IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 18, 2020, accepted November 29, 2020, date of publication December 2, 2020,
date of current version December 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3041847

F-2D-QPCA: A Quaternion Principal Component Analysis Method for Color Face Recognition

MINGHUI WANG[1], LILI SONG[1], KAISONG SUN[1], AND ZHIGANG JIA[2,3]
[1]Department of Mathematics, Qingdao University of Science and Technology, Qingdao 266061, China
[2]School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou 221116, China
[3]Research Institute of Mathematical Science, Jiangsu Normal University, Xuzhou 221116, China

Corresponding author: Zhigang Jia (jiazhigang1980@126.com)

This work was supported by the National Natural Science Foundation of China under Grant 11771188.

ABSTRACT Two-dimensional quaternion principal component analysis (2D-QPCA) is one of the successful dimensionality reduction methods for color face recognition. However, 2D-QPCA is sensitive to outliers. For solving this shortcoming, an efficient robust method(F-2D-QPCA) is presented by means of Frobenius norm(F-norm). The goal of F-2D-QPCA is to find the projection matrix such that the projected data has the maximum variance based on F-norm, and it is more robust to outliers and has higher recognition accuracy than other methods, such as 2D-QPCA, R_1-2-DPCA, F-norm 2DPCA and 2D-PCA, etc. Also, we study in detail a quaternion optimization problem, propose a nongreedy iterative algorithm and prove its convergence. Experiments on several color face databases illustrate the superiority of our proposed method.

INDEX TERMS Color face recognition, quaternion matrix, 2D-QPCA, F-norm 2DPCA, quaternion non-greedy iterative algorithm.

I. INTRODUCTION

Face recognition has always been a focus in recent years. Principal component analysis (PCA) and its various variants have been successfully used for grayscale face recognition [1]–[7]. Based on the Karhunen-Loeve procedure for the characterization of human faces [1], Turk and Pentland [2] presented the eigenface method for face recognition. Early various PCA methods mainly deal with grayscale images by using a vector to represent a grayscale image. As a result, color information and partial spatial information of images are not fully utilized. In order to make full use of the spatial information of images, Yang *et al.* [3] proposed a novel two-dimensional principal component analysis (2D-PCA) by using a matrix to represent a grayscale image. All these methods adopt *squared Euclidean norm* or *squared Frobenius norm* as the distance metric. However, these norms are very sensitive to outliers. Thus, these methods have great shortcomings when processing data sets with outliers [8], [9].

To improve robustness, ℓ_1-norm is adopted because it can suppress the outliers very well [9], [10]. Ke and Kanade [10] used ℓ_1-norm instead of squared Euclidean norm to construct the reconstruction error, and then presented the robust L1-PCA method by minimizing this error. Based on L1-norm maximization, Kwak [11] found the projection vectors, and presented the PCA-L1 method for image representation. For solving PCA-L1, Nie *et al.* [12] presented a nongreedy iterative algorithm. Correspondingly, to exploit spatial structure, many 2-D PCA methods based on ℓ_1-norm have been proposed. Inspired by 2D-PCA, Li *et al.* [13] and Wang *et al.* [14] extended PCA-L1 to 2-DPCA-L1 and gave the greedy algorithm and the nongreedy algorithm, respectively. Pang *et al.* [15] extended PCA-L1 to robust tensor analysis with ℓ_1-norm. By applying sparse constraint on 2-DPCA-L1, Wang and Wang [16] developed 2-DPCAL1-S for simultaneously robust and sparse modelling. Recently, Mi *et al.* presented a novel robust method, called nuclear norm based on PCA (N-PCA) to take full advantage of the structure information of error image [17] and a generalized robust 2-DPCA, which is named as 2-DPCA with $l_{2,p}$-norm minimization ($l_{2,p}$-2- DPCA), for image representation and recognition [18].

However, it is well known that the image covariance matrix characterizes the geometric structure of images, but these methods based on ℓ_1-norm do not involve the image covariance matrix [19], [22], [23]. Moreover, although ℓ_1-norm can suppress the role of outliers, but we do not know that whether ℓ_1-norm can enhance the role of small distance.

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

Thus, to partially overcome this shortcoming, based on F-norm, Li *et al.* [20] proposed a F-norm distance metric based robust 2DPCA (F-norm 2DPCA) for face recognition, and based on R_1-norm, Ding *et al.* [22] developed a rotational invariant ℓ_1-norm (R_1-norm) PCA (R_1-PCA). Motivated by R_1-PCA and 2D-PCA, Gao *et al.* [24] maximized the image covariance with R_1-norm and proposed R_1-2-DPCA method for grayscale face recognition.

None of the above methods use the color information of the image. In [4], Torres *et al.* pointed the importance of color information in face recognition and extended traditional PCA to color face recognition by using the R, G, B color channels, respectively. But, this method do not consider the relationship between three channels. In order to overcome this shortcoming, Yang and Liu [5] used a set of color component combination coefficients to convert three color channels into one channel D by $D = x_1R + x_2G + x_3B$ and presented a general discriminant model for color face recognition, but the optimal coefficients x_1, x_2 and x_3 are difficultly obtained. Xiang *et al.* [6] used a row vector to denote a color channel and then presented a color image as a $3 \times n$ matrix. Then by utilizing both the spatial and color information, they proposed a color 2D-PCA (C2DPCA) method for color face recognition. However, all these methods do not directly use color information.

To directly deal with three channels of color image, the quaternion with zero real part was used to represent the color pixel consisting of three components [25]–[34]. Based on quaternion matrix theory, Jia *et al.* [35] presented the color two-dimensional principal component analysis (2D-QPCA) method for color face recognition. With the aid of two-dimensional quaternion matrices rather than one-dimensional quaternion vectors, 2D-QPCA utilizes the color information and the spatial characteristics simultaneously and mathematically. Recently, Xiao and Zhou [44] proposed a novel quaternion ridge regression (QRR) model for two-dimensional QPCA (QRR-2D-QPCA) and mathematically proved that this QRR model is equivalent to the QCM model of 2D-QPCA.

In this paper, inspired by F-norm 2DPCA method and 2D-QPCA method, we propose the F-2D-QPCA method, which maximizes image covariance based on the quaternion F-norm, and obtains *the eigenface subspace* by a nongreedy algorithm. Compared to most existing 2D-PCA methods and 2D-QPCA method, our method has the following advantages. First, our method treats a color image as a quaternion matrix, which makes full use of the color and spatial information of the image. Second, our method is more robust to outliers because F-norm weakens the role of outliers. Third, our method is based on the image covariance matrix, so it makes good use of the geometric structure of images.

The paper is organized as follows. In Section II, we review quaternion matrices and elaborate the principle of 2D-QPCA method and QSR-2D-QSPCA method for color face recognition. In Section III, we propose a quaternion optimization problem and develop a nongreedy iterative algorithm. And then, we propose a new color two-dimensional quaternion principal component analysis method for color face recognition, which is based F-norm and denoted as F-2D-QPCA method. In Section IV, experiments verify the efficiency of our method. Finally, the conclusion is presented in Section V.

II. PRELIMINARY

In this section, we review the relationship between quaternion matrices and color images, give some properties of quaternion matrices and elaborate the principle of the 2D-QPCA method for color face recongnition.

A. QUATERNION MATICES AND COLOR IMAGES

In 1843, William Rowan Hamilton found the quaternion:

$$q = q_1 + q_2\mathrm{i} + q_3\mathrm{j} + q_4\mathrm{k},$$

where q_1, q_2, q_3, q_4 are real and $\mathrm{i}, \mathrm{j}, \mathrm{k}$ are three imaginary units stasfying

$$\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = \mathrm{ijk} = -1.$$

The set of all quaternions is denoted by $\mathbf{Q}$. The conjugate of q is defined as $q^* = q_1 - q_2\mathrm{i} - q_3\mathrm{j} - q_4\mathrm{k}$ and the modulus $|a|$ is defined as $|a| = \sqrt{aa^*} = \sqrt{a_1^2 + a_2^2 + a_3^2 + a_4^2}$. If the real part is zero, we call $q = r\mathrm{i} + g\mathrm{j} + b\mathrm{k}$ as the pure quaternion, which can represent a pixel of the RGB color space, where R, G, B stand for the values of Red, Green, Blue components, respectively. So, an $m \times n$ color image can be saved as an $m \times n$ pure quaternion matrix $A = (a_{ij})_{m\times n} = R\mathrm{i} + G\mathrm{j} + B\mathrm{k}$ with the nonnegative integer matrix R, G and B.

For $A = (a_{ij}) \in \mathbf{Q}^{m\times n}, a \in \mathbf{Q}^{n\times 1}$, we list several quaternion matrix and vector norms which will be used in this paper [41].

(a). Euclidean norm: $\|a\|_2 = \|a^*\|_2 = \sqrt{a^*a}$, where $(\cdot)^*$ denotes conjugate transpose operation;

(b). Frobenius norm: $\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2} = \sqrt{\mathbf{Tr}(A^*A)}$, where $\mathbf{Tr}(M)$ denotes the trace of M;

(c). 2-norm or spectral norm: $\|A\|_2 = \max_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sqrt{\sigma_1(A)}$, where $\sigma_1(A)$ is the largest singular value of A.

For any $A = A_1 + A_2\mathrm{i} + A_3\mathrm{j} + A_4\mathrm{k} \in \mathbf{Q}^{m\times n}$, $A_l \in \mathbf{R}^{m\times n} (l = 1, 2, 3, 4)$, define

$$A^{\mathcal{R}} \equiv \begin{pmatrix} A_1 & -A_2 & -A_3 & -A_4 \\ A_2 & A_1 & -A_4 & A_3 \\ A_3 & A_4 & A_1 & -A_2 \\ A_4 & -A_3 & A_2 & A_1 \end{pmatrix} \in \mathbf{R}^{4m\times 4n}.$$

The real matrix $A^{\mathcal{R}}$ is known as *real representation* of the quaternion matrix A.

For $A, B \in \mathbf{Q}^{m\times n}$, $C \in \mathbf{Q}^{n\times s}$, the following properties are well known [21].

(a). $(A + B)^{\mathcal{R}} = A^{\mathcal{R}} + B^{\mathcal{R}}$, $(AC)^{\mathcal{R}} = A^{\mathcal{R}}C^{\mathcal{R}}$;

(b). $(A^*)^{\mathcal{R}} = (A^{\mathcal{R}})^T$, where $(\cdot)^T$ denotes transpose operation;

(c). A is a column unitary matrix if and only if $A^{\mathcal{R}}$ is a column orthogonal matrix.

Inspired by [22], we present a new quaternion matrix norm.

Definition 1: For $A \in \mathbf{Q}^{m\times n}$, its R_1-norm is

$$\|A\|_{R_1} = \sum_{i=1}^{m} \|A(i,:)\|_2,$$

where $A(i,:)$ denotes the ith row of A.

Let $\mathbf{Re}(A)$ denote the real part of A. After the simple derivation, we can get the following result.

Theorem 1: For $A \in \mathbf{Q}^{m\times n}, B \in \mathbf{Q}^{n\times m}, C \in \mathbf{Q}^{n\times n}$, we have

(1). $\mathbf{Re}(\mathbf{Tr}(AB)) = \mathbf{Re}(\mathbf{Tr}(BA))$, $\mathbf{Tr}(AB) \neq \mathbf{Tr}(BA)$;
(2). $\mathbf{Tr}(AA^*) = \mathbf{Tr}(A^*A) = \|A\|_F^2$;
(3). $\mathbf{Re}(\mathbf{Tr}(AB)) \leq \|A\|_F\|B\|_F$(Cauchy-Schwarz inequality);
(4). $\|A\|_F \leq \|A\|_{R_1}$;
(5). $\|A\|_F = \frac{1}{2}\|A^{\mathcal{R}}\|_F$;
(6). $\mathbf{Re}(\mathbf{Tr}(C)) = \frac{1}{4}\mathbf{Tr}(C^{\mathcal{R}})$.

B. 2D-QPCA

Let $V = (v_1, v_2, \cdots, v_k)$ denote an $n \times k$ quaternion matrix with unitary column vectors. Our idea is to project $m \times n$ color image A onto V [42], [43]:

$$B = AV, \tag{1}$$

where $m \times k$ projected matrix B is called the projected feature image of image A. A good projection matrix V can be determined by the total scatter of the projected samples. That is, the following function is adopted:

$$J(V) = \mathbf{Tr}(G_V), \tag{2}$$

where the covariance matrix G_V of the projected feature images of the training samples can be denoted by

$$\begin{aligned} G_V &= \mathbf{E}[(B - \mathbf{E}B)(B - \mathbf{E}B)^*] \\ &= \mathbf{E}[(AV - \mathbf{E}(AV))(AV - \mathbf{E}(AV))^*] \\ &= \mathbf{E}[(A - \mathbf{E}A)V((A - \mathbf{E}A)V)^*] \\ &= \mathbf{E}[(A - \mathbf{E}A)VV^*(A - \mathbf{E}A)^*], \end{aligned}$$

and $\mathbf{E}(\cdot)$ denotes the mathematical expectation. The physical significance of maximizing (2) is to find projection directions $v_1, v_2, \cdots, v_k$ such that the total scatter of the resulting projected samples is maximized.

Because

$$\mathbf{Tr}(G_V) = \mathbf{Tr}[V^*\mathbf{E}((A - \mathbf{E}A)^*(A - \mathbf{E}A))V],$$

we can define the color image covariance matrix (QCM)

$$G = \mathbf{E}((A - \mathbf{E}A)^*(A - \mathbf{E}A)),$$

which is an $m \times m$ nonnegative definite matrix and can be evaluated directly using the training image samples.

Let $\{A_i^{(j)} \in \mathbf{Q}^{m\times n}\}_{i=1}^{l_j}$ denote the set of the training color image samples, where class index $j = 1, 2, \cdots, M$.

We compute the average image $\bar{A}$ and the color image covariance (scatter) matrix (QCM) G of training samples by

$$\bar{A} = \frac{1}{N}\sum_{j=1}^{M}\sum_{i=1}^{l_j} A_i^{(j)} \in \mathbf{Q}^{m\times n}, \tag{3}$$

and

$$G = \frac{1}{N}\sum_{j=1}^{M}\sum_{i=1}^{l_j}(A_i^{(j)} - \bar{A})^*(A_i^{(j)} - \bar{A}) \in \mathbf{Q}^{n\times n}. \tag{4}$$

The aim of 2D-QPCA is to find a set of unitary projection basis vectors $v_1, \cdots, v_k$, where $\hat{V} = \mathbf{Span}(v_1, \cdots, v_k)$ is often called *the eigenface subspace or the projection matrix*, such that, when projected onto $\hat{V}$, the projected sample

$$P_s = (A_s - \bar{A})\hat{V} \tag{5}$$

of A_s has the maximal scatter. $\hat{V}$, which maximizes the trace of the generalized total scatter criterion V^*GV, meets this requirement. In other words, $\hat{V}$ is the solution of the following problem:

$$\begin{aligned} &\max_{V^*V=I} \mathbf{Tr}\sum_{j=1}^{M}\sum_{i=1}^{l_j} V^*(A_i^{(j)} - \bar{A})^*(A_i^{(j)} - \bar{A})V \\ &= \max_{V^*V=I} \mathbf{Tr}\sum_{j=1}^{M}\sum_{i=1}^{l_j} V^*(\hat{A}_i^{(j)})^*\hat{A}_i^{(j)}V \\ &= \max_{V^*V=I} \sum_{j=1}^{M}\sum_{i=1}^{l_j} \|\hat{A}_i^{(j)}V\|_F^2, \end{aligned} \tag{6}$$

where $\hat{A}_i^{(j)} = A_i^{(j)} - \bar{A}$.

In the 2D-QPCA method [35], the columns $v_1, \cdots, v_k$ of $\hat{V}$ are the eigenvectors (called *eigenfaces*) of G corresponding to the first k largest eigenvalues.

The procedure of 2D-QPCA [35] is given in Table 1.

C. QSR-2D-QSPCA

In this subsection we recall the QSR model for 2D-QSPCA from [44].

Let $\{X_i \in \mathbf{Q}^{m\times n}\}_{i=1}^{h}$ and $\{Y_i \in \mathbf{Q}^{k\times n}\}_{i=1}^{h}$ be the set of 2D quaternion samples and the set of projected quaternion samples, respectively. Here all quaternion samples are mean-centered, i.e., $\mathbf{E}(X_i) = 0$. As mentioned in the previous section, the objective of 2D-QPCA is to find an orthonormal quaternion basis $V = (v_1, \cdots, v_k)$ so that the projected quaternion samples have the largest scatter after projection. What we should note here is that the QCM model for 2D-QPCA in [44] is working in the column direction. That is $Y_i = V^*X_i$ is defined as the projected sample of X_i. Under the constraint of least-squares error, maximizing the scatter of projected quaternion samples is equivalent to minimizing the reconstruction error between projected quaternion samples and the input quaternion samples. Hence, the solution of

TABLE 1. 2D-QPCA algorithm.

Input: Training set $\{A_i^{(j)} \in \mathbf{Q}^{m\times n}\}_{i=1}^{l_j}$ (the set of the training color image samples with $j = 1, 2, \cdots, M$ and $N = l_1 + l_2 + \cdots + l_M$). Test image A for recognition.
Output: K such that A belongs to class K.

% Compute the eigenface subspace $\hat{V}$
(i) Compute the average image $\bar{A}$ and the quaternion covariance matrix G of the training set by (3) and (4).
(ii) Perform QED for the eigenface subspace: $V^*GV = D$ with $V^*V = I, D = \mathbf{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$.
(iii) For given $r(1 \le r \le n)$, take the eigenface subspace $\hat{V} = V(:, 1:r)$.
% Compute the feature matrix or feature image $P_i^{(j)}$ of the sample image $A_i^{(j)}$
(iv) Compute the feature matrix $P_i^{(j)} = (A_i^{(j)} - \bar{A})\hat{V} \in \mathbf{Q}^{n\times r}$ for $j = 1, \cdots, M, i = 1, \cdots, l_j$.
% Use a nearest neighbor classifier for color face recognition
(v) Compute the feature matrix $P = (A - \bar{A})\hat{V}$ for the test image A.
(vi) Find the nearest feature matrix $P_s^{(K)}$ satisfying $\|P - P_s^{(K)}\|_F = \min_{j,i} \|P - P_i^{(j)}\|_F$.
(vii) Output K.

TABLE 2. QSR-2D-QPCA algorithm.

Input: Training set $\{X_i \in \mathbf{Q}^{m\times n}\}_{i=1}^{h}$, the dimension k, and parameters $\lambda_2, \lambda_{1,k}, j = 1, \cdots, k, \varepsilon^{outer}$.
Output: Optimal quaternion saprse projection basis $(\dot{v}_{s1}, \cdots, \dot{v}_{sk})$.

Rewrite 2D-QSPCA into its complex form.
Convert the task to find $\alpha = (\alpha_1, \cdots, \alpha_{2k})$ and $\beta = (\beta_1, \cdots, \beta_{2k})$, where $\alpha = \chi_A, \beta = \chi_B$ are the complex adjoint form of A and B.
Initialize α to the complex adjoint form of the solution of 2D-QPCA
repeat
  (1) Fixing α, find optimal β.
  for $j = 1, \cdots, k$, **do**
    Compute β_j using Algorithm 2 in [44].
    Compute residual $\xi(\beta_j)$.
  end
  for $j = k+1, \cdots, 2k$, **do**
    Derive β_j from β_{j-k}.
  end
  (2) Fixing β, find optimal α. Let $\varphi\beta = U\Sigma V^*$, set $\hat{\alpha} = UV^*$, where $\varphi = \sum_{i=1}^{h} \chi_{X_i}\chi_{X_i^*}$.
until $\xi(\beta_j) < \varepsilon^{outer}$ for $j = 1, 2, \cdots, k$;
for $j = 1, \cdots, k$, **do**
  Recover v_{sj} from β_j using $v_{sj} = \gamma(\frac{\beta_j}{\|\beta_j\|_2})$, where $\gamma(\cdot)$ is a recover operator in Definition 5 in [44].
end
Output $(v_{s1}, \cdots, v_{sk})$.

QCM model [44] is equals to the solution of the following problem

$$B = \arg\min_{B}\Big(\sum_{i=1}^{h} \|X_i - BB^*X_i\|_F^2\Big), \quad s.t. \quad B^*B = I_k.$$

Based on this observation, the quaternion ridge regression (QRR) model has been proposed. Taking the advantages of sparse regularization, the QSR model for 2D-QSPCA has been further advanced by regularizing the QRR model with the l_1-norm penalties. Theorem 2 presents the QSR model for 2D-QSPCA.

Theorem 2: Let $\{X_i \in \mathbf{Q}^{m\times n}\}_{i=1}^{h}$ be a set of 2D quaternion samples and the columns of $V_s = (v_{s1}, \cdots, v_{sk})$ be the quaternion sparse basis vectors of 2D-QSPCA. V_s can be obtained as follows.

For any $\lambda_2 \ge 0$ and $\lambda_{1,j}, j \ge 0, j = 1, \cdots, k$, suppose that $A = (a_1, \cdots, a_k) \in \mathbf{Q}^{m\times k}$ and $B = (b_1, \cdots, b_k) \in \mathbf{Q}^{m\times k}$ satisfy

$$(A, B) = \arg\min_{A,B}\left(\sum_{i=1}^{h} \|X_i - AB^*X_i\|_F^2 + \lambda_2\|B\|_F^2 + \sum_{j=1}^{k} \lambda_{1,j}\|b_j\|_1\right) \quad (7)$$
$$\text{s.t. } A^*A = I_k$$

Then, $v_{sj} = (b_j/\|b_j\|_2), j = 1, 2, \cdots, k$.

The penalty term $\lambda_2\|B\|_F^2$ is used to avoid the potential colinearity problem when the number of input samples is much smaller than the dimension of input samples. In the QSR model, it is obvious that if $\lambda_{1,j} = 0$, for $j = 1, 2, \cdots, k$, equation (7) reduces to the QRR model [44]:

$$(A, B) = \arg\min_{A,B}\left(\sum_{i=1}^{h} \|X_i - AB^*X_i\|_F^2 + \lambda_2\|B\|_F^2\right)$$
$$\text{s.t. } A^*A = I_k.$$

that is, the obtained basis vectors $v_j = (b_j/\|b_j\|_2), j = 1, 2, \cdots, k$ are not sparse. Xiao *et al.* still mathematically prove that the QRR model is equivalent to the QCM model of 2D-QPCA.

In [44], an alternating minimization algorithm was developed to iteratively compute the solution of QSR model in the equivalent complex domain. The procedure is given in Table 2.

III. F-2D-QPCA

In this section we propose a new color two-dimensional quaternion principal component analysis method for color face recognition, which is based F-norm and denoted as F-2D-QPCA method. For this, we need to solve a quaternion optimization problem.

A. A QUATERNION OPTIMIZATION PROBLEM

In this subsection, we propose a quaternion optimization problem and develop a nongreedy iterative algorithm, which has not only a closed-form solution in each iteration but also a good convergence.

Given $X_i \in \mathbf{Q}^{m\times n} (i = 1, 2, \cdots, N)$, we discuss the optimization problem

$$\max_{\substack{W\in\mathbf{Q}^{n\times d}\\ W^*W=I}} \sum_{i=1}^{N} \|X_iW\|_F. \quad (8)$$

Theorem 3: Suppose that $H \in \mathbf{Q}^{d\times n}(d \le n)$ and its singular value decomposition (SVD) is $H = U\wedge V^*$ with $U \in \mathbf{Q}^{d\times d}$, $U^*U = UU^* = I_d$, $V \in \mathbf{Q}^{n\times n}$, $V^*V = VV^* = I_n$ and $\wedge = [\mathbf{diag}(\sigma_1, \cdots, \sigma_d), 0_{d\times(n-d)}] \in \mathbf{R}^{d\times n}$, where I and $\mathbf{diag}(\sigma_1, \cdots, \sigma_d)$ denote the identity matrix and the diagonal matrix with $\sigma_1, \cdots, \sigma_d$ as diagonal elements, respectively. Then $W = VI_{n\times d}U^*$ is the solution of the optimization problem

$$\max_{\substack{W\in\mathbf{Q}^{n\times d}\\ W^*W=I_d}} \mathbf{Re}(\mathbf{Tr}(HW)), \tag{9}$$

with

$$I_{n\times d} = \begin{pmatrix} I_d \\ 0_{(n-d)\times d} \end{pmatrix}.$$

Proof: First, we have

$$\begin{aligned}
&\mathbf{Re}(\mathbf{Tr}(HW)) \\
&\quad = \mathbf{Re}(\mathbf{Tr}(U \wedge V^*W)) = \mathbf{Re}(\mathbf{Tr}(\wedge V^*WU)) \\
&\quad = \mathbf{Re}(\mathbf{Tr}(\wedge Q)) = \mathbf{Re}(\sum_{k=1}^{d} \sigma_k Q(k,k)) \\
&\quad = \sum_{k=1}^{d} \sigma_k \mathbf{Re}(Q(k,k)),
\end{aligned}$$

where $Q = V^*WU \in \mathbf{Q}^{n\times d}$, $Q^*Q = I_d$.

Since Q is column unitary orthogonal, we can obtain that

$$\sum_{k=1}^{d} \sigma_k \mathbf{Re}(Q(k,k)) \le \sum_{k=1}^{d} \sigma_k$$

and the equality holds only and only if $Q(k,k) = 1$ for all $k = 1, 2, \cdots, d$, that is, $Q = I_{n\times d}$ and $W = VI_{n\times d}U^*$. □

Now we consider how to find the optimal solution of (8). First, we have

$$\begin{aligned}
\sum_{i=1}^{N} \|X_iW\|_F &= \sum_{i=1}^{N} \frac{\|X_iW\|_F^2}{\|X_iW\|_F} \\
&= \sum_{i=1}^{N} \mathbf{Tr}\left(\frac{(X_iW)^*X_iW}{\|X_iW\|_F}\right) \\
&= \mathbf{Tr}(HW) \xrightarrow{\mathbf{Tr}(HW)\text{ is real}} \mathbf{Re}(\mathbf{Tr}(HW)),
\end{aligned} \tag{10}$$

where

$$H = \sum_{i=1}^{N} \beta_i^* X_i, \tag{11}$$

$$\beta_{ik} = \begin{cases} \dfrac{X_iW}{\|X_iW\|_F}, & \|X_iW\|_F \ne 0; \\ 0, & \|X_iW\|_F = 0. \end{cases} \tag{12}$$

In summary, the optimization problem (8) finally becomes the optimization problem

$$\max_{\substack{W\in\mathbf{Q}^{n\times d}\\ W^*W=I_d}} \mathbf{Re}(\mathbf{Tr}(HW)), \tag{13}$$

where H is the function of W. To solve this problem, we present a nongreedy iterative algorithm(See Table 3).

For QPCA-F algorithm, we have the following results.

Theorem 4: For the sequence $\{W^{(t)}\}_{t=1}^{\infty}$ obtained by QPCA-F algorithm, we have

$$\sum_{i=1}^{N} \|X_iW^{(t+1)}\|_F \ge \sum_{i=1}^{N} \|X_iW^{(t)}\|_F.$$

That is, the sequence $\{\sum_{i=1}^{N} \|X_iW^{(t)}\|_F\}_{t=1}^{\infty}$ monotonically increases.

Proof: From Theorem 3, we have

$$\mathbf{Re}(\mathbf{Tr}(H^{(t)}W^{(t+1)})) \ge \mathbf{Re}(\mathbf{Tr}(H^{(t)}W^{(t)})). \tag{14}$$

Then from (10)-(12), (14) is equivalent to

$$\begin{aligned}
&\mathbf{Re}\left(\mathbf{Tr}\left(\sum_{i=1}^{N}(\beta_i^{(t)})^*X_iW^{(t+1)}\right)\right) \\
&\qquad \ge \mathbf{Re}\left(\mathbf{Tr}\left(\sum_{i=1}^{N}(\beta_i^{(t)})^*X_iW^{(t)}\right)\right).
\end{aligned} \tag{15}$$

According to Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\|X_iW^{(t)}\|_F\|X_iW^{(t+1)}\|_F \\
&\qquad \ge \mathbf{Re}\left(\mathbf{Tr}\left((X_iW^{(t)})^*X_iW^{(t+1)}\right)\right),
\end{aligned} \tag{16}$$

i.e.,

$$\begin{aligned}
&\|X_i(k,:)W^{(t+1)}\|_F \\
&\quad \ge \mathbf{Re}\left(\mathbf{Tr}\left(\frac{(X_iW^{(t)})^*X_iW^{(t+1)}}{\|X_iW^{(t)}\|_F}\right)\right) \\
&\quad = \mathbf{Re}\left(\mathbf{Tr}\left((\beta_i^{(t)})^*X_iW^{(t+1)}\right)\right).
\end{aligned} \tag{17}$$

From (15),(17) and (10), we have

$$\begin{aligned}
\sum_{i=1}^{N} \|X_iW^{(t+1)}\|_F &\ge \sum_{i=1}^{N} \mathbf{Re}\left(\mathbf{Tr}\left((\beta_i^{(t)})^*X_iW^{(t+1)}\right)\right) \\
&\ge \mathbf{Re}\left(\mathbf{Tr}\left(\sum_{i=1}^{N}(\beta_i^{(t)})^*X_iW^{(t)}\right)\right) \\
&= \sum_{i=1}^{N} \|X_iW^{(t)}\|_F. \square
\end{aligned}$$

Theorem 5: The sequence $\{\sum_{i=1}^{N} \|X_iW^{(t)}\|_F\}_{t=1}^{\infty}$ is upper bounded.

Proof: For all t, because $(W^{(t)})^*W^{(t)} = I_d$, we have $\|W^{(t)}\|_F = \sqrt{d}$ and

$$\sum_{i=1}^{N} \|X_iW^{(t)}\|_F \le \sum_{i=1}^{N} \|X_i\|_F\|W^{(t)}\|_F = \sqrt{d}\sum_{i=1}^{N} \|X_i\|_F. \square$$

Because a monotone bounded sequence must have the limit, we can know that the sequence $\{\sum_{i=1}^{N} \|X_iW^{(t)}\|_F\}_{t=1}^{\infty}$ has

TABLE 3. QPCA-F algorithm.

Input: $X_i \in \mathbf{Q}^{m\times n}(i=1,2,\cdots,N)$, dimension $d(d\le n)$, accuracy ϵ, initial $t=1$ and $W^{(t)}\in\mathbf{Q}^{n\times d}$ which satisfies $(W^{(t)})^*W^{(t)}=I$.
Output: W.

For $t=1,2,\cdots$ until convergence or
$$\frac{\left|\sum_{i=1}^{N}\|X_iW^{(t+1)}\|_F-\sum_{i=1}^{N}\|X_iW^{(t)}\|_F\right|}{\sum_{i=1}^{N}\|X_iW^{(t)}\|_F}<\epsilon$$
(i)Compute $H^{(t)}\in\mathbf{Q}^{d\times n}$ according to (12) and (11);
(ii)Compute the SVD of $H^{(t)}$: $H^{(t)}=U\wedge V^*$;
(iii)Update $W^{(t)}$ by $W^{(t+1)}=VI_{n\times d}U^*$;
(iv)$t\leftarrow t+1$.
end
Output $W\leftarrow W^{(t+1)}$.

the limitation $\sum_{i=1}^{N}\|X_i\hat{W}\|_F$. But is $\hat{W}$ the solution of (8)? Due to the imperfect theory of quaternion matrix calculus, we cannot discuss this problem on the quaternion ring. Next, we transform this problem into a problem in the real field through the real representation of the quaternion matrix.

Because

$$\mathbf{Re}(\mathbf{Tr}(HW))=\frac{1}{4}\mathbf{Tr}(H^{\mathcal{R}}W^{\mathcal{R}})=\mathbf{Tr}(\hat{H}\hat{W}),$$

where $\hat{H}\in\mathbf{R}^{d\times 4n}$ is the first d rows of $H^{\mathcal{R}}$ and $\hat{W}\in\mathbf{R}^{4n\times d}$ is the first d columns of $W^{\mathcal{R}}$, the optimization problem (13) finally becomes the optimization problem

$$\max_{\substack{\hat{W}\in\mathbf{R}^{4n\times d}\\ \hat{W}^*\hat{W}=I_d}}\mathbf{Tr}(\hat{H}\hat{W}). \tag{18}$$

$$\sum_{i=1}^{N}\|X_iW\|_F=\frac{1}{2}\sum_{i=1}^{N}\|X_i^{\mathcal{R}}W^{\mathcal{R}}\|_F=\sum_{i=1}^{N}\|X_i^{\mathcal{R}}\hat{W}\|_F$$

tells us that the optimization problem (8) is equivalent to the optimization problem

$$\max_{\substack{\hat{W}\in\mathbf{R}^{4n\times d}\\ \hat{W}^*\hat{W}=I}}\sum_{i=1}^{N}\|X_i^{\mathcal{R}}\hat{W}\|_F. \tag{19}$$

From (18), (19) and Theorem 2 of [23], we can obtain the following result.

Theorem 6: The sequence $\{\sum_{i=1}^{N}\|X_iW^{(t)}\|_F\}_{t=1}^{\infty}$ has the limitation $\sum_{i=1}^{N}\|X_i\hat{W}\|_F$, where $\hat{W}$ is a local solution of (8).

B. F-2D-QPCA

In this subsection, based on QPCA-F algorithm proposed in the previous subsection, we propose a new color two-dimensional quaternion principal component analysis method for color face recognition.

F-norm is the unitary invariant norm and can retain traditional 2D-QPCA's nice properties such as geometric structure and rotational invariance. Moreover, compared to squared F-norm and R_1- norm, F-norm weakens the large distance, and compared to squared Frobenius norm, F-norm enhances the role of small distance. Thus, to improve robustness of the 2D-QPCA method and the accuracy of the R_1-2-DPCA method, we employ F-norm instead of squared Euclidean distance and R_1- norm as the distance metric in the model (6) and obtain the following objective function:

$$\max_{V^*V=I}\sum_{j=1}^{M}\sum_{i=1}^{l_j}\|\hat{A}_i^{(j)}V\|_F. \tag{20}$$

Then, we can get the following the F-2D-QPCA algorithm(See Table 4).

TABLE 4. F-2D-QPCA algorithm.

Input: Training set $\{A_i^{(j)}\in\mathbf{Q}^{m\times n}\}_{i=1}^{l_j}$ (the set of the training color image samples with $j=1,2,\cdots,M$ and $N=l_1+l_2+\cdots+l_M$). Test image A for recognition. Dimension $d(d\le n)$, accuracy $\epsilon=10^{-4}$, initial $t=1$ and $V^{(t)}=I_{n\times d}$.
Output: K such that A belongs to class K.

(i)Compute the mean image of the training color image samples from the j-th class by $\bar{A}_j=\frac{1}{l_j}\sum_{i=1}^{l_j}A_i^{(j)}\in\mathbf{Q}^{m\times n}$ and let $\hat{A}_i^{(j)}=A_i^{(j)}-\bar{A}_j$.
(ii) Perform QPCA-F algorithm for the eigenface subspace V.
(iii) Compute the feature matrix $P_i^{(j)}=\hat{A}_i^{(j)}V$ for $j=1,\cdots,M,i=1,\cdots,l_j$.
(iv) Compute the feature matrix $P=(A-\bar{A})V$ for the test image A.
(v) Find the nearest feature matrix $P_s^{(K)}$ satisfying $\|P-P_s^{(K)}\|_F=\min_{i,j}\|P-P_i^{(j)}\|_F$.
(ix) Output K.

IV. EXPERIMENTS

In this section, we test F-2D-QPCA method by the famous Georgia Tech face(GT) database[1] and the color Face Recognition Technology database (FT),[2] and compare with 2D-QPCA method, R_1-2D-PCA method, F-norm 2DPCA method, 2D-PCA method and QRR-2D-QPCA method.

All experiments in this section are performed on a personal computer with 3.2 GHz Intel Core i5-6500 and 16 GB 2400 MHz DDR4 using MATLAB-R2018b and Quaternion toolbox for Matlab(QTFM 2.6) [45]. It is worth pointing out that Jia *et al.* [37]–[39] have developed a quaternion calculation toolbox based on the real number field, which has achieved higher accuracy and calculation efficiency.

The GT database are composed of color images of 50 individuals with 15 views per individual, and with no specific order in their viewing direction.

All images in the Georgia Tech face database are manually cropped, and then resized to 44 × 33 pixels. There are 50 persons to be used.

[1]The Georgia Tech face database. http://www.anefian.com/research/facereco.htm
[2]The color FERET database. https://www.nist.gov/itl/iad/image-group/color-feret-database.

FIGURE 1. First row is some images in the GT database. The second row is the corresponding noised images.

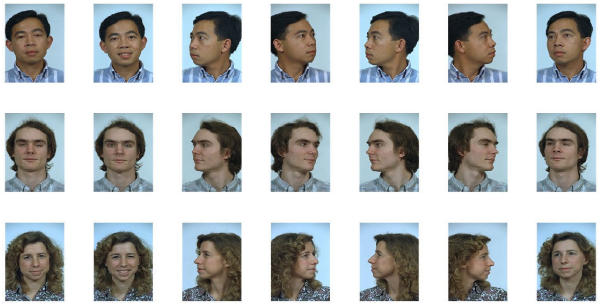


FIGURE 2. Sample images for three persons of the color FT face database.

We randomly select 4 images per person and then place salt and pepper noise. The noise is random distribution and accounted for 0.0172 to 0.1550 of the image area (see Figure 1). Thus, we get a new gallery for the experiments, recorded as GT-noise. In this new dataset, The first x face images per individual person are chosen for training and the remaining five face images are used for testing. The number of chosen eigenfaces or projection vectors is recorded as r. Our approach and the aforementioned three approaches are employed to extract low-dimensional representations, respectively. This process is repeated 5 times.

Also, we randomly select 3 images per person, place salt and pepper noise to the entire area, and get a new gallery for the experiments, recorded as GT-noise1.

The FT database contains 14126 color face images of 1199 individuals. The minimal number of face images for one person is 6, and the maximal one is 44. The size of each cropped color face image is 192 × 128 pixels. We choose 219 persons with 10 views per individual as samples. Some samples are shown in Figure 2.

To further validate the robustness of R1-2-DPCA, we randomly choose three images from each class (person) in the GT database and add a 22 × 16 pixels object image (outlier) in the chosen images. Combining the remaining images, we get a new gallery for the experiments, recorded as GT-outlier. Figure 3 shows some images of this new gallery. Also, we randomly choose two images from each class (person) in the FT database and add a 48 × 32 pixels object image (outlier) in the chosen images. Combining the remaining images, we get a new gallery for the experiments, recorded as FT-outlier. In the experiments, 21 images per person, which include 16 noise-free images and five noised images, are randomly chosen for training, and the remaining images are used as probe images. We employ four approaches

FIGURE 3. First row is some face images in the GT database. The second row is the corresponding face images + object image (outliers).

TABLE 5. Average classification accuracy on the GT, GT-noise, GT-outlier, FT, and FT-outlier databases.

	$\frac{GT}{(x=10,r=2)}$	$\frac{GT-noise}{(x=10,r=2)}$	$\frac{GT-outlier}{(x=10,r=2)}$	$\frac{FT}{(x=7,r=1)}$	$\frac{FT-outlier}{(x=7,r=1)}$
2D-PCA	0.8320	0.7892	0.6240	0.7542	0.5012
R_1-2-DPCA	0.8240	0.7840	0.6600	0.7610	0.5403
2D-QPCA	0.8520	0.8200	0.6600	0.8326	0.6149
F-norm 2DPCA	0.8320	0.7904	0.6400	0.7641	0.5419
F-2D-QPCA	0.8560	0.8232	0.6720	0.8387	0.6219

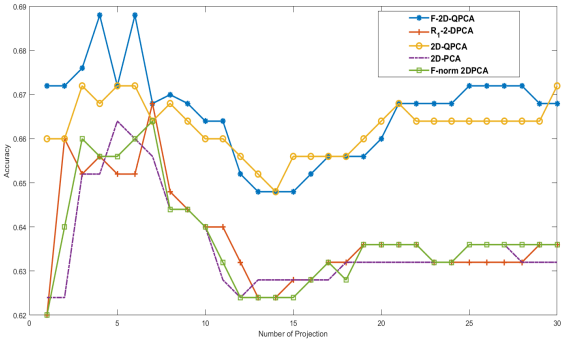


FIGURE 4. Accuracy versus number of projection vectors on the GT-outlier database with $x = 10$.

to extract low-dimensional representations of images, respectively. We do it five times to evaluate performance of each method.

Table 4 shows the average recognition accuracy of each approach on the the GT, GT-noise, GT-outlier, FT, and FT-outlier databases.

Figure 4 plots the average classification curve versus the number projection vectors on the GT-outlier database. For these four methods, a large number of experiments on the above five databases show that the best recognition accuracy generally appears in $r \leq 6$. Figure 5 plots the average classification curve versus the number projection vectors on the GT-noise1 database.

Figure 6 plots the convergence curve of our method on the GT and FT databases, respectively, which shows that our algorithm has good convergence. For the computational complexity, compared with 2D-QPCA, F-2D-QPCA has one more singular value decomposition of quaternion matrix and the product of two quaternion matrices, and compared with F-norm 2DPCA, the computation of F-2D-QPCA is not more than four times higher because F-2D-QPCA processes color images, and F-norm 2DPCA processes grayscale images.

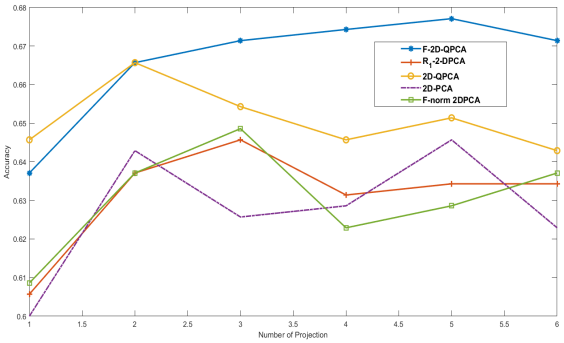


FIGURE 5. Accuracy versus number of projection vectors on the GT-noise1 database with $x = 8$.

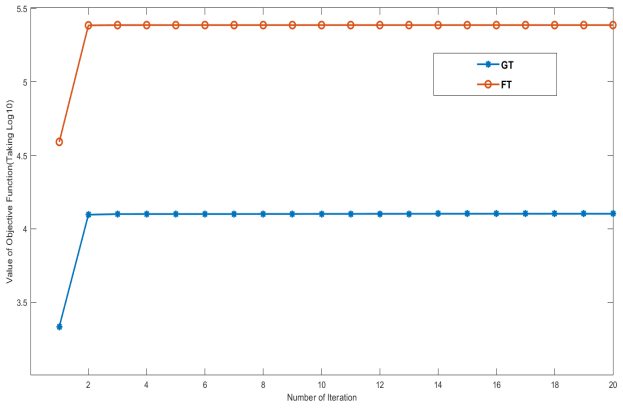


FIGURE 6. Convergence curve of our method on two databases.

From the above experimental results, we can obtain the following conclusions.

1) F-2D-QPCA and 2D-QPCA are superior to and R_1-2-DPCA and 2D-PCA. Because the first two methods use the color information of the image, the last two methods only deal with the grayscale image.

2) F-2D-QPCA is slightly superior to 2D-QPCA on four databases and better than 2D-QPCA on GT-noise1 and GT-outlier database. This is probably because that 2D-QPCA is sensitive to the small variation due to the illumination, pose and occlusion. It results in unstable representation for face images.

Compared with 2D-QPCA, our approach is slightly more accurate. But in [20], for grayscale images, F-norm 2DPCA is superior to 2D-PCA on the modified Extended Yale B, AR and CMU PIE databases.

At the end of this section, we test QRR-2D-QPCA [44] and F-2D-QPCA on AR face database,[3] which contains color face images of 126 individuals recorded in two sessions. In each session, a neutral color face image is recorded sequentially by images with different expressions and illumination, and images occluded by sunglasses and scarf (three images

[3]A. M. Martinez, "The AR face database," CVC, New Delhi, India, Tech. Rep. 24, 1998.

FIGURE 7. Sample images of one person from AR face database.

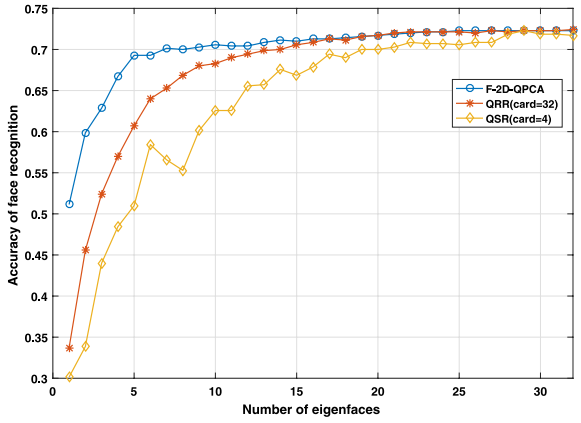


FIGURE 8. Accuracy versus number of projection vectors on AR face database.

per condition). We use a popular subset of AR containing 100 individuals with 26 views per individual. Some samples are shown in Figure 7. We use seven nonoccluded color face images in session one for training and the corresponding seven nonoccluded images in session two for testing. All face images are resized to 32×32 pixels.

In our experiments, we compare the performance of F-2D-QPCA with QSR-2D-QSPCA. We refer to [44] and specify the parameter setting in the QSR model. Firstly, λ_2 is set to 0.001. As is shown in Table 2, the sparsity of the basis of QSR model is controlled via the parameter $\lambda_{1,j}$. According to [44], we do not explicitly preassign the value of $\lambda_{1,j}$. Instead, we specify the cardinality (the number of nonzero elements, denoted by card) of the basis. Card is set to 4 and 32 in our test. Obviously, QSR model reduces to the QRR model when card $= 32$.

The number of the chosen eigenfaces are from 1 to 32. The face recognition rates of two methods are shown in Figure 8. For the three methods, the projected testing samples are all classified based on the nearest-neighbor classifier using F-norm distance. It is obvious that the difference between F-2D-QPCA and QSR-2D-QSPCA is that the methods of computing the eigenface subspace V. Table 6 presents the recognition rates and time of calculating the eigenfaces for cases that the number of features are chosen as 10, 20 and 30. From the results we can see that F-2D-QPCA reaches the highest face recognition rate, and costs less time than QSR and QRR models. The reason is that the QSR model needs to be solved iteratively, which will take a lot of time.

F-2D-QPCA has the higher recognition rate than QRR-2D-QPCA. Moreover, due to the determination of many

TABLE 6. Face recognition rate(RR) and time(second) for computing the eigenfaces.

	x	10	20	30
F-2D-QPCA	RR	70.57%	71.71%	72.29%
	time	3.41	4.01	4.27
QRR(card=32)	RR	68.29%	71.71%	72.29%
	time	10.20	14.77	14.88
QSR(card=4)	RR	62.57%	70.00%	71.86%
	time	355.58	295.53	312.45

parameters, the complexity of the latter is much higher than that of the former.

V. CONCLUSION

In this paper, a robust quaternion-matrix-based subspace learning method is presented. F-2D-QPCA employs F-norm as the distance metric to measure image covariance matrix. Compared to F-norm 2DPCA method and most existing 2D-PCA methods, F-2D-QPCA method is more accurate. Also, compared to 2D-QPCA method, F-2D-QPCA method is slightly more robust to outliers. Moreover, our method retains 2D-QPCA's desirable properties such as rotational invariance and geometric structure. To solve F-2D-QPCA, we propose a nongreedy iterative algorithm which has good convergence. Experimental results on several color face image databases show the effectiveness of F-2D-QPCA method.

REFERENCES

[1] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 103–108, Jan. 1990.

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

[3] 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.

[4] L. Torres, J. Y. Reutter, and L. Lorente, "The importance of the color information in face recognition," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3, Oct. 1999, pp. 627–631.

[5] J. Yang and C. Liu, "A general discriminant model for color face recognition," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–6.

[6] X. Xiang, J. Yang, and Q. Chen, "Color face recognition by PCA-like approach," *Neurocomputing*, vol. 152, pp. 231–235, Mar. 2015.

[7] Y. Zhu, C. Zhu, and X. Li, "Improved principal component analysis and linear regression classification for face recognition," *Signal Process.*, vol. 145, pp. 175–182, Apr. 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] 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.

[10] 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 Conf. Comput. Vis. Pattern Recognit.*, vol. 1, San Diego, CA, USA, Jun. 2005, pp. 739–746.

[11] N. Kwak, "Principal component analysis based on L_1-norm maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1672–1680, Sep. 2008.

[12] 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.*, vol. 22, Barcelona, Spain, 2011, pp. 1433–1438.

[13] X. Li, Y. Pang, and Y. Yuan, "L_1-norm-based 2DPCA," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 4, pp. 1170–1175, Aug. 2010.

[14] R. Wang, F. Nie, X. Yang, F. Gao, and M. Yao, "Robust 2DPCA with non-greedy L_1-norm maximization for image analysis," *IEEE Trans. Cybern.*, vol. 45, no. 5, pp. 1108–1112, Dec. 2015.

[15] Y. Pang, X. Li, and Y. Yuan, "Robust tensor analysis with L_1-norm," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 2, pp. 172–178, Feb. 2010.

[16] H. Wang and J. Wang, "2DPCA with L_1-norm for simultaneously robust and sparse modelling," *Neural Netw.*, vol. 46, pp. 190–198, Oct. 2013.

[17] J.-X. Mi, Y.-N. Zhang, Z. Lai, W. Li, L. Zhou, and F. Zhong, "Principal component analysis based on nuclear norm minimization," *Neural Netw.*, vol. 118, pp. 1–16, Oct. 2019.

[18] J. Mi, Y.-N. Zhang, Y. Li, and Y. Shu, "Generalized two dimensional PCA based on L_2, p-norm minimization," *Int. J. Mach. Learn. Cybern.*, vol. 11, pp. 2421–2438, May 2020, doi: 10.1007/s13042-020-01127-1.

[19] Q. Wang, Q. Gao, X. Gao, and F. Nie, "Angle principal component analysis," in *Proc. Int. Joint Conf. Artif. Intell.*, vol. 2, Jul. 2017, pp. 1201–1207.

[20] 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.

[21] T. Jiang, "Algebraic methods for diagonalization of a quaternion matrix in quaternionic quantum theory," *J. Math. Phys.*, vol. 46, no. 5, May 2005, Art. no. 052106.

[22] C. Ding, D. Zhou, X. He, and H. Zha, "R1-PCA: Rotational invariant L_1-norm principal component analysis for robust subspace factorization," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, Pittsburgh, PA, USA, 2006, pp. 281–288.

[23] 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.

[24] 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.

[25] N. Le Bihan and S. J. Sangwine, "Quaternion principal component analysis of color images," in *Proc. Int. Conf. Image Process.*, Barcelona, Spain, 2003, pp. 809–812.

[26] Z. Jia, *The Eigenvalue Problem of Quaternion Matrix: Structure-Preserving Algorithms and Applications*. Beijing, China: Science Press, Dec. 2019.

[27] P. Denis, P. Carre, and C. Fernandez-Maloigne, "Spatial and spectral quaternionic approaches for colour images," *Comput. Vis. Image Understand.*, vol. 107, nos. 1–2, pp. 74–87, Jul. 2007.

[28] L. Shi and B. Funt, "Quaternion color texture segmentation," *Comput. Vis. Image Understand.*, vol. 107, nos. 1–2, pp. 88–96, Jul. 2007.

[29] Y. Li, M. Wei, F. Zhang, and J. Zhao, "A new double color image watermarking algorithm based on the SVD and arnold scrambling," *J. Appl. Math.*, vol. 2016, pp. 1–9, Nov. 2016.

[30] L. Liu, S. Li, and C. L. P. Chen, "Quaternion locality-constrained coding for color face hallucination," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1474–1485, May 2018.

[31] C. Zou, K. I. Kou, and Y. Wang, "Quaternion collaborative and sparse representation with application to color face recognition," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3287–3302, Jul. 2016.

[32] X. Xiao, Y. Chen, Y.-J. Gong, and Y. Zhou, "2D quaternion sparse discriminant analysis," *IEEE Trans. Image Process.*, vol. 29, pp. 2271–2286, Feb. 2020.

[33] Y. Chen, X. Xiao, and Y. Zhou, "Low-rank quaternion approximation for color image processing," *IEEE Trans. Image Process.*, vol. 29, pp. 1426–1439, Feb. 2020.

[34] Z. Jia, M. K. Ng, and G. Song, "Robust quaternion matrix completion with applications to image inpainting," *Numer. Linear Algebra Appl.*, vol. 26, no. 4, p. e2245, Aug. 2019.

[35] Z. Jia, S. Ling, and M. Zhao, "Color two-dimensional principal component analysis for face recognition based on quaternion model," in *Proc. Int. Conf. Intell. Comput.* Cham, Switzerland: Springer, 2017, pp. 177–189.

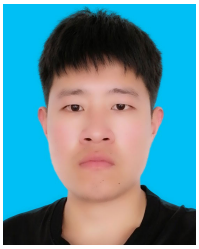
[36] M. Zhao, Z. Jia, and D. Gong, "Improved two-dimensional quaternion principal component analysis," *IEEE Access*, vol. 7, pp. 79409–79417, Nov. 2019.

[37] Z. Jia, M. Wei, and S. Ling, "A new structure-preserving method for quaternion hermitian eigenvalue problems," *J. Comput. Appl. Math.*, vol. 239, pp. 12–24, Jan. 2013.

[38] Z. Jia, M. Wei, M.-X. Zhao, and Y. Chen, "A new real structure-preserving quaternion QR algorithm," *J. Comput. Appl. Math.*, vol. 343, pp. 26–48, Feb. 2018.
[39] Z. Jia, K. M. Ng, and G. Song, "Lanczos method forlarge-scale quaternion singular value decomposition," *Numer. Algorithms*, vol. 82, no. 2, pp. 699–717, Apr. 2019.
[40] D. Zhang and Z. Zhou, "$(2D)^2$PCA: Two-directional two-dimensional PCA for efficient face representation and recognition," *Neurocomputing*, vol. 69, pp. 224–231, Feb. 2005.
[41] T. Jiang and L. Chen, "Algebraic algorithms for least squares problem in quaternionic quantum theory," *Comput. Phys. Commun.*, vol. 176, no. 7, pp. 481–485, Apr. 2007.
[42] K. Liu, Y.-Q. Cheng, and J.-Y. Yang, "Algebraic feature extraction for image recognition based on an optimal discriminant criterion," *Pattern Recognit.*, vol. 26, no. 6, pp. 903–911, Jun. 1993.
[43] J. Yang and J.-Y. Yang, "From image vector to matrix: A straightforward image projection technique—IMPCA vs. PCA," *Pattern Recognit.*, vol. 35, no. 9, pp. 1997–1999, Sep. 2002.
[44] X. Xiao and Y. Zhou, "Two-dimensional quaternion PCA and sparse PCA," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 7, pp. 2028–2042, Jul. 2019.
[45] S. Sangwine and N. Le Bihan. *Quaternion Toolbox for MATLAB*. Accessed: Sep. 2015. [Online]. Available: http://qtfm.sourceforge.net/
[46] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.

MINGHUI WANG received the B.A., M.S., and Ph.D. degrees from Qufu Normal University and East China Normal University in 2000, 2003, and 2008, respectively. He is currently a Professor of mathematics with the School of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao, China. His current research interests include numerical algebra and image processing.

LILI SONG is currently pursuing the M.S. degree with Qingdao University of Science and Technology, Qingdao, China. Her current research interest includes image processing.



KAISONG SUN is currently pursuing the B.S. degree from the Qingdao University of Science and Technology, Qingdao, China. His current research interest includes image processing.

ZHIGANG JIA received the Ph.D. degree in mathematics from East China Normal University, Shanghai, China, in 2009. He is currently a Professor of mathematics with the School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou, China. His current research interests include numerical linear algebra, machine learning, eigenvalue theory and algorithms, quaternion matrix computations, color image recognition, and imaging science.