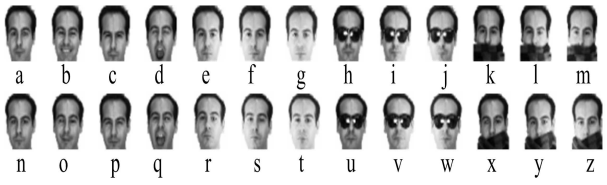

FIGURE 4. Images of one person in the AR database.

The Extended Yale B database [34] includes 2,414 images for 38 people. These images contain 64 different lighting conditions and 9 different postures. In this database, there are 60 images in category 11 and 13, 59 in category 12, 62 in category 15, 63 in category 14 and 64 in other categories.

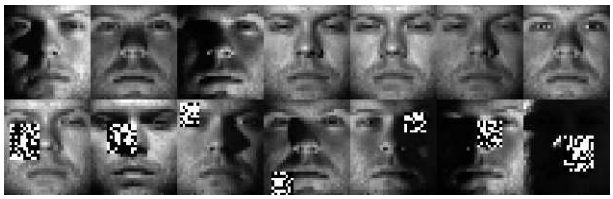

FIGURE 5. The first row is the original pictures in the Extended Yale B database and second row is some noisy images.

Fig.5 shows some of the sample images. The first row in the Fig.5 is the original pictures with varying lighting and the second row is the images after adding the random noise block. In the experiments, we resized each picture to 32×32 pixels and 14 pictures of each class were randomly selected for adding noise blocks. Randomly selected 32 pictures from each class for training that contains 7 noisy images, and the remaining for the test. Repeat all the experiments 10 times.

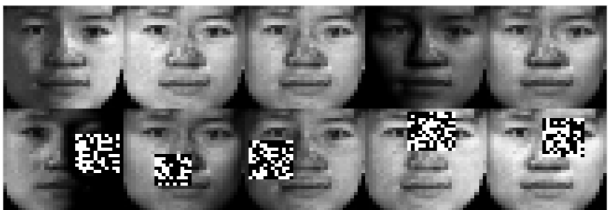

FIGURE 6. The first row is the original pictures in the CMU PIE database and second row is some noisy images.

The CMU PIE database [35] consists of more than 41,368 face images taken by 68 volunteers in 13 different poses, 4 different facial expressions and 43 different light conditions. In this experiment, a sub-dataset of the PIE database was used, which includes 2,856 images of 68 individuals and 42 images per person, and each image is cropped to 32×32 pixels. In Fig.6, we show some samples in the PIE database (See first row) and noisy images (See second row). In our experiment, 10 images in each class were randomly selected to add noise, and 21 images in each class were randomly selected for training, which include 5 noisy images and 16 noiseless images, and the remaining for the test. We repeat this process randomly 10 times.

FIGURE 7. Some sample instances in the COIL20 dataset.

The COIL20 database [36], also known as the Columbia Object Image Library. This database contains 20 objects, and the object has a variety of complex geometric and reflective properties. Each object which is stable at the approximate center of the turntable rotates horizontally through 360 degrees and is taken a picture every 5 degrees, so each object has a total of 72 pictures, which is 64×64 in size. Fig.7 shows some samples in the COIL20 database.

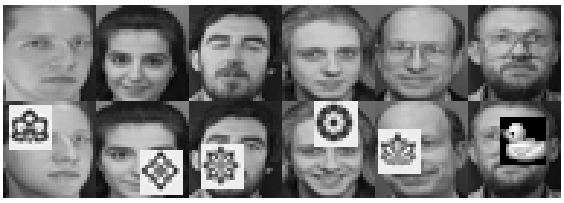

FIGURE 8. Some sample instances in the ORL dataset.

20 images of each object are randomly selected for training, and the remaining for the test. We repeat it 10 times.

Face images in the ORL database [37] are taken by the University of Cambridge Laboratory from April 1992 to April 1994. it has a total of 40 subjects of different ages, races and genders. In this database, 10 images per person for a total of 400 grayscale images and the image size is 92×112 and the image background is black. Some facial expressions and details have some changes, such as laughing and not laughing, eyes open or closed, with or without glasses and so on. Fig. 8 shows some samples in this database. In our experiment, each picture is cropped to 32×32 pixels, we randomly selected half of the pictures of each category for training and the remaining as the test, and then repeat 10 times.

LFWcrop [38] is a clipped version of the Labeled Faces in the Wild (LFW) [39] database, reserving only the center part of each image (i.e, face), with the majority of the background in the dataset clipped off. Then the remaining area was scaled to 64×64 pixels. The cropped face in LFWcrop shows real-life conditions including misalignment, scale variations, in-plane and out-of-plane rotations due to the fact that the position and size of the faces in LFW was determined by using an automatic face locator (detector). In the LFWcrop database, the number of each person is unequal. We show some samples in Fig.9. In our experiments, we chose the person whose pictures are more than 20 but less than 100 as the gallery. This gallery has a total of 1883 samples of 57 classes. We randomly select ninety percent of images of each person for training, the rest of the samples for testing, and then repeat this process 10 times.

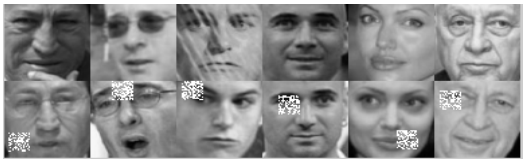

FIGURE 9. Some sample instances in the LFWCrop dataset.

B. EXPERIMENTAL RESULTS

Based on the aforementioned experiments, we list the average classification accuracy, recall, precision and the corresponding standard deviation (Std) of five algorithms on several database (AR, CMU PIE and Yale B) in Table 1, Table 2 and 3, respectively. Table 4 and 5 list the average reconstruction error and the corresponding std of five algorithms on the ORL, COIL20, AR and LFWCrop databases with black and while dot noise and object noise, respectively. To evaluate the complexity of our algorithm, we show the average running time of each algorithm on the AR, COIL20, ORL, and LFWcrop databases in Table 6. Note that, the aforementioned five algorithms run on the PC computer with Intel Core i7-4770 CPU M620 @ 3.40 GHz 8 GB RAM, simultaneously we plot the convergence curve of our algorithm on the four databases (AR, ORL, COIL20 and LFWcrop). Comparing with the aforementioned experimental results, we have several interesting observations as follows:

TABLE 1. The average classification accuracy (%) and std on the PIE, AR and Extended Yale B databases.

Methods	2DPCA	2DPCAL1	N-2DPCA	2DPCA L1-S	Area-2DPCA
Yale B	59.92 ±0.42	60.23 ±0.38	59.92 ±0.42	60.28 ±0.36	**65.11 ±0.61**
AR	83.25 ±0.83	83.26 ±0.84	83.20 ±0.80	83.26 ±0.84	**89.72 ±0.72**
PIE	85.59 ±0.47	86.00 ±0.53	85.60 ±0.51	85.59 ±0.50	**87.27 ±0.71**

TABLE 2. The average recall (%) and std on the PIE, AR and Extended Yale B databases.

Methods	2DPCA	2DPCAL1	N-2DPCA	2DPCA L1-S	Area-2DPCA
Yale B	57.97 ±0.85	58.66 ±0.82	57.97 ±0.85	58.57 ±0.75	**65.06 ±1.21**
AR	82.44 ±0.76	82.44 ±0.76	82.44 ±0.76	82.44 ±0.76	**88.24 ±1.72**
PIE	85.45 ±1.02	85.77 ±0.97	85.45 ±1.02	85.77 ±1.02	**93.10 ±0.97**

TABLE 3. The average precision (%) and Std on the PIE, AR and Extended Yale B databases.

Methods	2DPCA	2DPCAL1	N-2DPCA	2DPCA L1-S	Area-2DPCA
Yale B	56.49 ±1.41	57.00 ±1.39	59.49 ±1.41	56.49 ±1.41	**64.36 ±1.38**
AR	83.50 ±1.41	83.52 ±1.39	83.50 ±1.41	83.52 ±1.39	**90.11 ±0.47**
PIE	85.18 ±0.63	85.67 ±0.68	85.18 ±0.63	85.70 ±0.72	**94.87 ±0.52**

TABLE 4. The average reconstruction error ($\times 10^{-2}$) and std of five algorithms on the ORL, COIL20 and LFW databases with black and white dot noise.

Methods	2DPCA	2DPCAL1	N-2DPCA	2DPCA L1-S	Area-2DPCA
COIL20	6.34±0.05	6.33±0.06	6.29±0.06	6.25±0.05	**6.24±0.06**
ORL	3.75±0.05	3.76±0.06	3.64±0.03	3.65±0.03	**3.62±0.02**
LFWCrop	1.78±0.10	1.57±0.03	1.63±0.12	1.56±0.01	**1.54±0.01**

TABLE 5. The average reconstruction error ($\times 10^{-2}$) and std of five algorithms on the AR, ORL, and LFWcrop databases with image noise.

Methods	2DPCA	2DPCAL1	N-2DPCA	2DPCA L1-S	Area-2DPCA
AR	3.02±0.10	3.02±0.12	3.01±0.10	3.01±0.11	**3.00±0.10**
ORL	3.96±0.07	3.87±0.08	3.73±0.04	3.75±0.05	**3.71±0.03**
LFWCrop	1.58±0.01	1.56±0.01	1.54±0.01	1.55±0.01	**1.53±0.01**

TABLE 6. The average running time ($\times 10^{-1}$) and std of each method on the AR, ORL, COIL20 and LFWCrop databases.

Methods	AR	ORL	COIL20	LFWCrop
2DPCA	0.31±0.05	**0.04±0.08**	0.13±0.01	0.58±0.03
2DPCAL1	349.02±37.27	**13.16±0.83**	123.13±7.66	709.63±41.73
N-2DPCA	134.84±20.78	**13.50±2.73**	59.67±18.08	230.52±46.60
2DPCAL1-S	109.62±9.26	**6.30±0.56**	76.88±5.87	274.26±18.50
Area-2DPCA	71.89±0.69	**6.18±0.02**	26.46±0.30	112.21±2.51

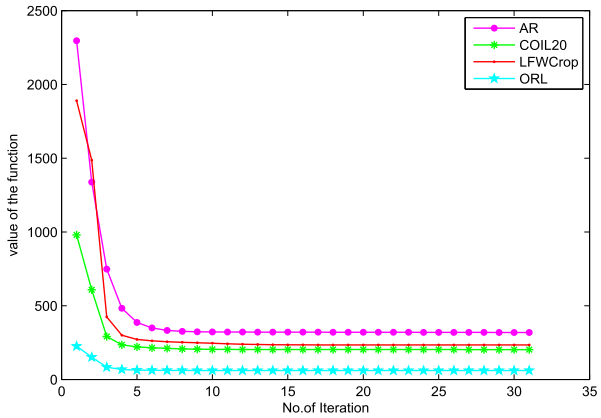

FIGURE 10. Convergence of Area-2DPCA on the AR, COIL20, ORL, and LFWcrop databases.

- Our approach Area-2DPCA is overall superior to robust methods 2DPCAL1, N-2DPCA and 2DPCAL1-S. The reason may be that the solution of our model relates to the weighted covariance matrix, which can characterize the data structure well. Moreover, our model assigns different weighted coefficients, which relate to variance, to each reconstruction error and then get optimal solution by minimizing it, while the other robust methods do not have these advantages.
- Area 2DPCA works better than 2DPCA. It may be that Area-2DPCA uses F-norm to characterize the variance and reconstruction error, which has an inhibitory effect on outliers, compared with squared F-norm that is used in traditional 2DPCA.
- Table 6 illustrates that traditional 2DPCA is the faster among all the five methods. It may be that 2DPCA directly solves the optimal solution without iteration, while other methods do not. Our proposed algorithm is faster than 2DPCAL1, N-2DPCA and 2DPCAL1-S. The reason may be that L2-norm optimization problem is easier to be solved with low computational complexity and our algorithm has good convergence, compared with nuclear-norm and L1-norm based optimization problem. Fig.10 illustrates that our model will converge with a few iteration.

V. CONCLUSIONS

A novel robust 2DPCA objective function for image feature extraction and representation is developed, namely Area-2DPCA. Area-2DPCA uses F-norm to characterize reconstruction error and low-dimensional representation, and its purpose is to seek the projection directions which minimize the total area between projection directions and reconstruct error of data. To find the solution of our model, an iterative algorithm is represented. In each iteration, our algorithm has a closed solution. Compared with 2DPCA and other robust 2DPCA methods, Area 2DPCA not only reduces the impact of outliers in the objective function but also has rotational invariance. Moreover, our model well preserves the geometric structure of data due to the fact that the solution of our model depends on the weighted covariance matrix. The extensive experiments show that our model is more effective and robust.

REFERENCES

[1] Z. Lai, Y. Xu, Q. Chen, J. Yang, and D. Zhang, "Multilinear sparse principal component analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 10, pp. 1942–1950, Oct. 2014.

[2] Z. Fan *et al.*, "Modified principal component analysis: An integration of multiple similarity subspace models," *Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 8, pp. 1538–1552, Aug. 2014.

[3] Y. Xu, Z. Li, J. Yang, and D. Zhang, "A survey of dictionary learning algorithms for face recognition," *IEEE Access*, vol. 5, pp. 8502–8514, 2017.

[4] F. Nie, H. Zhang, R. Zhang, and X. Li, "Robust multiple rank-k bilinear projections for unsupervised learning," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2574–2583, May 2019.

[5] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognit. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[6] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.

[7] Q. Wang, Q. Gao, X. Gao, and F. Nie, "$\ell_{2,p}$-norm based PCA for image recognition," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1336–1346, Mar. 2018.

[8] Q. Gao, F. Gao, H. Zhang, X.-J. Hao, and X. Wang, "Two-dimensional maximum local variation based on image Euclidean distance for face recognition," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3807–3817, Oct. 2013.

[9] Q. Wang and Q. Gao, "Robust 2DPCA and its application," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun./Jul. 2016, pp. 1152–1158.

[10] S. Wang, D. Xie, F. Chen, and Q. Gao, "Dimensionality reduction by LPP-L21," *IET Comput. Vis.*, vol. 12, no. 5, pp. 659–665, 2018.

[11] Y. Liu, Q. Gao, S. Miao, X. Gao, F. Nie, and Y. Li, "A non-greedy algorithm for L1-norm LDA," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 684–695, Feb. 2017.

[12] R. Zhang, F. Nie, and X. Li, "Auto-weighted two-dimensional principal component analysis with robust outliers," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 6065–6069.

[13] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, p. 11, May 2011.

[14] Q. Gao, S. Xu, F. Chen, C. Ding, X. Gao, and Y. Li, "R_1-2-DPCA and Face Recognition," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1212–1223, Apr. 2019.

[15] Y. Chen, Z. Lai, J. Wen, and C. Gao, "Nuclear norm based two-dimensional sparse principal component analysis," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 16, no. 2, 2018, Art. no. 1840002.
[16] Y. Lu, C. Yuan, Z. Lai, X. Li, W. K. Wong, and D. Zhang, "Nuclear norm-based 2DLPP for image classification," *IEEE Trans. Multimedia*, vol. 19, no. 11, pp. 2391–2403, Nov. 2017.
[17] F. Zhang, J. Yang, J. Qian, and Y. Xu, "Nuclear norm-based 2-DPCA for extracting features from images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2247–2260, Oct. 2015.
[18] B.-K. Bao, G. Liu, C. Xu, and S. Yan, "Inductive robust principal component analysis," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3794–3800, Aug. 2012.
[19] Q. Ke and T. Kanade, "Robust L_1 norm factorization in the presence of outliers and missing data by alternative convex programming," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 739–746.
[20] 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.
[21] F. Nie, H. Huang, C. Ding, D. Luo, and H. Wang, "Robust principal component analysis with non-greedy ℓ_1-norm maximization," in *Proc. Int. Joint Conf. Artif. Intell.*, Barcelona, Spain, 2011, pp. 1433–1438.
[22] G.-F. Lu, J. Zou, Y. Wang, and Z. Wang, "L1-norm-based principal component analysis with adaptive regularization," *Pattern Recognit.*, vol. 60, pp. 901–907, Dec. 2016.
[23] X. Li, Y. Pang, and Y. Yuan, "L1-norm-based 2DPCA," *IEEE Trans. Syst., Man, Cybern., B*, vol. 40, no. 4, pp. 1170–1175, Aug. 2010.
[24] R. Wang, F. Nie, X. Yang, F. Gao, and M. Yao, "Robust 2DPCA with non-greedy ℓ1-norm maximization for image analysis," *IEEE Trans. Cybern.*, vol. 45, no. 5, pp. 1108–1112, May 2015.
[25] H. Wang and J. Wang, "2DPCA with L1-norm for simultaneously robust and sparse modelling," *Neural Netw.*, vol. 46, pp. 190–198, Oct. 2013.
[26] S. Liao, Q. Gao, Z. Yang, F. Chen, F. Nie, and J. Han, "Discriminant analysis via joint euler transform and $\ell_{2,1}$-norm," *IEEE Trans. Image Process*, vol. 27, no. 11, pp. 5668–5682, Nov. 2018.
[27] C. Ding, D. Zhou, X. He, and H. Zha, "*R*1-PCA: Rotational invariant L1-norm principal component analysis for robust subspace factorization," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 281–288.
[28] A. Y. Ng, "Feature selection, $L1$ vs. $L2$ regularization, and rotational invariance," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, p. 78.
[29] Y. Liu, Q. Gao, X. Gao, and L. Shao, "L_{21}-norm discriminant manifold learning," *IEEE Access*, vol. 6, pp. 40723–40734, 2018.
[30] T. Li, M. Li, Q. Gao, and D. Xie, "F-norm distance metric based robust 2DPCA and face recognition," *Neural Netw.*, vol. 94, pp. 204–211, Oct. 2017.
[31] Q. Wang, Q. Gao, X. Gao, and F. Nie, "Optimal mean two-dimensional principal component analysis with F-norm minimization," *Pattern Recognit.*, vol. 68, pp. 286–294, Aug. 2017.
[32] Q. Gao, L. Ma, Y. Liu, X. Gao, and F. Nie, "Angle 2DPCA: A new formulation for 2DPCA," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1672–1678, May 2018.
[33] A. M. Martinez and R. Benavente, "The AR face database," Autonomous Univ. Barcelona, CVC, Barcelona, Spain, Tech. Rep. 24, Jun. 1998.
[34] A. S. Georghiades, P. N. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.
[35] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proc. 5th IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2002, pp. 53–58.
[36] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia object image library (coil-100)," Dept. Comput. Sci., Columbia Univ., New York, NY, USA, Tech. Rep. CUCS-005-96, 1996.
[37] B. Anton, J. Fein, T. To, X. Li, L. Silberstein, and C. J. Evans, "Immunohistochemical localization of ORL-1 in the central nervous system of the rat," *J. Comparative Neurol.*, vol. 368, no. 2, pp. 229–251, 1996.
[38] C. Sanderson and B. C. Lovell, "Multi-region probabilistic histograms for robust and scalable identity inference," in *Proc. Int. Conf. Biometrics*. Berlin, Germany: Springer, 2009, pp. 199–208.
[39] G. B. Huang, M. Matter, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, Amherst, MA, USA, ECCV, Tech. Rep. 07-49, Oct. 2007.

YONG WANG received the Ph.D. degree from the Northwestern Polytechnical University, China. He is currently with the Intelligence Department, 10th Research Institute of China Electronic Group Corporation, Chengdu, China. His current research interests include intelligence information systems, image processing, and pattern recognition.

QIN LI received the Ph.D. degree from the Hong Kong polytechnic university, China, in 2010. He is currently a Professional Teacher with the Shenzhen Institute of Information Technology, a Senior Engineer, and a Shenzhen Peacock Scholars. His current research interests include image processing, pattern recognition, and biometrics based on mobile terminals.