Two-Dimensional Quaternion PCA and Sparse PCA

Xiaolin Xiao and Yicong Zhou, *Senior Member, IEEE*

***Abstract*— Benefited from quaternion representation that is able to encode the cross-channel correlation of color images, quaternion principle component analysis (QPCA) was proposed to extract features from color images while reducing the feature dimension. A quaternion covariance matrix (QCM) of input samples was constructed, and its eigenvectors were derived to find the solution of QPCA. However, eigen-decomposition leads to the fixed solution for the same input. This solution is susceptible to outliers and cannot be further optimized. To solve this problem, this paper proposes a novel quaternion ridge regression (QRR) model for two-dimensional QPCA (2D-QPCA). We mathematically prove that this QRR model is equivalent to the QCM model of 2D-QPCA. The QRR model is a general framework and is flexible to combine 2D-QPCA with other technologies or constraints to adapt different requirements of real-world applications. Including sparsity constraints, we then propose a quaternion sparse regression model for 2D-QSPCA to improve its robustness for classification. An alternating minimization algorithm is developed to iteratively learn the solution of 2D-QSPCA in the equivalent complex domain. In addition, 2D-QPCA and 2D-QSPCA can preserve the spatial structure of color images and have a low computation cost. Experiments on several challenging databases demonstrate that 2D-QPCA and 2D-QSPCA are effective in color face recognition, and 2D-QSPCA outperforms the state of the arts.**

***Index Terms*— 2D-principle component analysis (2D-PCA), 2D-quaternion PCA (2D-QPCA), 2D-quaternion sparse PCA (2D-QSPCA), color face recognition, feature extraction, partial occlusions, QPCA.**

I. Introduction

PRINCIPAL component analysis (PCA) [1] is an unsupervised learning approach for feature extraction and dimension reduction. The core of PCA is to reduce the dimension of input samples while preserving the variations in these samples as many as possible. This is achieved by seeking an orthonormal basis, such that when projecting samples onto this basis, the first few dimensions retain most of the original variations. PCA has been widely used in the fields of computer vision and pattern recognition [2]–[7].

When being applied to extract feature from images, PCA treats each sample as a vector, and hence, 2D images are converted to high-dimensional (HD) vectors prior to feature extraction [2]. To avoid the intensive computation of HD data, 2D-PCA [3] was proposed. 2D-PCA is computationally efficient because it constructs the sample covariance matrix directly from 2D images, relieving the burden of calculating the sample covariance matrix from HD vectors. Moreover, 2D-PCA copes with 2D matrices and, thus, preserves the spatial structure of images [3]–[5].

Both PCA and 2D-PCA use l_2-norm as measurement. This brings at least two limitations. First, the projected images (extracted features) are linear combinations of all original variables, making it difficult to interpret the extracted features; second, the l_2-norm measurement is sensitive to outliers and noise [8]. To overwhelm these limitations, the l_1-norm measurement was employed [9]–[12]. According to the aim of using l_1-norm measurement, the PCA-based and 2D-PCA-based algorithms can be classified into: 1) robust algorithms, e.g., PCA based on l_1-norm [13] and 2D-PCA based on l_1-norm (2D-PCA-L1) [7]; 2) sparse algorithms, e.g., sparse PCA (SPCA) [9]; and 3) their combinations, such as 2D-PCA-L1 with sparsity (2D-PCA-L1S) [14] and generalized 2D-PCA (G2D-PCA) (l_p-norm, $0 < p < 2$) [15]. The robust algorithms use l_1-norm to measure errors, alleviating the influence of outliers [13]. However, they do not have rotation invariance, which is an essential property of l_2-norm-based PCA algorithms and is in favor with learning algorithms [16], [17]. The sparse algorithms regularize the objective functions using l_1-norm penalties for sparse feature extraction. Projecting images onto the sparse basis, the sparse algorithms rely on the important variables and ignore the less important variables. This is helpful to discover the underlying patterns on the images. The combination algorithms minimize l_1-norm-based errors while extracting sparse features, inheriting the strengths and drawbacks of the robust and the sparse algorithms.

To process color images, the aforementioned methods treat three color channels independently or concatenate the representations of different color channels into large matrices and, thus, fail to consider the cross-channel correlation [18]–[20]. Nevertheless, this correlation is important for recognition [21]. To preserve the cross-channel correlation, tensor representation (TR) [22], reduced biquaternion representation (BR) [23], and quaternion representation (QR) [25]–[27] were utilized to represent color images. TR uses third-order tensors to represent color images, and multilinear PCA (MPCA) [28] and multilinear SPCA (MSPCA) [29] were proposed based on TR.

BR and QR employ the Clifford algebra to represent color images. The Clifford algebra yields an excellent representation for the rotation group [30]. It has shown promising performance in the application of computer vision [30].

Manuscript received December 12, 2017; revised April 17, 2018 and July 6, 2018; accepted September 24, 2018. Date of publication November 6, 2018; date of current version June 14, 2019. This work was supported in part by the Macau Science and Technology Development Fund under Grant FDCT/189/2017/A3 and in part by the Research Committee at the University of Macau under Grant MYRG2016-00123-FST and Grant MYRG2018-00136-FST. *(Corresponding author: Yicong Zhou.)*

The authors are with the Department of Computer and Information Science, University of Macau, Macau 999078, China (e-mail: yicongzhou@umac.mo).

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

Digital Object Identifier 10.1109/TNNLS.2018.2872541

However, the reduced biquaternions cannot form a division system [23], [30]. This results in the inelegant forms in the mathematical derivations of BR. Reduced biquaternion PCA (BPCA) and 2D-BPCA [23] represent color images using BR. Meanwhile, QR transforms color images into an orthogonal color space, resulting in the superiority for the discriminative task [31]. Quaternion PCA (QPCA) [25] converts each color image into an HD quaternion vector and, hence, endures high computation cost and the loss of the spatial structure of color images. Row-based 2D quaternion PCA (R2D-QPCA) [32] solves the above-mentioned problems by directly processing quaternion matrices. Essentially, R2D-QPCA works in the row direction of images [5], [33].

This paper first proposes 2D-QPCA to preserve the spatial structure of color images. To find optimal projection bases, existing QPCA-based algorithms (see [25], [32]) construct the quaternion covariance matrices (QCMs) and find their solutions via quaternion eigen-decomposition (QED). However, the solution of eigen-decomposition is fixed and is inflexible to adapt different applications. To address this issue, we further propose 2D-QSPCA. Our contributions are listed as follows.

1) In addition to the QCM model, we propose a novel quaternion ridge regression (QRR) model for 2D-QPCA, and then mathematically prove that this QRR model is equivalent to the QCM model of 2D-QPCA. To the best of our knowledge, we are the first to introduce the QRR model for the QPCA-based algorithms. The QRR model is a general framework that offers the great flexibility to combine 2D-QPCA with additional assumptions to regularize the solution of 2D-QPCA to fit various real-world applications.
2) Combining the QRR model with sparse regularization, we further propose a quaternion sparse regression (QSR) model for 2D-QSPCA to improve the robustness of 2D-QPCA.
3) We propose an alternating minimization algorithm to iteratively learn the solution of 2D-QSPCA in the equivalent complex space. In each iteration, the optimization is designed under the framework of the complex-valued Alternating Direction Method of Multipliers (complex ADMM).
4) The effectiveness of 2D-QPCA and 2D-QSPCA is verified by the application of color face recognition. Experiments show that 2D-QPCA and 2D-QSPCA can well recognize color face images with varying expressions, and 2D-QSPCA is robust to partial occlusions.

In the rest of this paper, Section II provides the background knowledge. Section III introduces two equivalent models for 2D-QPCA. Section IV proposes a QSR model for 2D-QSPCA, elaborates its solution, and discusses the benefit from sparse regularization. The effectiveness of the proposed algorithms is demonstrated by the application of color face recognition in Section V. We discuss and compare the proposed methods in Section VI. Finally, conclusions are drawn in Section VII.

II. Preliminaries

In this paper, we represent scalars, vectors, and matrices in real space ($\mathbb{R}$) and complex space ($\mathbb{C}$) using lowercase letters, bold lowercase letters, and bold uppercase letters, respectively, e.g., a, $\mathbf{a}$, and $\mathbf{A}$. A dot above the variable is used to denote variables in quaternion space ($\mathbb{H}$), e.g., $\dot{a}$, $\dot{\mathbf{a}}$, and $\dot{\mathbf{A}}$.

$\mathrm{Tr}(\cdot)$ represents the trace of a matrix and $\mathrm{Re}(\cdot)$ denotes the real (scalar) part of a complex or quaternion variable.

A. Quaternion Representation

1) Quaternion Numbers: The quaternions are a 4D number system [34], which is an extension of the complex number system. A quaternion number ($\dot{q} \in \mathbb{H}$) has four parts, i.e., one real part and three imaginary parts. It can be represented as

$$\dot{q} = q_0 + q_1 i + q_2 j + q_3 k \tag{1}$$

where $q_0, q_1, q_2, q_3 \in \mathbb{R}$ and $\{1, i, j, k\}$ is the basis satisfying

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

$$ij = -ji = k, \quad jk = -kj = i, \quad ki = -ik = j. \tag{3}$$

Following (2) and (3) and the distributive law, the multiplication of two quaternion numbers is noncommutative, i.e., $\dot{p}\dot{q} \neq \dot{q}\dot{p}$ in general. Two widely used operators of quaternion numbers are:

1) conjugate: $\overline{\dot{q}} = q_0 - (q_1 i + q_2 j + q_3 k)$;
2) modulus: $|\dot{q}| = (\dot{q}\overline{\dot{q}})^{1/2} = (\overline{\dot{q}}\dot{q})^{1/2} = (q_0^2 + q_1^2 + q_2^2 + q_3^2)^{1/2}$.

2) Quaternion Vectors and Matrices: Let $\dot{\mathbf{Q}} = (\dot{q}_{s,t}) \in \mathbb{H}^{m\times n}$, where $s = 1, \ldots, m$ and $t = 1, \ldots, n$ are the row and column indices, respectively. Then,

1) conjugate: $\overline{\dot{\mathbf{Q}}} = (\overline{\dot{q}_{st}}) \in \mathbb{H}^{m\times n}$.
2) transpose: $\dot{\mathbf{Q}}^T = (\dot{q}_{ts}) \in \mathbb{H}^{n\times m}$.
3) conjugate transpose: $\dot{\mathbf{Q}}^* = (\overline{\dot{\mathbf{Q}}})^T = \overline{\dot{\mathbf{Q}}^T} = (\overline{\dot{q}_{ts}}) \in \mathbb{H}^{n\times m}$.
4) $\dot{\mathbf{Q}}$ is Hermitian if $\dot{\mathbf{Q}}^* = \dot{\mathbf{Q}}$.

The following gives some definitions and properties on quaternion vectors and matrices.

Definition 1: Let $\dot{\mathbf{q}} = (\dot{q}_s) \in \mathbb{H}^m$, where $s = 1, \ldots, m$ is a position index. The l_1-norm and l_2-norm of $\dot{\mathbf{q}}$ are defined as $\|\dot{\mathbf{q}}\|_1 = \sum_{s=1}^m |\dot{q}_s|$ and $\|\dot{\mathbf{q}}\|_2 = (\sum_{s=1}^m |\dot{q}_s|^2)^{(1/2)}$, respectively; let $\dot{\mathbf{Q}} = (\dot{q}_{s,t}) \in \mathbb{H}^{m\times n}$. The F-norm of $\dot{\mathbf{Q}}$ is defined as $\|\dot{\mathbf{Q}}\|_F = (\sum_{s=1}^m \sum_{t=1}^n |\dot{q}_{s,t}|^2)^{(1/2)} = [\mathrm{Tr}(\dot{\mathbf{Q}}^*\dot{\mathbf{Q}})]^{(1/2)}$.

The noncommutativity of quaternion multiplication makes it quite difficult to cope with quaternion matrices. As pointed in [35], one of the effective approaches to process quaternion matrices is to convert them into pairs of complex matrices.

Definition 2: Let $\dot{\mathbf{Q}} = \mathbf{Q}_0 + \mathbf{Q}_1 i + \mathbf{Q}_2 j + \mathbf{Q}_3 k \in \mathbb{H}^{m\times n}$, $\mathbf{Q}_0, \mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3 \in \mathbb{R}^{m\times n}$. The Cayley–Dickson construction [36] represents $\dot{\mathbf{Q}}$ using an ordered pair of complex matrices

$$\dot{\mathbf{Q}} = \mathbf{Q}_a + \mathbf{Q}_b j \tag{4}$$

where $\mathbf{Q}_a = \mathbf{Q}_0 + \mathbf{Q}_1 i$, $\mathbf{Q}_b = \mathbf{Q}_2 + \mathbf{Q}_3 i$, and $\mathbf{Q}_a, \mathbf{Q}_b \in \mathbb{C}^{m\times n}$.

Definition 3: Let $\dot{\mathbf{Q}} = \mathbf{Q}_a + \mathbf{Q}_b j \in \mathbb{H}^{m\times n}$. The complex adjoint form [35] of $\dot{\mathbf{Q}}$ is formulated as

$$\chi_{\dot{\mathbf{Q}}} = \begin{bmatrix} \mathbf{Q}_a & \mathbf{Q}_b \\ -\overline{\mathbf{Q}}_b & \overline{\mathbf{Q}}_a \end{bmatrix} \tag{5}$$

where $\chi_{\dot{\mathbf{Q}}} \in \mathbb{C}^{2m\times 2n}$. $\dot{\mathbf{Q}}$ and $\chi_{\dot{\mathbf{Q}}}$ are isomorphic [35]. The complex adjoint form has been widely used for quaternion matrix analysis, including QED [35] and quaternion singular value decomposition (QSVD) [37].

Properties: Let $\dot{\mathbf{P}}, \dot{\mathbf{Q}} \in \mathbb{H}^{m\times m}$. Then [35],

1. $(\dot{\mathbf{P}}\dot{\mathbf{Q}})^* = \dot{\mathbf{Q}}^*\dot{\mathbf{P}}^*$.
2. $(\chi_{\dot{\mathbf{Q}}})^* = \chi_{\dot{\mathbf{Q}}^*}$.
3. $\chi_{(\dot{\mathbf{P}}+\dot{\mathbf{Q}})} = \chi_{\dot{\mathbf{P}}} + \chi_{\dot{\mathbf{Q}}}$.
4. $\chi_{\dot{\mathbf{P}}\dot{\mathbf{Q}}} = \chi_{\dot{\mathbf{P}}}\chi_{\dot{\mathbf{Q}}}$.
5. $2\|\dot{\mathbf{Q}}\|_F^2 = 2\mathrm{Tr}(\dot{\mathbf{Q}}^*\dot{\mathbf{Q}}) = \|\chi_{\dot{\mathbf{Q}}}\|_F^2 = \mathrm{Tr}(\chi_{\dot{\mathbf{Q}}^*}\chi_{\dot{\mathbf{Q}}})$.

B. 2D-PCA

Let $\{\mathbf{X}_i \in \mathbb{R}^{m\times n}\}_{i=1}^h$ denote a set of 2D samples. The aim of 2D-PCA is to find a set of orthonormal projection basis vectors, denoted by the columns of $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_k]$, such that, when projected onto $\mathbf{V}$, the projected samples have the maximal scatter [5]. We define $\mathbf{Y}_i = \mathbf{V}^T\mathbf{X}_i$ as the projected sample of $\mathbf{X}_i$. The scatter of the projected samples, denoted by $J(\mathbf{V})$, can be characterized by the trace of the covariance matrix of these projected samples. Assuming that all samples are mean-centered, i.e., $E(\mathbf{X}_i) = \mathbf{0}$, the objective of 2D-PCA is to maximize

$$\begin{aligned} J(\mathbf{V}) &= \mathrm{Tr}\{E[(\mathbf{Y} - E\mathbf{Y})(\mathbf{Y} - E\mathbf{Y})^T]\} \\ &= \mathrm{Tr}\{E[(\mathbf{V}^T\mathbf{X}_i - E(\mathbf{V}^T\mathbf{X}_i))(\mathbf{V}^T\mathbf{X}_i - E(\mathbf{V}^T\mathbf{X}_i))^T]\} \\ &= \mathrm{Tr}\{\mathbf{V}^T E[(\mathbf{X}_i - E(\mathbf{X}_i))(\mathbf{X}_i - E(\mathbf{X}_i))^T]\mathbf{V}\} \\ &= \mathrm{Tr}[\mathbf{V}^T E(\mathbf{X}_i\mathbf{X}_i^T)\mathbf{V}] \end{aligned} \tag{6}$$

in which

$$\Phi = E\left(\mathbf{X}_i\mathbf{X}_i^T\right) = \frac{1}{h}\sum_{i=1}^{h}\left(\mathbf{X}_i\mathbf{X}_i^T\right) \tag{7}$$

is actually the covariance matrix of the input samples.

Rewriting (6) with (7), together with the orthonormal constraint on the projection basis, the objective function of 2D-PCA is expressed by

$$\begin{aligned} \hat{\mathbf{V}} &= \arg\max_{\mathbf{V}} \ [\mathrm{Tr}(\mathbf{V}^T\Phi\mathbf{V})] \\ &\text{s.t. } \mathbf{V}^T\mathbf{V} = \mathbf{I}_k \end{aligned} \tag{8}$$

where the optimal columns of $\mathbf{V}$ are the eigenvectors of Φ corresponding to the first k largest eigenvalues [3], [5].

III. 2D-QPCA

This section first constructs a QCM model for 2D-QPCA. We then introduce a novel QRR model for 2D-QPCA and prove that the solution of this QRR model is mathematically equivalent to that of the QCM model. Finally, we compare and discuss the two models.

A. QCM Model for 2D-QPCA

Similar to 2D-PCA, the objective of 2D-QPCA is to find an orthonormal quaternion basis so that the projected quaternion samples have the largest scatter after projection. Let the columns of $\dot{\mathbf{V}} = [\dot{\mathbf{v}}_1, \ldots, \dot{\mathbf{v}}_k]$ be the optimal quaternion basis vectors, $\{\dot{\mathbf{X}}_i \in \mathbb{H}^{m\times n}\}_{i=1}^h$ and $\{\dot{\mathbf{Y}}_i \in \mathbb{H}^{k\times n}\}_{i=1}^h$ be the set of 2D quaternion samples and the set of projected quaternion samples, respectively. Without the loss of generality, all quaternion samples are mean-centered, i.e., $E(\dot{\mathbf{X}}_i) = \mathbf{0}$. 2D-QPCA seeks optimal $\dot{\mathbf{V}}$ that maximizes the scatter of the projected quaternion samples. This scatter can be characterized by the trace of the QCM of the projected quaternion samples as

$$\begin{aligned} J(\dot{\mathbf{V}}) &= \mathrm{Tr}\{E[(\dot{\mathbf{Y}} - E\dot{\mathbf{Y}})(\dot{\mathbf{Y}} - E\dot{\mathbf{Y}})^*]\} \\ &= \mathrm{Tr}\{E[(\dot{\mathbf{V}}^*\dot{\mathbf{X}}_i - E(\dot{\mathbf{V}}^*\dot{\mathbf{X}}_i))(\dot{\mathbf{V}}^*\dot{\mathbf{X}}_i - E(\dot{\mathbf{V}}^*\dot{\mathbf{X}}_i))^*]\} \\ &= \mathrm{Tr}\{\dot{\mathbf{V}}^* E[(\dot{\mathbf{X}}_i - E(\dot{\mathbf{X}}_i))(\dot{\mathbf{X}}_i - E(\dot{\mathbf{X}}_i))^*]\dot{\mathbf{V}}\} \\ &= \mathrm{Tr}[\dot{\mathbf{V}}^* E(\dot{\mathbf{X}}_i\dot{\mathbf{X}}_i^*)\dot{\mathbf{V}}] \end{aligned} \tag{9}$$

in which

$$\dot{\Phi} = E\left(\dot{\mathbf{X}}_i\dot{\mathbf{X}}_i^*\right) = \frac{1}{h}\sum_{i=1}^{h}\left(\dot{\mathbf{X}}_i\dot{\mathbf{X}}_i^*\right) \tag{10}$$

is the QCM of the input quaternion samples.

The QCM model of 2D-QPCA can be expressed by

$$\begin{aligned} \hat{\dot{\mathbf{V}}} &= \arg\max_{\dot{\mathbf{V}}} \ [\mathrm{Tr}(\dot{\mathbf{V}}^*\dot{\Phi}\dot{\mathbf{V}})] \\ &\text{s.t. } \dot{\mathbf{V}}^*\dot{\mathbf{V}} = \mathbf{I}_k. \end{aligned} \tag{11}$$

The solution of (11) can be computed via QED on $\dot{\Phi}$ [25]. Let the Hermitian matrix $\dot{\Phi}$ admit a QED as

$$\dot{\Phi} = \dot{\mathbf{W}}\Sigma_r\dot{\mathbf{W}}^* \tag{12}$$

where r is the rank of $\dot{\Phi}$, Σ_r is a real diagonal matrix, and the diagonal elements of Σ_r are arranged in a decreasing order. Then, the eigenvectors of $\dot{\Phi}$ are the solution of 2D-QPCA. That is, $\dot{\mathbf{V}} = \dot{\mathbf{W}}(:, 1:k)$, $k \le r$. The procedures to solve the QCM model of 2D-QPCA are given in Algorithm 1.

Algorithm 1: Solving the QCM Model of 2D-QPCA

Input : Training set $\{\dot{\mathbf{X}}_i\}_{i=1}^h$ and the dimension k.
Output: Optimal projection basis $\dot{\mathbf{V}} = [\dot{\mathbf{v}}_1, ..., \dot{\mathbf{v}}_k]$.

1 Calculate the quaternion covariance matrix of input samples: $\dot{\Phi} = \frac{1}{h}\sum_{i=1}^{h}(\dot{\mathbf{X}}_i\dot{\mathbf{X}}_i^*)$.
2 Perform QED: $\dot{\Phi} = \dot{\mathbf{W}}\Sigma_r\dot{\mathbf{W}}^*$.
3 $\dot{\mathbf{V}} = \dot{\mathbf{W}}(:, 1:k)$, $k \le r$.

A similar idea that constructs the QCM model directly from quaternion matrices also exists in R2D-QPCA [32]. Essentially, R2D-QPCA operates on the image rows, while our 2D-QPCA is working in the column direction. Thus, our 2D-QPCA and R2D-QPCA are suitable to capture different patterns. This can be demonstrated in [32] and our experiments in Section V-B3.

B. QRR Model for 2D-QPCA

For the QCM model of 2D-QPCA, QED is an efficient and off-the-shelf technique to find the solution. In this section, we further design a more flexible QRR model for 2D-QPCA and prove that the solution of this QRR model is equivalent to that of the QCM model (Theorem 1).

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 [37]. Hence, the solution of (11) equals to the solution of its dual problem: $\hat{\dot{\mathbf{B}}} = \arg\min_{\dot{\mathbf{B}}}(\sum_{i=1}^{h} \|\dot{\mathbf{X}}_i - \dot{\mathbf{B}}\dot{\mathbf{B}}^*\dot{\mathbf{X}}_i\|_F^2)$ subject to $\dot{\mathbf{B}}^*\dot{\mathbf{B}} = \mathbf{I}_k$. Based on this observation, we propose the QRR model of 2D-QPCA.

Theorem 1: Let $\{\dot{\mathbf{X}}_i \in \mathbb{H}^{m\times n}\}_{i=1}^{h}$ be a set of 2D quaternion samples and $\dot{\mathbf{V}} = [\dot{\mathbf{v}}_1, \ldots, \dot{\mathbf{v}}_k]$ be the solution of (11).

For any $\lambda_2 \geq 0$, suppose that $\dot{\mathbf{A}} = [\dot{\mathbf{a}}_1, \ldots, \dot{\mathbf{a}}_k] \in \mathbb{H}^{m\times k}$ and $\dot{\mathbf{B}} = [\dot{\mathbf{b}}_1, \ldots, \dot{\mathbf{b}}_k] \in \mathbb{H}^{m\times k}$ satisfy

$$(\hat{\dot{\mathbf{A}}}, \hat{\dot{\mathbf{B}}}) = \arg\min_{\dot{\mathbf{A}},\dot{\mathbf{B}}} \left(\sum_{i=1}^{h} \|\dot{\mathbf{X}}_i - \dot{\mathbf{A}}\dot{\mathbf{B}}^*\dot{\mathbf{X}}_i\|_F^2 + \lambda_2 \|\dot{\mathbf{B}}\|_F^2 \right) \quad \text{s.t. } \dot{\mathbf{A}}^*\dot{\mathbf{A}} = \mathbf{I}_k. \tag{13}$$

Then, $\dot{\mathbf{v}}_j = (\dot{\mathbf{b}}_j / \|\dot{\mathbf{b}}_j\|_2)$, $j = 1, 2, \ldots, k$. The proof of Theorem 1 is given in the Appendix.

Equation (13) gives the QRR model of 2D-QPCA. The penalty term $\lambda_2\|\dot{\mathbf{B}}\|_F^2$ is used to avoid the potential colinearity problem when the number of input samples is much smaller than the dimension of input samples [38].

C. Discussion

Theorem 1 reveals that the solution of the QRR model is equivalent to that of the QCM model. For 2D-QPCA, we can directly obtain its solution of the QCM model via eigen-decomposition. However, this solution lacks of flexibility. For example, the data may contain noise and the sparse regularization is desired to correctly discover the underlying patterns. In this case, the QCM model fails. Fortunately, the QRR model provides a feasible and flexible framework to deal with this problem since the regression model can be easily combined with different regularizers for further optimization.

IV. 2D-QSPCA

As verified in [8] and [28], the sparse PCA-based algorithms help to interpret the projected samples and are robust to outliers and noise. Taking the advantages of sparse regularization, this section proposes 2D-QSPCA to improve the robustness of 2D-QPCA. The 2D-QSPCA is formulated as a QSR model by combining the QRR model with sparse regularization. This also shows the advantages of the QRR model as a general framework. We design a novel algorithm to solve this QSR model. Afterward, we provide a simulation example to demonstrate that 2D-QSPCA can discover the underlying patterns in the presence of noise.

A. QSR Model for 2D-QSPCA

Regularizing the QRR model of 2D-QPCA with the l_1-norm penalties, we propose a QSR model for 2D-QSPCA.

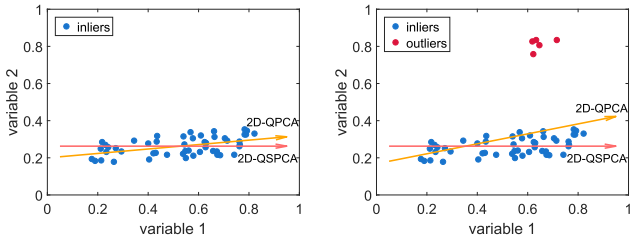

Fig. 1. Comparison of the projection bases of 2D-QPCA and 2D-QSPCA with (a) inliers and (b) inliers and outliers.

Theorem 2: Let $\{\dot{\mathbf{X}}_i \in \mathbb{H}^{m\times n}\}_{i=1}^{h}$ be a set of 2D quaternion samples and the columns of $\dot{\mathbf{V}}_s = [\dot{\mathbf{v}}_{s1}, \ldots, \dot{\mathbf{v}}_{sk}]$ be the quaternion sparse basis vectors of 2D-QSPCA. $\dot{\mathbf{V}}_s$ can be obtained as follows.

For any $\lambda_2 \geq 0$ and $\lambda_{1,j} \geq 0$, $j = 1, \ldots, k$, suppose that $\dot{\mathbf{A}} = [\dot{\mathbf{a}}_1, \ldots, \dot{\mathbf{a}}_k] \in \mathbb{H}^{m\times k}$ and $\dot{\mathbf{B}} = [\dot{\mathbf{b}}_1, \ldots, \dot{\mathbf{b}}_k] \in \mathbb{H}^{m\times k}$ satisfy

$$(\hat{\dot{\mathbf{A}}}, \hat{\dot{\mathbf{B}}}) = \arg\min_{\dot{\mathbf{A}},\dot{\mathbf{B}}} \left(\sum_{i=1}^{h} \|\dot{\mathbf{X}}_i - \dot{\mathbf{A}}\dot{\mathbf{B}}^*\dot{\mathbf{X}}_i\|_F^2 + \lambda_2 \|\dot{\mathbf{B}}\|_F^2 + \sum_{j=1}^{k} \lambda_{1,j} \|\dot{\mathbf{b}}_j\|_1 \right) \quad \text{s.t. } \dot{\mathbf{A}}^*\dot{\mathbf{A}} = \mathbf{I}_k. \tag{14}$$

Then, $\dot{\mathbf{v}}_{sj} = (\dot{\mathbf{b}}_j / \|\dot{\mathbf{b}}_j\|_2)$, $j = 1, 2, \ldots, k$.

Theorem 2 presents the QSR model for 2D-QSPCA. It is obvious that if $\lambda_{1,j} = 0$ for $j = 1, \ldots, k$, (14) reduces to the QRR model, and the obtained basis vectors are not sparse. Applying different coefficients to the l_1-norm penalties, this QSR model can provide a flexible control on the sparsity of its basis vectors.

Compared with 2D-QPCA, the sparse regularization on the projection basis of 2D-QSPCA brings at least the following three benefits:

1) The projected quaternion samples are obtained from only a few original variables, making it easier to interpret the projected samples.
2) Adopting sparse feature extraction, 2D-QSPCA is robust to outliers and, thus, beneficial for classification.
3) Sparse regularization helps to identify the important variables, which are associated with the underlying patterns in the presence of noise.

The first two properties can be illustrated in Fig. 1, and property (3) will be examined in Section IV-C. In Fig. 1, the data points represent quaternion samples whose values are denoted by the quaternion modulus for visualization, and each arrowed line represents the direction of the basis vector corresponding to the first principle component. Considering property (1), the basis vector of 2D-QPCA is the combination of the original variables 1 and 2, whereas that of 2D-QSPCA is associated only with the original variable 1. Thus, the projected samples of 2D-QSPCA can be represented using less original variables. For property (2), the direction of the basis vector of 2D-QPCA is easily influenced by outliers while that of 2D-QSPCA is resistant to outliers.

B. Solution of 2D-QSPCA

Owning to the noncommutativity of quaternion multiplication, directly solving the QSR model in quaternion space is quite difficult. To solve quaternion-valued models, different transformations have been utilized. For instance, Zou *et al*. [39] defined specific operators to convert the models in quaternion space into the models in real space, and QED [35] and QSVD [37] were implemented by transferring the quaternion-valued models into complex-valued models. In this paper, we reformulate (14) into an equivalent complex form. To start with, the converting of the quaternion l_1-norm to its equivalent complex form is given in Definition 4.

Definition 4: Let $\dot{\mathbf{q}} = \mathbf{q}_a + \mathbf{q}_b j \in \mathbb{H}^m$ and $\mathbf{q}$ be the first column of $\chi_{\dot{\mathbf{q}}}$, i.e., $\mathbf{q} = \chi_{\dot{\mathbf{q}}}(:, 1) = [\mathbf{q}_a; -\overline{\mathbf{q}_b}] \in \mathbb{C}^{2m}$. We define an operator as

$$\xi(\mathbf{q}) = \left[\mathbf{q}_a^T; \mathbf{q}_b^T\right]$$

where $\xi(\mathbf{q}) \in \mathbb{C}^{2\times m}$. Then, the l_1-norm of $\dot{\mathbf{q}}$ equals the $l_{2,1}$-norm of matrix $\xi(\mathbf{q})$, i.e.,

$$\|\dot{\mathbf{q}}\|_1 = \|\xi(\mathbf{q})\|_{2,1}$$

where $\|\mathbf{M}\|_{2,1}$ denotes the $l_{2,1}$-norm of $\mathbf{M} \in \mathbb{C}^{n\times m}$, and it is defined as $\|\mathbf{M}\|_{2,1} = \sum_{j=1}^{m} \|\mathbf{M}(:, j)\|_2$.

The quaternion F-norm terms in (14) can be easily transformed into complex space using the complex adjoint form (see the Appendix).

Let $\alpha = \chi_{\dot{\mathbf{A}}} \in \mathbb{C}^{2m\times 2k}$, $\beta = \chi_{\dot{\mathbf{B}}} \in \mathbb{C}^{2m\times 2k}$, and $\varphi = \sum_{i=1}^{h} \chi_{\dot{\mathbf{X}}_i} \chi_{\dot{\mathbf{X}}_i^*} \in \mathbb{C}^{2m\times 2m}$. There are two observations that are useful to rewrite (14): 1) φ is Hermitian and $\mathrm{Tr}(\alpha^*\varphi\beta)$ is the conjugate of $\mathrm{Tr}(\beta^*\varphi\alpha)$ and 2) the complex adjoint form has a redundant structure, and we can derive the right half columns of any complex adjoint matrix from its left half columns.

Equation (14), thus, can be converted into an complex form as

$$\begin{aligned}(\hat{\alpha}, \hat{\beta}) &= \arg\min_{\alpha,\beta}\Bigg\{ \mathrm{Tr}\varphi - 2\mathrm{Re}[\mathrm{Tr}(\alpha^*\varphi\beta)] + \mathrm{Tr}[\beta^*(\varphi + \lambda_2\mathbf{I})\beta] \\ &\qquad + 2\sum_{j=1}^{k} \lambda_{1,j}\|\xi(\beta_j)\|_{2,1}\Bigg\} \\ &= \arg\min_{\alpha,\beta}\Bigg\{ \sum_{j=1}^{k}[\beta_j^*(\varphi + \lambda_2\mathbf{I})\beta_j - 2\mathrm{Re}(\alpha_j^*\varphi\beta_j) \\ &\qquad + \lambda_{1,j}\|\xi(\beta_j)\|_{2,1}]\Bigg\} \\ &\text{s.t. } \alpha^*\alpha = \mathbf{I}_{2k} \end{aligned} \tag{15}$$

where β_j is the jth column of β, $j = 1, \ldots, k$.

We propose an alternating minimization algorithm to iteratively solve (15).

1) Fixing α, Find Optimal β: Given α, (15) reduces to individually solve the subproblem

$$\begin{aligned}&\hat{\beta}_j = \arg\min_{\beta_j}\ \left[\beta_j^*(\varphi + \lambda_2\mathbf{I})\beta_j - 2\mathrm{Re}\left(\alpha_j^*\varphi\beta_j\right) + \lambda_{1,j}\|\xi(\beta_j)\|_{2,1}\right] \\ &\text{for } j = 1, \ldots, k. \end{aligned} \tag{16}$$

Due to the sum-of-norms regularization ($l_{2,1}$-norm penalties), there is no closed-form expression for optimal β_j. We devise an algorithm to solve (16) under the complex ADMM framework [40]. Setting $\mathbf{Z} = \xi(\beta_j)$, the idea of complex ADMM is to convert (16) to a constrained optimization problem as

$$\begin{aligned}&\arg\min_{\beta_j,\mathbf{Z}}\ \left\{\beta_j^*(\varphi + \lambda_2\mathbf{I})\beta_j - 2\mathrm{Re}\left(\alpha_j^*\varphi\beta_j\right) + \lambda_{1,j}\|\mathbf{Z}\|_{2,1}\right\} \\ &\text{s.t. } \mathbf{Z} = \xi(\beta_j) \end{aligned} \tag{17}$$

and optimize its augmented Lagrangian function

$$\begin{aligned}L(\beta_j, \mathbf{Z}, \mathbf{y}) &= \beta_j^*(\varphi + \lambda_2\mathbf{I})\beta_j - 2\mathrm{Re}\left(\alpha_j^*\varphi\beta_j\right) + \lambda_{1,j}\|\mathbf{Z}\|_{2,1} \\ &\quad + \mathbf{y}^*[\beta_j - \xi^{-1}(\mathbf{Z})] + \frac{\rho}{2}\|\beta_j - \xi^{-1}(\mathbf{Z})\|_2^2 \end{aligned} \tag{18}$$

where $\xi^{-1}(\cdot)$ is the inverse operator of $\xi(\cdot)$ that concatenates the transpose of the rows of a $2 \times m$ matrix into a $2m \times 1$ vector, $\mathbf{y}$ is the Lagrangian multiplier, and $\rho > 0$ is the penalty parameter. The iterative scheme to optimize $L(\beta_j, \mathbf{Z}, \mathbf{y})$ is

$$\beta_j^{\tau+1} = \arg\min_{\beta_j}\ L(\beta_j, \mathbf{Z}^\tau, \mathbf{y}^\tau) \tag{19}$$

$$\mathbf{Z}^{\tau+1} = \arg\min_{\mathbf{Z}}\ L(\beta_j^{\tau+1}, \mathbf{Z}, \mathbf{y}^\tau) \tag{20}$$

$$\mathbf{y}^{\tau+1} = \mathbf{y}^\tau + \rho\left[\beta_j^{\tau+1} - \xi^{-1}(\mathbf{Z}^{\tau+1})\right] \tag{21}$$

where in each iteration, β_j, $\mathbf{Z}$,and$\mathbf{y}$ are updated one by one when other two variables are fixed.

Based on the above-mentioned scheme, we devise the iterative procedure to solve (18). Specifically, given the result of the τth iteration, the variables in the $(\tau + 1)$th iteration are updated by following the three steps.

1) Update $\beta_j^{\tau+1}$ by minimizing L with respect to β_j when $(\mathbf{Z}^\tau, \mathbf{y}^\tau)$ are fixed. This is implemented by setting the gradient of L with respect to β_j to 0

$$\beta_j^{\tau+1} = [\varphi + (\lambda_2 + \rho^\tau)\mathbf{I}]^{-1}[\varphi\alpha_j + \rho^\tau\xi^{-1}(\mathbf{Z}^\tau) - \mathbf{y}^\tau]. \tag{22}$$

2) Update $\mathbf{Z}^{\tau+1}$ by minimizing L with respect to $\mathbf{Z}$ when $(\beta_j^{\tau+1}, \mathbf{y}^\tau)$ are fixed. Considering $\mathbf{Z}$, the optimization of L is to find

$$\begin{aligned}&\min_{\mathbf{Z}}\Bigg\{\frac{\rho^\tau}{2}\left\|\beta_j^{\tau+1} - \xi^{-1}(\mathbf{Z})\right\|_2^2 \\ &\qquad + \mathbf{y}^*\left[\beta_j^{\tau+1} - \xi^{-1}(\mathbf{Z})\right] + \lambda_{1,j}\|\mathbf{Z}\|_{2,1}\Bigg\} \\ &= \min_{\mathbf{Z}}\left\{\frac{1}{2}\left\|\beta_j^{\tau+1} - \xi^{-1}(\mathbf{Z}) + \frac{\mathbf{y}^\tau}{\rho^\tau}\right\|_2^2 + \frac{\lambda_{1,j}}{\rho^\tau}\|\mathbf{Z}\|_{2,1}\right\} \\ &= \min_{\mathbf{Z}}\left\{\frac{1}{2}\left\|\xi^{-1}(\mathbf{Z}) - \left(\beta_j^{\tau+1} + \frac{\mathbf{y}^\tau}{\rho^\tau}\right)\right\|_2^2 + \frac{\lambda_{1,j}}{\rho^\tau}\|\mathbf{Z}\|_{2,1}\right\} \\ &= \min_{\mathbf{Z}}\left\{\frac{1}{2}\left\|\mathbf{Z} - \xi\left(\beta_j^{\tau+1} + \frac{\mathbf{y}^\tau}{\rho^\tau}\right)\right\|_F^2 + \frac{\lambda_{1,j}}{\rho^\tau}\|\mathbf{Z}\|_{2,1}\right\}. \end{aligned} \tag{23}$$

This way, the solution of (23) can be found using Lemma 1, which is a soft-thresholding process derived from the sum-of-norms regularized optimization in [39] and [40].

Lemma 1: If a problem considering $\mathbf{Z} \in \mathbb{C}$ is to find

$$\hat{\mathbf{Z}} = \arg\min_{\mathbf{Z}} \left\{ \frac{1}{2}\|\mathbf{Z} - \mathbf{R}\|_F^2 + \sigma \|\mathbf{Z}\|_{2,1} \right\}. \quad (24)$$

The optimal $\mathbf{Z}$ is obtained at

$$\hat{\mathbf{Z}}(:,i) = \begin{cases} \dfrac{\|\mathbf{R}(:,i)\|_2 - \sigma}{\|\mathbf{R}(:,i)\|_2}\mathbf{R}(:,i), & \|\mathbf{R}(:,i)\|_2 > \sigma \\ \mathbf{0}, & \text{otherwise.} \end{cases} \quad (25)$$

3) Update $\mathbf{y}^{\tau+1}$ using (21) when $(\beta_j^{\tau+1}, \mathbf{Z}^{\tau+1})$ are fixed.

Let the primal residual $r_{\text{pri}}^{\tau+1} = \beta_j^{\tau+1} - \xi^{-1}(\mathbf{Z}^{\tau+1})$ and the dual variable residual $r_{\text{dual}}^{\tau+1} = \rho(\mathbf{Z}^{\tau+1} - \mathbf{Z}^{\tau})$. The stopping criterion of the complex ADMM framework is set to $\|r_{\text{pri}}^{\tau+1}\|_2 < \varepsilon^{\text{pri}}$ and $\|r_{\text{dual}}^{\tau+1}\|_F < \varepsilon^{\text{dual}}$, and ε^{pri} and $\varepsilon^{\text{dual}}$ are small numbers [40].

The initialization of the penalty parameter ρ should be small to obtain an optimal solution [41]. However, this may lead the proposed algorithm converges too slowly. In this paper, we employ a simple yet effective scheme introduced in [41] to adjust ρ according to the primal and dual variable residuals. The scheme is defined as

$$\rho^{\tau+1} = \begin{cases} \nu^{\text{incr}}\rho^{\tau}, & \text{if } \|r_{\text{pri}}^{\tau}\|_2 > \mu\|r_{\text{dual}}^{\tau}\|_F \\ \rho^{\tau}/\nu^{\text{decr}}, & \text{if } \|r_{\text{dual}}^{\tau}\|_2 > \mu\|r_{\text{pri}}^{\tau}\|_F \\ \rho^{\tau}, & \text{otherwise} \end{cases} \quad (26)$$

where μ, ν^{incr},$\nu^{\text{decr}} > 1$ are predefined parameters. We set $\mu = 10$ and $\nu^{\text{incr}} = \nu^{\text{decr}} = 2$ as recommended in [41].

Finally, the proposed method to solve (16) is summarized in Algorithm 2. The same operations are performed independently to find the optimal β_j, $j = 1, \ldots, k$. Due to the redundant structure of complex adjoint form, for $j = k+1, \ldots, 2k$, we can deduce the optimal β_j from the optimal β_{j-k}.

2) Fixing β, Find Optimal α: Given β, the minimization of (15) with respect to α is equivalent to find

$$\hat{\alpha} = \arg\max_{\alpha} \text{Re}[\text{Tr}(\alpha^*\varphi\beta)] \quad \text{s.t. } \alpha^*\alpha = I_k \quad (27)$$

Algorithm 2: Computing β_j

Input : φ, α_j, λ_2, $\lambda_{1,j}$, ε^{pri}, and ε^{dual}.
Output: Optimal β_j.

1 Convert (16) to a constrained problem (17), and construct the augmented Lagrangian function as (18).
2 Initialize $\beta_j^0 = \mathbf{0}$, $\mathbf{Z}^0 = \mathbf{0}$, $\mathbf{y}^0 = \mathbf{0}$, and $\rho^0 = 10^{-3}$.
3 **repeat**
4 Update $\beta_j^{\tau+1}$ using (22).
5 Update $\mathbf{Z}^{\tau+1}$ using (25).
6 Compute $r_{pri}^{\tau+1} = \beta_j^{\tau+1} - \xi^{-1}(\mathbf{Z}^{\tau+1})$, and $r_{dual}^{\tau+1} = \rho^{\tau}(\mathbf{Z}^{\tau+1} - \mathbf{Z}^{\tau})$.
7 Update $\mathbf{y}^{\tau+1}$ using (21).
8 Adjust $\rho^{\tau+1}$ using (26).
9 **until** $\|r_{pri}^{\tau+1}\|_2 \le \varepsilon^{pri}$ *and* $\|r_{dual}^{\tau+1}\|_F \le \varepsilon^{dual}$;
10 Output $\beta_j^{\tau+1}$.

where α can be solved using Lemma 2 by setting $\eta = \varphi\beta$.

Lemma 2: Let $\alpha, \eta \in \mathbb{C}^{m\times k}$ and the rank of η be k ($k < m$). Consider the optimization

$$\hat{\alpha} = \arg\max_{\alpha} \text{Re}[\text{Tr}(\alpha^*\eta)] \quad \text{s.t. } \alpha^*\alpha = I_k. \quad (28)$$

Suppose that the SVD of η is $\eta = U_\eta \Sigma_\eta V_\eta^*$, then $\hat{\alpha} = U_\eta V_\eta^*$.

Please refer to the orthogonal Procrustes problem in the complex domain [42] for the proof of Lemma 2.

The alternating minimization algorithm starts at any $\alpha^*\alpha = \mathbf{I}_{2k}$. To make this algorithm converge fast, the initialization of α is set to the complex adjoint form of the solution of 2D-QPCA. The stopping condition is the convergence of all columns of β. Let $\xi(\beta_j)$ be the residual of the jth column of β. The alternating minimization algorithm stops when $\xi(\beta_j) < \varepsilon^{\text{outer}}$ for $j = 1, \ldots, k$, where $\varepsilon^{\text{outer}}$ is a small number.

Afterward, we can recover the quaternion-valued solution from the complex-valued solution using the operator in Definition 5 [37].

Definition 5: Given a complex column vector $\mathbf{c} = [c_1, c_2, \ldots, c_n, c_{n+1}, c_{n+2}, \ldots, c_{2n}]^T \in \mathbb{C}^{2n}$, we define an operator $\gamma(\mathbf{c})$: $\mathbb{C}^{2n} \to \mathbb{H}^n$ such that

$$\gamma(\mathbf{c}) = \begin{bmatrix} c_1 \\ c_2 \\ \cdots \\ c_n \end{bmatrix} + \begin{bmatrix} -\overline{c_{n+1}} \\ -\overline{c_{n+2}} \\ \cdots \\ -\overline{c_{2n}} \end{bmatrix} j. \quad (29)$$

The quaternion sparse projection basis $[\dot{\mathbf{v}}_{s1}, \ldots, \dot{\mathbf{v}}_{sk}]$ of 2D-QSPCA can be recovered from the columns of β using: $\dot{\mathbf{v}}_{sj} = \gamma(\beta_j/\|\beta_j\|_2)$, $j = 1, \ldots, k$. The detail procedures of 2D-QSPCA are listed in Algorithm 3.

C. Discussion

In real applications, the data set often contains noise. The important variables are associated with the underlying patterns in the presence of noise. The original PCA algorithm uses l_2-norm measure. Its basis vectors are linear combinations of all variables and, hence, may misidentify the important variables and underlying patterns. SPCA can identify the sets of important variables [9], so can 2D-QSPCA.

A straightforward way to obtain the sparse basis is to first find the original basis, manually set a threshold, and then discard the variables whose absolute values are smaller than the threshold. However, this simple thresholding approach also suffers from the misidentification problem [9].

To examine the ability of 2D-QSPCA in identifying the sets of important variables, we extend the synthetic example in [9] to quaternion domain and compare the results of 2D-QSPCA with 2D-QPCA and the simple thresholding method. First, we generate three quaternion hidden factors following the quaternion normal distribution ($\dot{N}$) [43]:

$$\dot{f}_1 \sim \dot{N}(0, 290), \quad \dot{f}_2 \sim \dot{N}(0, 300)$$
$$\dot{f}_3 = -0.3\dot{f}_1 + 0.925\dot{f}_2 + \dot{\epsilon}$$
$$\dot{\epsilon} \sim \dot{N}(0, 1)$$
$$\dot{f}_1, \dot{f}_2, \text{ and } \dot{\epsilon} \text{ are independent.}$$

TABLE I
COMPARISON OF 2D-QPCA, 2D-QSPCA, AND THE SIMPLE THRESHOLDING METHODS

	True Positive Rate			True Negative Rate			Variance		
	$\dot{\mathbf{v}}_1$	$\dot{\mathbf{v}}_2$	Mean	$\dot{\mathbf{v}}_1$	$\dot{\mathbf{v}}_2$	Mean	PC1	PC2	Cumulative Variance
2D-QPCA	1	1	1	0	0	0	0.599	0.3974	0.9964
2D-QSPCA	0.961	1	0.9805	0.9587	0.9874	0.9717	0.3921	0.392	0.7841
Simple Thresholding	0.57	1	0.785	0.7134	1	0.8567	0.39	0.386	0.776

Algorithm 3: 2D-QSPCA

Input : Training set $\{\dot{\mathbf{X}}_i\}_{i=1}^h$, the dimension k, and parameters λ_2, $\lambda_{1,j}$, $j = 1, ..., k$, $\varepsilon^{\text{outer}}$.
Output: Optimal quaternion sparse projection basis $[\dot{\mathbf{v}}_{s1}, ..., \dot{\mathbf{v}}_{sk}]$.

Rewrite 2D-QSPCA into its complex form (15). Convert the task to find $\alpha = [\alpha_1, ..., \alpha_{2k}]$ and $\beta = [\beta_1, ..., \beta_{2k}]$.
Initialize α to the complex adjoint form of the solution of 2D-QPCA.
repeat
  (1) Fixing α, find optimal β.
  for $j = 1, ..., k$, **do**
    Compute β_j using Algorithm 2.
    Compute residual $\xi(\beta_j)$.
  end
  for $j = k + 1, ..., 2k$, **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 $\varphi = \sum_{i=1}^{h} \chi_{\dot{\mathbf{X}}_i} \chi_{\dot{\mathbf{X}}_i^*}$.
until $\xi(\beta_j) < \varepsilon^{\text{outer}}$ *for* $j = 1, ..., k$;
for $j = 1, ..., k$, **do**
  Recover $\dot{\mathbf{v}}_{sj}$ from β_j using $\dot{\mathbf{v}}_{sj} = \gamma\left(\frac{\beta_j}{\|\beta_j\|_2}\right)$, where $\gamma(\cdot)$ is a recover operator in Definition 5.
end
Output $[\dot{\mathbf{v}}_{s1}, ..., \dot{\mathbf{v}}_{sk}]$.

Then, 25 observed quaternion variables are generated

$$\begin{aligned}
&\dot{x}_i = \dot{f}_1 + \dot{\epsilon}_i^1, \quad \dot{\epsilon}_i^1 \sim \dot{N}(0,1),\ i = 1, \ldots, 10\\
&\dot{x}_i = \dot{f}_2 + \dot{\epsilon}_i^2, \quad \dot{\epsilon}_i^2 \sim \dot{N}(0,1),\ i = 11, \ldots, 20\\
&\dot{x}_i = \dot{f}_3 + \dot{\epsilon}_i^3, \quad \dot{\epsilon}_i^3 \sim \dot{N}(0,1),\ i = 21, \ldots, 25\\
&\left\{\dot{\epsilon}_i^j\right\} \text{ are independent, } i = 1, \ldots, 25,\ j = 1, 2, 3.
\end{aligned}$$

$\dot{\mathbf{x}} = [\dot{x}_1, \ldots, \dot{x}_{25}]^T \in \mathbb{H}^{25}$ is used to generate synthetic samples. For simplicity, we use this 1D example because: 1) it can be regarded as a special case of a 2D model in which the column number equals one and 2) in this simulation, we aim to illustrate the ability of 2D-QSPCA in identifying the sets of important variables and, hence, the spatial structure of samples can be ignored. The spatial structure is more important when the samples are 2D.

The variances of hidden factors $\dot{f}_1$, $\dot{f}_2$, and $\dot{f}_3$ are 290, 300, and 283.79, respectively. Thus, $\dot{f}_2$ is slightly important than $\dot{f}_1$, and both of them are more important than $\dot{f}_3$. The numbers of associated variables with $\dot{f}_1$, $\dot{f}_2$, and $\dot{f}_3$ are 10, 10, and 5, respectively. Ideally, after dimension reduction, the first two principle components (denoted by PC1 and PC2) should explain 0.996 $((301 + 291)/(301 + 291 + 2))$ of the total variance. Based on the priori knowledge, if we set the number of basis vectors to $k = 2$ and restrict the nonzero elements on each basis vector to 10, then to correctly discover the underlying patterns, the first basis vector $(\dot{\mathbf{v}}_1)$ should recover $\dot{f}_2$ using the set of important variables $(0, \ldots, \dot{x}_{11}, \ldots, \dot{x}_{20}, 0, \ldots, 0)$, while the second basis vector $(\dot{\mathbf{v}}_2)$ should recover $\dot{f}_1$ using the set of important variables $(\dot{x}_1, \ldots, \dot{x}_{10}, 0, \ldots, 0)$.

Since quaternion numbers have four dimensions, it is hard to directly compare the absolute values. Instead, we use the norms of the quaternion numbers for comparison. Following [44], we record true positive rate (TPR) and true negative rate (TNR) values for each basis vector to evaluate the algorithms' ability in correctly identifying the sets of important variables, and rejecting the less important variables, respectively. For both measures, the higher value indicates the better performance. We generate 1000 quaternion samples from the above-mentioned model and run the simulation 100 times to avoid the interference of randomness. The TPR and TNR values for each basis vector and the variance of the corresponding principle component are reported in Table I. The 2D-QPCA can explain most of the variance in the data set. However, its basis vectors combine all variables and cannot reject the less important variables (TNR $= 0$). The 2D-QSPCA obtains TPR $= 0.9805$ and TNR $= 0.9717$, indicating that it can correctly identify the sets of important variables and reject the less important variables in most cases. For both TPR and TNR, the simple thresholding method achieves worse results than the proposed 2D-QSPCA. This simulation example verifies the effectiveness of 2D-QSPCA in identifying the sets of important variables.

The ability of 2D-QPCA and 2D-QSPCA in dimension reduction can also be verified using this example. We observed that the first two principle components of 2D-QPCA and 2D-QSPCA can explain 0.996 and 0.7841 of total variances, respectively. That is, they are able to use a few dimensions to represent most of the variations in the data set and, hence, can be employed for dimension reduction. Considering dimension reduction, 2D-QPCA performs better than 2D-QSPCA because 2D-QPCA can explain more variances using a fixed number of principle components. This is because 2D-QPCA provides the most compact representation of the data set under the constraint of least square error, while 2D-QSPCA balances the ability of maximizing the explained variances and that of discovering the underlying patterns.

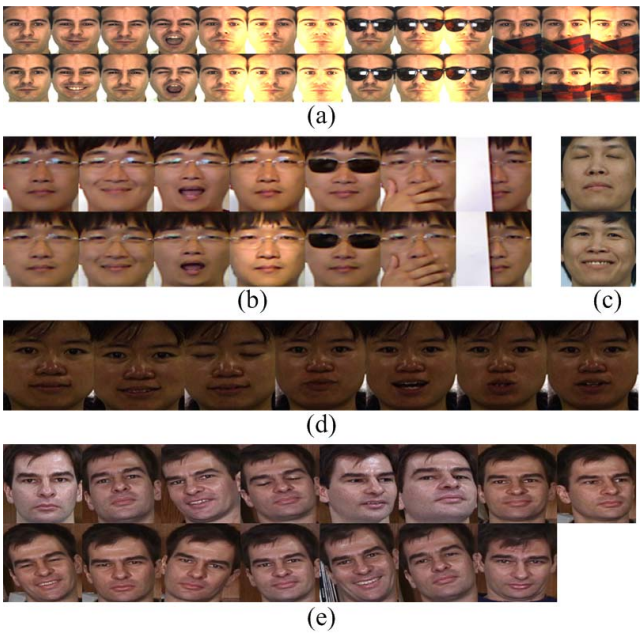

Fig. 2. Sample color face images of one person from (a) AR database, (b) EURECOM Kinect database, (c) color FERET database, (d) CMU PIE database, and (e) Georgia tech face database.

V. Experiments on Color Face Recognition

To validate the performance of 2D-QPCA and 2D-QSPCA, we employ color face recognition as an application example. In Section V-A, we describe the experimental settings. The recognition performance of 2D-QPCA and 2D-QSPCA is compared with that of peer algorithms in Section V-B. Section V-C statistically evaluates the performance of the competing algorithms.

A. Experiment Settings

In this section, we introduce the databases, the competing algorithms, and the parameter settings.

1) Databases: The following five widely used color face databases are employed for experiments.

a) AR face database [24]: It contains frontal color face images of 126 people recorded in two sessions. In each session, a neutral color face image is followed by images with different expressions and illumination, and images occluded by sunglasses and scarf (three images per condition). We employ a popular subset of AR containing 100 subjects [45], and hence, 2600 images in total are used. Sample images of one person from AR are given in Fig. 2(a).

b) EURECOM kinect database [46]: It provides color face images of 52 people with different expressions, lighting conditions, and occlusions. In our experiments, 728 frontal face images are used. Sample images of one person from EURECOM are given in Fig. 2(b).

c) Color FERET database [47]: It contains 14 126 color face images of 1199 individuals. We collect a subset that contains 265 subjects. Each subject has two frontal images with intense changes of expression (the images "fa" and "fb"). Fig. 2(c) provides example face images of one person from FERET.

d) CMU PIE database [48]: It provides color face images of 68 subjects with different poses, illumination, expressions, and frames from a talking sequence. We collect a subset (C27) of CMU PIE with, for each person, one neutral face image, two face images with expression changes (blinking and smiling), and four images from the talking sequence (frames 00, 19, 39, and 59). Fig. 2(d) gives the example face images of one person from CMU PIE.

e) Georgia tech face database [49]: It contains color face images of 50 individuals with 15 different poses per subject. Fig. 2(e) shows example face images of one person from GT.

The color face images in AR, EURECOM, FERET, and PIE are aligned and cropped based on the location of eyes. Then, all face images including that in Georgia tech face database (GT) are resized to 32×32 pixels. Each database is divided into a training set and a testing set. Optimal projection basis vectors are learned from the training set, and the testing samples are projected onto the basis vectors afterward. The projected testing samples are classified based on the nearest-neighbor classifier using l_1-norm distance.

2) Competing Algorithms: We compare the performance of 2D-QPCA and 2D-QSPCA with 13 state-of-the-art algorithms, namely, PCA [1], SPCA [9], linear discriminant analysis (LDA) [50], 2D-PCA [3], 2D-PCA-L1 [7], 2D-PCA-L1S [14], G2D-PCA [15], MPCA [28], MSPCA [29], QPCA [25], quaternion LDA (QLDA) [25], R2D-QPCA [32], and 2D-BPCA [23].

To represent color face images, MPCA and MSPCA utilize the third-order tensors, QPCA and QLDA use quaternion vectors, and 2D-BPCA utilizes reduced biquaternion matrices, while R2D-QPCA, 2D-QPCA, and 2D-QSPCA use quaternion matrices. Other algorithms employ the real domain 2D matrices or 1D vectors to represent samples. We extend them to process color images by concatenating the representations of three color channels into large vectors or matrices.

3) Parameter Settings: In the following, we specify the detailed parameter settings in the competing algorithms.

a) Parameters for 2D-QSPCA: To avoid the potential colinearity problem, we set $\lambda_2 = 0.001$.

The sparsity of the basis of 2D-QSPCA is controlled via the soft-thresholding process in Algorithm 2. According to (25), the threshold is set to $\sigma = (\lambda_{1,j}/\rho)$. In this paper, 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. Specifically, we sort the l_2-norm of the columns of $\mathbf{R}$ in a descending order and then set σ as the $(card + 1)$th largest value of sorted norms. Thus, after thresholding, only *card* columns of $\mathbf{Z}$ are kept, and the other columns are discarded. This way, when transferring back to the quaternion space, we have exact *card* nonzero quaternion elements per basis vector.

The alternating minimization algorithm (Algorithm 3) stops when the l_2-norm of the residual for each column of β is smaller than $\varepsilon^{\text{outer}} = 10^{-3}$, and the stopping criterion of Algorithm 2 is set to $\varepsilon^{\text{pri}} = \varepsilon^{\text{dual}} = 10^{-3}$.

b) Setups for peer algorithms: For PCA, SPCA, and QPCA, the projection dimension (k) is individually selected

TABLE II
EXPERIMENT RESULTS ON NONOCCLUDED COLOR FACE IMAGES

		PCA	SPCA	LDA	2D-PCA	2D-PCA-L1	2D-PCA-L1S	G2D-PCA	MPCA	MSPCA	QPCA	QLDA	R2D-QPCA	2D-BPCA	2D-QPCA	2D-QSPCA
AR	rate	0.7714	0.7557	0.4586	0.7957	0.7971	0.7971	0.8	0.7614	***0.8357***	0.8143	0.3514	0.81	0.7543	0.8343	**0.8914***
	dim	100	100	100	96*20	96*24	96*24	96*24	18*11*3	20*20*3	50*4	90*4	32*24*4	32*24*4	32*20*4	32*20*4
	card	1024	100	1024	32	32	–	–	32	8	32	32	32	32	32	4
EU	rate	0.726	0.7115	0.5	0.6923	0.6923	0.6923	0.6971	0.7548	***0.7933***	0.75	0.5144	0.7019	0.6635	***0.7933***	**0.8365**
	dim	90	80	50	96*16	96*16	96*16	96*16	20*10*3	20*20*3	20*4	50*4	32*16*4	32*20*4	32*20*4	32*20*4
	card	1024	100	1024	32	32	–	–	32	8	32	32	32	32	32	8
FERET	rate	0.7245	0.717	0.7359	0.7434	0.7547	0.7849	***0.8151***	0.7359	0.8113	0.7245	0.7170	0.7736	0.7811	***0.8151***	**0.834**
	dim	90	100	260	96*4	96*4	96*28	96*8	20*10*3	16*16*3	50*4	90*4	32*4*4	32*4*4	32*16*4	32*12*4
	card	1024	50	1024	32	32	–	–	32	2	32	32	32	32	32	4
PIE	rate	0.549	0.5074	0.5490	0.5735	0.5466	0.6544	0.6569	0.5613	0.6397	0.5441	0.5417	0.5833	0.6078	***0.6961****	**0.7034***
	dim	40	36	40	96*4	96*4	96*24	96*24	11*10*3	12*12*3	30*4	40*4	32*4*4	32*4*4	32*4*4	32*4*4
	card	1024	200	1024	32	32	–	–	32	4	32	32	32	32	32	12

Bold font indicates the best performance; italic bold font number represents the second best performance; an asterisk ($*$) designates a statistically significant improvement between the current algorithm and the best peer algorithm.

from $10, 20, 30, \ldots, h$, where h is the number of training samples; for LDA and QLDA, k is selected from $10, 20, 30, \ldots,$ ind, where ind is the number of individuals; for MPCA and MSPCA, the row (k_r) and column (k_c) dimensions are set to $k_r = k_c$, and they are selected from $1, 2, 3, \ldots, 32$, while the color dimension is chosen from $1, 2, 3$; for other 2D methods, k is set to $2, 4, \ldots, 32$.

We then need to consider the sparsity of the projection basis. For SPCA, the number of nonzero elements (cardinality, denoted by *card*) per basis vector is chosen from $50, 100, \ldots, 1024$, where $1024 = 32\times32$ is the dimension of the input samples after vectorization; for other sparse methods, *card* is set to $2, 4, \ldots, 32$. In all experiments, *card* is fixed for all basis vectors, namely, $card = card_1 = \cdots = card_k$. The sparsity of the basis vectors in 2D-PCA-L1S and G2D-PCA is controlled via the coefficients of the sparsity penalties. In our comparisons, we test all recommended coefficients.

For all competing algorithms, we report the best recognition rates, the corresponding dimension of features, and the cardinalities of basis vectors.

B. Recognition Performance

We compare the performance of 2D-QPCA and 2D-QSPCA with that of the state-of-the-arts. Generally speaking, (1) on clean color face images, 2D-QPCA is comparable to the state-of-the-arts and 2DQSPCA outperforms them; (2) on occluded color face images, 2D-QPCA is less effective, whereas 2D-QSPCA shows consistently superior performance; (3) 2D-QPCA and 2D-QSPCA are suitable for varying expressions but they are less effective for illumination and pose changes. The detailed experiment results are presented as follows.

1) Performance on Clean Face Images: The following four experiments compare the classification accuracies of different algorithms on the nonoccluded color face images. The results are reported in Table II.

1) For AR, we use the nonoccluded color face images in session one to train the projection basis. The corresponding seven images in session two are used for testing. The 2D-QSPCA has the highest recognition rate, while 2D-QPCA ranks third in this comparison.
2) On EURECOM, similar to AR, the nonoccluded color face images from session one and session two compose the training set and the testing set, respectively. The 2D-QSPCA has the best performance, followed by 2D-QPCA and MSPCA.
3) For FERET, we test the performance of different algorithms against the small sample size problem. Only one color face image per person is used for training. The 2D-QSPCA and 2D-QPCA rank first and second, respectively, validating their effectiveness when the number of training samples is small. This is because the proposed 2D-QSPCA and 2D-QPCA essentially work in the column direction of color images, and thus, the size of samples is sufficient compared with the dimension of samples. Moreover, setting $\lambda_2 > 0$, the proposed QSR model for 2D-QSPCA can further avoid the potential colinearity problem [38].
4) On CMU PIE, similar to FERET, we learn the projection basis from only one color face image per person and test the recognition performance with other six images. The 2D-QPCA and 2D-QSPCA have comparable performance, and they are much better than their competitors.

2) Performance on Partially Occluded Face Images: To evaluate the robustness of competing algorithms, we learn the projection basis from nonoccluded color face images and examine their performance on classifying color face images with 1) natural occlusions and 2) synthetic occlusions.

In the first experiment, we examine their robustness to real-world occlusions. Two databases are used. For AR,

TABLE III
EXPERIMENT RESULTS ON NATURAL-OCCLUDED COLOR FACE IMAGES

		PCA	SPCA	LDA	2D-PCA	2D-PCA-L1	2D-PCA-L1S	G2D-PCA	MPCA	MSPCA	QPCA	QLDA	R2D-QPCA	2D-BPCA	2D-QPCA	2D-QSPCA
AR	rate	0.2958	0.4442	0.15	0.7575	0.7592	0.7592	0.7592	0.5283	0.6667	0.4575	0.1867	***0.7817***	0.675	0.355	**0.845***
	dim	100	100	100	96*28	96*20	96*20	96*16	18*12*3	28*28*3	180*4	100*4	32*20*4	32*32*4	32*20*4	32*20*4
	card	1024	50	1024	32	32	–	–	32	6	32	32	32	32	32	4
EU	rate	0.3846	0.4455	0.1801	0.5032	0.5096	0.5513	0.5673	0.4281	0.5289	0.4327	0.1897	0.5096	0.5	***0.609***	**0.6987***
	dim	90	76	20	96*20	96*20	96*28	96*16	15*15*3	28*28*3	100*4	30*4	32*20*4	32*24*4	32*16*4	32*20*4
	card	1024	100	1024	32	32	–	–	32	6	32	32	32	32	32	6

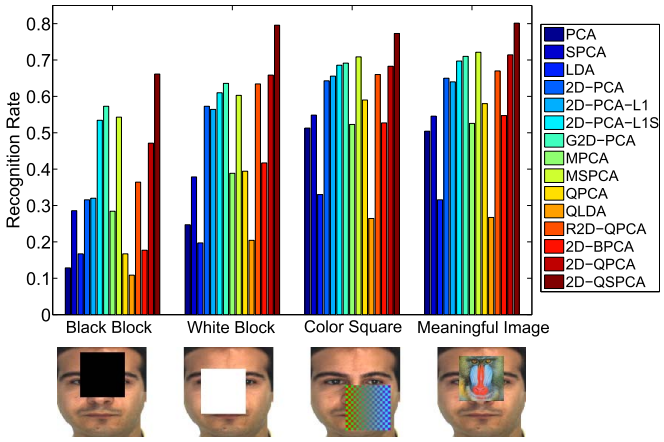

Fig. 3. Recognition rates under various contents of occlusions.

we use the 14 nonoccluded face images for training and the 12 natural-occluded images for testing (occluded by sunglasses and scarf). Similarly, for EURECOM, the training set is composed of eight nonoccluded face images per person, and the testing set contains the rest occluded face images (occluded by glasses, sunglasses, hands, and white paper). Table III presents the best recognition rates of all methods. We find that 2D-QSPCA has the best performance among all competing algorithms, and the sparse algorithms have improvements over their nonsparse counterparts.

In the following two experiments, we manually add synthetic occlusions on the testing image samples. The AR database is employed. Seven nonoccluded color face images from session one are used for training, and seven nonoccluded images from session two are added with different kinds of synthetic occlusions for testing.

We first examine the robustness of different algorithms over varying contents of occlusions. The occlusions are black blocks, white blacks, color squares, and random meaningful images, respectively. Each occlusion is scaled to 25% of the size of the color face images and is imposed at a random location, as illustrated in Figs. 3 and 4. According to Fig. 3, 2D-QSPCA achieves the best performance for all kind of occlusions.

To further investigate the robustness of competing algorithms, we add random white-and-black blocks with different sizes (from 10% to 50% of the size of the color face images) onto the testing images. The blocks are imposed at random

Fig. 4. Color face images occluded by different perceptually meaningful images.

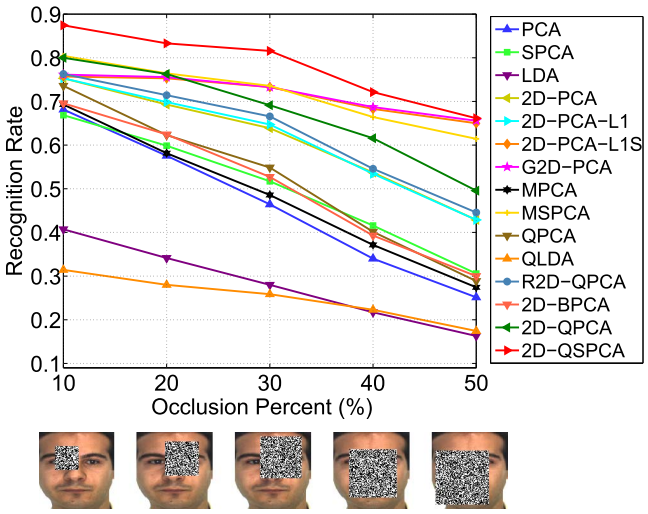

Fig. 5. Recognition rates with varying portions of occlusions.

positions. Fig. 5 shows the recognition rates of all algorithms. As can be seen, 2D-QSPCA has consistently best performance among all experiments

3) Performance Under Varying Expressions, Illumination and Poses: In this section, we examine the performance of the competing algorithms under different conditions. The AR database is used to evaluate the performance under variations in expressions and illumination. In both tests, two neural color face images per person are used for training, and the six images with corresponding variations are used for testing. GT is used to examine the performance with pose changes. For each person, we use the first seven color face images for training and the rest images for testing. The best recognition rates of different algorithms are reported in Table IV. These results show that 2D-QPCA and 2D-QSPCA are suitable for recognizing color face images with varying expressions, and that they are less effective for capturing variations of illumination and poses. In addition, operating on image columns and rows [5], [33], respectively, our 2D-QPCA and R2D-QPCA [32] are suitable to capture different patterns. Our 2D-QPCA performs better with varying expressions and illumination, while R2D-QPCA deals well with pose changes.

TABLE IV
EXPERIMENT RESULTS UNDER VARYING EXPRESSIONS, ILLUMINATION, AND POSES

		PCA	SPCA	LDA	2D-PCA	2D-PCA-L1	2D-PCA-L1S	G2DPCA	MPCA	MSPCA	QPCA	QLDA	2D-RQPCA	2D-BPCA	2D-QPCA	2D-QSPCA
Expr.	rate	0.8517	0.8517	0.4817	0.9017	0.905	0.905	0.9167	0.8517	***0.9217***	0.8583	0.4567	0.9167	0.885	0.9133	**0.9433**
	dim	90	100	90	32*12	32*12	32*24	32*4	18*10*3	24*24*3	90*4	80*4	32*12*4	32*32*4	32*8*4	32*24*4
	card	1024	100	1024	32	32	–	–	32	8	32	32	32	32	32	2
Illu.	rate	0.545	0.67	0.3583	0.3517	0.345	0.345	0.4367	0.5217	**0.7833**	0.55	0.34	0.3267	0.135	0.6617	***0.7083***
	dim	100	90	90	32*28	32*28	32*28	32*20	24*20*3	28*28*3	50*4	100*4	32*32*4	32*32*4	32*32*4	32*24*4
	card	1024	200	1024	32	32	–	–	32	8	32	32	32	32	32	20
Pose	rate	0.72	0.7375	0.58	0.765	0.77	0.78	***0.8125***	0.7575	0.8025	0.75	0.5825	0.775	**0.817**	0.76	0.755
	dim	50	70	20	32*4	32*4	32*4	32*20	10*10*3	8*8*3	30*4	10*4	32*4*4	32*20*4	32*4*4	32*12*4
	card	1024	50	1024	32	32	–	–	32	4	32	32	32	32	32	8

TABLE V
NOTATION SUMMARY FOR MCNEMAR'S TEST

Notation	Description
SF	2D-QSPCA recognizes the image and the best peer algorithm does not.
FS	the best peer algorithm recognizes the image and 2D-QSPCA does not.
n_{SF}	the number of times SF is observed.
n_{FS}	the number of times FS is observed.
n	$n = n_{SF} + n_{FS}$.
H_0	the null hypothesis.
p	the probability value (p-value).

C. Statistical Evaluation of Recognition Performance

Section V-B demonstrates that 2D-QPCA is comparable to the state-of-the-art methods and 2D-QSPCA outperforms them in most cases. In this section, we statistically evaluate whether the improvement of 2D-QSPCA over the best competing algorithm is significant. Following [51], the McNemar's test is used for the comparison of different algorithms. The notations for McNemar's test are listed in Table V.

We extract SF and FS from 2D-QSPCA and the best peer algorithm. The null hypothesis H_0 is the probability of observing SF equals FS. When the p-value of H_0 is small enough, H_0 will be rejected and the probability of observing SF will be statistically higher than that of observing FS. Thus, 2D-QSPCA obtains statistically significant improvement over the best peer algorithm. The p-value of H_0 is calculated as

$$p = \sum_{i=0}^{n_{\text{FS}}} \frac{n!}{i!(n-1)!} 0.5^n. \tag{30}$$

As recommended in [51], H_0 will be rejected when $p < 1e-4$. For the experiments in Section V-B, we report the number of times when SF and FS are observed and the corresponding p-values in Table VI. The experiments on varying illumination and poses are excluded since 2D-QSPCA does not yield the best performance. The results of McNemar's test indicate that 2D-QSPCA is significantly better than the best peer algorithm in most cases.

VI. DISCUSSION AND COMPARISONS

To provide a comprehensive understanding of the proposed algorithms, this section discusses the properties of 2D-QPCA and 2D-QSPCA and compares them with several newly proposed algorithms.

A. Properties of 2D-QPCA and 2D-QSPCA

In this section, we first explore the convergence property of 2D-QSPCA. Moreover, we also compare 2D-QPCA with 2D-QSPCA considering their: 1) computation costs; 2) discriminability for classification; and 3) relationship between classification accuracy and feature dimension. These observations coincide the experiments in Section V-B. That is, for classification: 1) the sparse regularization in 2D-QSPCA improves its discriminability and 2) the best recognition rates of 2D-QPCA and 2D-QSPCA are generally obtained with approximately half of the original dimensions.

1) Convergence of 2D-QSPCA: To solve 2D-QSPCA, we develop an alternating minimization algorithm (Algorithm 3). Fig. 6(a) and (b) verifies that Algorithm 3 converges within a few iterations. In each iteration, k subproblems are optimized using Algorithm 2, which is designed under the complex ADMM framework. The convergence of the complex ADMM algorithm has been studied in [40]. Fig. 6(c) and (d) confirms the convergence of Algorithm 2.

2) Computational Cost: To evaluate the efficiency of the proposed algorithms, we examine the computational costs of 2D-QPCA and 2D-QSPCA. For 2D-QPCA (Algorithm 1), it requires $\mathcal{O}(hnm^2)$ operations to calculate $\dot{\Phi}$ and the cost of applying QED to $\dot{\Phi}$ is at most $\mathcal{O}(m^3)$. Therefore, the total computational cost of 2D-QPCA is $\mathcal{O}(hnm^2 + m^3)$. Considering 2D-QSPCA (Algorithm 3), the computational cost of each step is listed in Table VII. As we can see, compared with the computational cost of β_j-update: 1) the costs of updating $\mathbf{Z}$, $\mathbf{y}$, and adjusting ρ are negligible and 2) the cost of β-update is negligible compared with that of α-update. Suppose that the number of iterations of alternating minimization is T_1, the number of iterations for β-update is T_2, and there are k basis vectors, the total computational cost of 2D-QSPCA is $\mathcal{O}(hnm^2 + kT_1T_2m^3)$.

TABLE VI
MCNEMAR'S TEST ON 2D-QSPCA AND THE BEST PEER ALGORITHM

		AR	EU	FERET	PIE	AR_occ	EU_occ	Black Block	White Block	Color Square	Meaningful Image	Occ= 10%	Occ= 20%	Occ= 30%	Occ= 40%	Occ= 50%	Expr.
G2D-PCA	SF/FS	–	–	13/6	24/4	–	56/15	100/37	137/20	–	–	–	–	–	58/34	46/40	–
	p-value	–	–	8.35e-2	**8.99e-5**	–	**5.20e-7**	**3.50e-8**	**6.15e-23**	–	–	–	–	–	8.00e-3	2.95e-1	–
MSPCA	SF/FS	47/8	18/5	–	–	–	–	–	–	58/23	72/16	61/12	65/17	62/15	–	–	20/7
	p-value	**4.03e-8**	5.30e-3	–	–	–	–	–	–	**6.33e-5**	**5.96e-10**	**2.40e-9**	**4.48e-8**	**3.04e-8**	–	–	9.60e-3
R2D-QPCA	SF/FS	–	–	–	–	91/45	–	–	–	–	–	–	–	–	–	–	–
	p-value	–	–	–	–	**4.95e-5**	–	–	–	–	–	–	–	–	–	–	–

Bold font indicates that 2D-QSPCA obtains statistically significant improvement over current algorithm (the best peer algorithm).

TABLE VII
COMPUTATIONAL COST OF EACH STEP IN 2D-QSPCA

Rewriting	Initilization	Alternating minimization					Recover operation
		β-update				α-update	
		β_j-update	$\mathbf{Z}$-update	$\mathbf{y}$-update	ρ-update		
$\mathcal{O}(hnm^2)$	$\mathcal{O}(m^3)$	$\mathcal{O}(m^3)$	$\mathcal{O}(m)$	$\mathcal{O}(m)$	$\mathcal{O}(m)$	$\mathcal{O}(m^3)$	$\mathcal{O}(km)$

β_j is the j-th column of β, $j = 1, \ldots, k$.

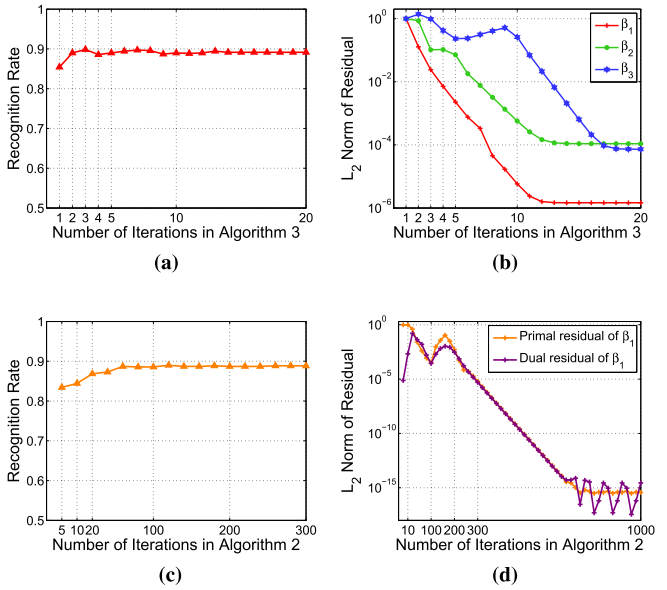

Fig. 6. Convergence analysis of 2D-QSPCA. (a) Recognition rates versus the number of iterations in Algorithm 3. (b) l_2-norm of the residuals of the first three columns of β in Algorithm 3. (c) Recognition rates versus the number of iterations in Algorithm 2. (d) l_2-norm of the residuals of the first column of β in Algorithm 2.

3) Discriminability for Classification: Projecting input samples onto their basis vectors, 2D-QPCA and 2D-QSPCA transform the color face images into different feature spaces. A discriminative feature space should contribute to the classification of color face images. That is, given a testing sample, its distances to the training samples of the same class should be small, while the distances to the training samples of different classes should be large. To quantitatively evaluate the discriminability for the task of classification, we define the relative distance θ as the ratio of two terms: the minimum distance to the uncorrect classes over the minimum distance to the correct class. If $\dot{\mathbf{X}}_i$ is a testing color face image from class c, i.e., $\dot{\mathbf{X}}_i \in c$, $\theta(\dot{\mathbf{X}}_i)$ is calculated as

$$\theta(\dot{\mathbf{X}}_i) = \frac{\min\limits_{j}[d(\dot{\mathbf{X}}_i, \dot{\mathbf{X}}_j)]}{\min\limits_{k}[d(\dot{\mathbf{X}}_i, \dot{\mathbf{X}}_k)]} \tag{31}$$

for all $\dot{\mathbf{X}}_j \notin c$ and all $\dot{\mathbf{X}}_k \in c$, $k \neq i$. $d(\dot{\mathbf{X}}_i, \dot{\mathbf{X}}_j)$ measures the l_1 norm distance between $\dot{\mathbf{X}}_i$ and $\dot{\mathbf{X}}_j$. Since we adopt the nearest-neighbor classifier for our experiments, it is obvious that if $\theta > 1$, $\dot{\mathbf{X}}_i$ will be correctly classified; otherwise, it will be misclassified. The discriminability of 2D-QPCA and 2D-QSPCA is examined under two conditions: classification on: 1) nonoccluded color face images and 2) color face images occluded by random white-and-black blocks. The blocks are set to 50% of the size of the color face images. We plot the average values of θ for all classes (persons) from AR database in Table VIII. The larger distance (θ) is the better discriminability achieves. The 2D-QSPCA achieves the largest θ using only a few nonzero elements (denoted by card) for each base. The results demonstrate that the sparsity constraints improve the discriminability of 2D-QSPCA for classification.

4) Classification Accuracy Versus Feature Dimension: The 2D-QPCA and 2D-QSPCA can be used to extract features from color images while reducing the dimension of features. To examine the relationship of the classification accuracy and the dimension of features, we record the recognition rates of 2D-QPCA and 2D-QSPCA with varying dimensions. The results are plotted in Fig. 7. Both 2D-QPCA and 2D-QSPCA achieve satisfactory recognition rates within half (less than 16*32) of the original dimensions (32*32). We observed that the recognition accuracy does not monotonously increase with the dimension of features. This is because some insignificant variations are characterized into these dimensions.

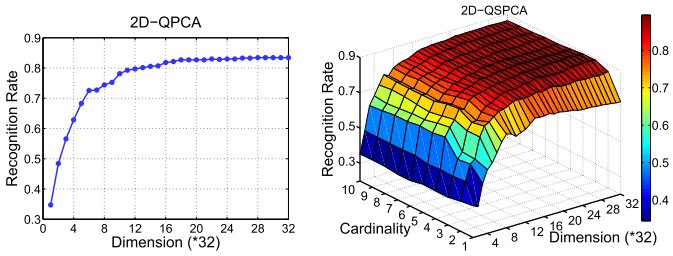

Fig. 7. Classification accuracy versus feature dimension.

TABLE VIII
DISCRIMINABILITY OF 2D-QPCA AND 2D-QSPCA FOR CLASSIFICATION

	2D-QPCA	2D-QSPCA			
		card=2	card=4	card=6	card=8
Non-occluded	1.2851	**1.3287**	1.3251	1.3182	1.3152
Occluded	0.9834	**1.043**	1.024	0.9965	0.9912

Bold font indicates the best performance.

We also find that 2D-QSPCA stably achieves good performance using basis with high sparsity, e.g., less than 10 nonzero elements (cardinality) are enough for each basis vector of 2D-QSPCA, while the basis vectors of 2D-QPCA usually contain 32 nonzero elements.

B. Comparison

In this section, we compare 2D-QPCA and 2D-QSPCA with several newly proposed methods, namely, BPCA-based methods, QLDA, and PCANet.

1) Comparison With BPCA-Based Methods: BPCA and 2D-BPCA [23] represent color images using reduced biquaternion vectors and matrices, respectively. The 2D-BPCA outperforms BPCA when applied to color images. This is because 2D-BPCA can well preserve the spatial structure of color images and is computationally more efficient by avoiding the processing of HD vectors.

Similar to 2D-BPCA, 2D-QPCA and 2D-QSPCA use 2D matrices to preserve the spatial structure of color images and to relieve the computation burden. Both quaternions and biquaternions are based on the theory of Clifford algebra. However, they have different definitions of their operations, e.g., multiplication, division, norm, and conjugation. The multiplication of quaternions is noncommutative while the multiplication of reduced biquaternions is commutative. Thus, the reduced biquaternions have the advantage in computing multiplication, whereas quaternions are more efficient when calculating the norm and conjugation [23]. Due to different mathematical operations, in the application of color face recognition, BPCA-based methods can well capture the pose changes [23] while 2D-QPCA and 2D-QSPCA are more suitable for expression and illumination variations. In addition, benefited from sparse regularization, 2D-QSPCA is robust for partial occlusions, while BPCA-based methods are not.

2) Comparison With QLDA: The 2D-QPCA and 2D-QSPCA are unsupervised techniques for dimension reduction, while QLDA [52] is a supervised technique. To extract features from color images, they all use QR to preserve the cross-channel correlation. However, they explore different perspectives of the samples. QLDA exploits the labels (classes) of the samples and aims to find the "discriminative" components to distinguish these classes [45]. Meanwhile, without considering the labels (classes) of samples, 2D-QPCA aims to obtain the "principle" components of all samples under the constraint of least-squares error; 2D-QSPCA seeks the "principle" components with an additional assumption that the projection basis should be sparse to improve classification robustness.

TABLE IX
COMPARISON OF 2D-QPCA AND 2D-QSPCA WITH PCANET

		AR	FERET	PIE	Occ= 10%	Occ= 20%	Occ= 30%	Occ= 40%	Occ= 50%
2D-QPCA	rate	0.8343	*0.8151*	*0.6961*	0.8	0.7629	0.6914	0.6157	0.4957
	dim	2560	2048	512	2560	3072	3072	2560	2560
2D-QSPCA	rate	0.8914	0.834	0.7034	0.8743	0.8329	**0.8157**	**0.7214**	**0.6614**
	dim	2560	1536	512	2560	2560	2560	2560	2560
PCANet	rate	**0.9671**	**0.9132**	**0.723**	**0.9257**	**0.8714**	0.8143	0.6372	0.4557
	dim	100352							

Since each quaternion number has four real-valued elements, the dimension of feature for 2D-QPCA and 2D-QSPCA has been multiplied by four for fair comparison.

3) Comparison With PCANet: Recently, network-based algorithms show superior performance over the statistical methods for face recognition. Among them, PCANet [53] is a deep learning network that employs the PCA technique to learn its filter banks. It benefits from the architecture of cascade networks and, thus, extracts high-dimension features. Meanwhile, 2D-QPCA and 2D-QSPCA are statistical methods to find compact representations of the data while reducing the dimension of features.

Table IX shows the comparison of the performance of 2D-QPCA and 2D-QSPCA with that of PCANet on the clean AR, FERET, PIE, and AR with different portions of occlusions. The original PCANet was trained on multiPIE [54]. However, we did not get the access to this database. For fair comparison, we train PCANet using the training sets of the aforementioned databases, as we described in Section V-B. As can be seen, PCANet outperforms 2D-QPCA and 2D-QSPCA in recognizing clean face images and the face images with moderate occlusions; however, 2D-QSPCA is more robust than PCANet when the face images suffer from a large portion of occlusions.

VII. CONCLUSION

In this paper, we developed a novel QRR model for 2D-QPCA. Different from the QCM models [25], [32] whose solutions are rigid and inflexible, the QRR model, as a general framework, can be combined with extra constraints to fit different applications. Including sparse regularization, we further proposed a QSR model for 2D-QSPCA to improve the robustness of 2D-QPCA. We also designed a novel algorithm to solve this QSR model under the framework of the complex ADMM. Extensive experiments on color face recognition verified the effectiveness of 2D-QPCA and 2D-QSPCA.

Although the successful application in color face recognition has been investigated, there are several limitations for the proposed algorithms. The 2D-QPCA is not robust for classification. The 2D-QSPCA solves this limitation. However, because 2D-QSPCA works in the column direction of color images, after sparse projection, it can discover only the underlying row patterns. Thanks for the flexibility of the QRR model, our future work will investigate the structured sparsity-induced norms [55] to regularize the QRR model to correctly identify structured patterns.

APPENDIX
PROOF OF THEOREM 1

We rewrite (13) to its equivalent complex form and prove it in complex space

$$2\left(\sum_{i=1}^{h}\|\dot{\mathbf{X}}_i-\dot{\mathbf{A}}\dot{\mathbf{B}}^*\dot{\mathbf{X}}_i\|_F^2+\lambda_2\|\dot{\mathbf{B}}\|_F^2\right)=\sum_{i=1}^{h}\|\chi_{\dot{\mathbf{X}}_i}-\chi_{\dot{\mathbf{A}}}\chi_{\dot{\mathbf{B}}^*}\chi_{\dot{\mathbf{X}}_i}\|_F^2+\lambda_2\|\chi_{\dot{\mathbf{B}}}\|_F^2. \quad (32)$$

Let

$$\alpha=\chi_{\dot{\mathbf{A}}},\quad \beta=\chi_{\dot{\mathbf{B}}},\ \phi_i=\chi_{\dot{\mathbf{X}}_i},\ i=1,\ldots,h.$$

Equation (32) can be represented as

$$\begin{aligned}&\sum_{i=1}^{h}\|\phi_i-\alpha\beta^*\phi_i\|_F^2+\lambda_2\|\beta\|_F^2\\&=\sum_{i=1}^{h}\mathrm{Tr}[\phi_i^*(\mathbf{I}-\alpha\beta^*)^*(\mathbf{I}-\alpha\beta^*)\phi_i]+\lambda_2\mathrm{Tr}(\beta\beta^*)\\&=\mathrm{Tr}\left[(\mathbf{I}-\beta\alpha^*)(\mathbf{I}-\alpha\beta^*)\left(\sum_{i=1}^{h}\phi_i\phi_i^*\right)\right]+\lambda_2\mathrm{Tr}(\beta\beta^*). \quad (33)\end{aligned}$$

Let

$$\varphi=\sum_{i=1}^{h}\phi_i\phi_i^*.$$

Equation (33) can be expressed by α, β, and φ

$$\mathrm{Tr}\varphi-\mathrm{Tr}(\alpha^*\varphi\beta)-\mathrm{Tr}(\beta^*\varphi\alpha)+\mathrm{Tr}[\beta^*(\varphi+\lambda_2\mathbf{I})\beta]. \quad (34)$$

Equation (34) is a complex-valued scalar function (the trace function yields scalar), and the optimal value of (34) can be achieved using the Wirtinger derivatives [56], [57].

For a fixed α, differentiating (34) with respect to β^* and formally treating β as a constant vector, we have

$$\frac{\partial f}{\partial\beta^*}=-\varphi\alpha+(\varphi+\lambda_2\mathbf{I})\beta. \quad (35)$$

Setting (35) to 0 gives the optimal β as

$$\hat{\beta}=(\varphi+\lambda_2\mathbf{I})^{-1}\varphi\alpha. \quad (36)$$

Substituting (36) back to (34) gives

$$\mathrm{Tr}\varphi-\mathrm{Tr}\{\alpha^*[\varphi(\varphi+\lambda_2\mathbf{I})^{-1}\varphi]\alpha\}. \quad (37)$$

Therefore,

$$\hat{\alpha}=\arg\max_{\alpha}\ \mathrm{Tr}\{\alpha^*[\varphi(\varphi+\lambda_2\mathbf{I})^{-1}\varphi]\alpha\}$$
$$\text{s.t. } \alpha^*\alpha=\mathbf{I}_{2k}. \quad (38)$$

The optimal columns of α are the eigenvectors of $\varphi(\varphi+\lambda_2\mathbf{I})^{-1}\varphi$.

Assuming that the Hermitian matrix φ admits an eigendecomposition, $\varphi=\mathbf{W}\Sigma\mathbf{W}^*$, we have

$$\varphi(\varphi+\lambda_2\mathbf{I})^{-1}\varphi=\mathbf{W}\left(\frac{\Sigma^2}{\Sigma+\lambda_2\mathbf{I}}\right)\mathbf{W}^*. \quad (39)$$

Let $\alpha=[\alpha_1,\ldots,\alpha_{2k}]$ and $\mathbf{W}=[\mathbf{w}_1,\ldots,\mathbf{w}_{2k}]$. Then, $\hat{\alpha}_j=\pm\mathbf{w}_j$, $j=1,\ldots,2k$. Recall that $\hat{\beta}=(\varphi+\lambda_2\mathbf{I})^{-1}\varphi\alpha$. Let $\beta=[\beta_1,\ldots,\beta_{2k}]$. Then, $(\hat{\beta}_j/\|\hat{\beta}_j\|_2)=\pm\mathbf{w}_j$, $j=1,\ldots,2k$.

We also have

$$\frac{1}{h}\varphi=\frac{1}{h}\sum_{i=1}^{h}\chi_{\dot{\mathbf{X}}_i}\chi_{\dot{\mathbf{X}}_i^*}=\chi_{\dot{\Phi}}. \quad (40)$$

Thus, φ and $\chi_{\dot{\Phi}}$ have the same eigenvectors. Because the columns of β are proportional to the eigenvectors of φ, they are also proportional to the eigenvectors of $\chi_{\dot{\Phi}}$, where $\dot{\Phi}=\sum_{i=1}^{h}\dot{\mathbf{X}}_i\dot{\mathbf{X}}_i^*$ is the QCM of the input quaternion samples.

Thus, we get the following results:

1) The optimal columns of β is proportional to the eigenvector of $\chi_{\dot{\Phi}}$.
2) β and $\chi_{\dot{\Phi}}$ are the complex adjoint forms of $\dot{\mathbf{B}}$ and $\dot{\Phi}$, respectively. Hence, the eigenvectors of $\dot{\mathbf{B}}$ and $\dot{\Phi}$ can be recovered from the eigenvectors of β and $\chi_{\dot{\Phi}}$.
3) The eigenvectors of $\dot{\Phi}$ are the solution of the QCM model [see (11)] of 2D-QPCA.

Let the columns of $[\dot{\mathbf{v}}_1,\ldots,\dot{\mathbf{v}}_k]$ be the projection bases of 2D-QPCA and $\dot{\mathbf{B}}=[\dot{\mathbf{b}}_1,\ldots,\dot{\mathbf{b}}_k]$. We have $\dot{\mathbf{v}}_j=(\dot{\mathbf{b}}_j/\|\dot{\mathbf{b}}_j\|_2)$, $j=1,2,\ldots,k$.

REFERENCES

[1] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.
[2] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 1991, pp. 586–591.
[3] J. Yang *et al.*, "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] J. Ye, "Generalized low rank approximations of matrices," *Mach. Learn.*, vol. 61, nos. 1–3, pp. 167–191, 2005.
[5] D. Zhang and Z.-H. Zhou, "$(2D)^2$PCA: Two-directional two-dimensional PCA for efficient face representation and recognition," *Neurocomputing*, vol. 69, no. 1, pp. 224–231, 2005.
[6] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear principal component analysis for unsupervised multilinear subspace learning," *IEEE Trans. Neural Netw.*, vol. 20, no. 11, pp. 1820–1836, Nov. 2009.
[7] 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.
[8] S.-J. Wang *et al.*, "STPCA: Sparse tensor principal component analysis for feature extraction," in *Proc. IEEE Int. Conf. Pattern Recogit.*, Nov. 2012, pp. 2278–2281.
[9] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Comput. Graph. Statist.*, vol. 15, no. 2, pp. 265–286, Jun. 2006.
[10] Z. Lai *et al.*, "Sparse approximation to the eigensubspace for discrimination," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 12, pp. 1948–1960, Dec. 2012.
[11] Z. Lai *et al.*, "Sparse alignment for robust tensor learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 10, pp. 1779–1792, Oct. 2014.
[12] Z. Lai *et al.*, "Approximate orthogonal sparse embedding for dimensionality reduction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 4, pp. 723–735, Apr. 2016.

[13] 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.
[14] H. Wang and J. Wang, "2-DPCA with L_1-norm for simultaneously robust and sparse modelling," *Neural Netw.*, vol. 46, pp. 190–198, Oct. 2013.
[15] J. Wang, "Generalized 2D principal component analysis by Lp-norm for image analysis," *IEEE Trans. Cybern.*, vol. 46, no. 3, pp. 792–803, 2016.
[16] A. Y. Ng, "Feature selection, L_1 vs. L_2 regularization, and rotational invariance," in *Proc. Int. Conf. Mach. Intell.*, 2004, p. 78.
[17] T. Li *et al.*, "F-norm distance metric based robust 2DPCA and face recognition," *Neural Netw.*, vol. 94, no. 10, pp. 204–211, 2017.
[18] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process.*, vol. 17, no. 1, pp. 53–69, Jan. 2008.
[19] R. Lan, Y. Zhou, and Y. Y. Tang, "Quaternionic local ranking binary pattern: A local descriptor of color images," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 566–579, Feb. 2016.
[20] R. Lan and Y. Zhou, "Quaternion-michelson descriptor for color image classification," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5281–5292, Nov. 2016.
[21] A. W. Yip and P. Sinha, "Contribution of color to face recognition," *Perception*, vol. 31, no. 8, pp. 995–1003, 2002.
[22] S.-J. Wang *et al.*, "Sparse tensor discriminant color space for face verification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 6, pp. 876–888, Jun. 2012.
[23] M. T. El-Melegy and A. T. Kamal, "Color image processing using reduced biquaternions with application to face recognition in a PCA framework," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Oct. 2017, pp. 3039–3046.
[24] A. M. Martinez, "The AR face database," CVC, New Delhi, India, Tech. Rep. 24, 1998.
[25] N. Le Bihan and S. J. Sangwine, "Quaternion principal component analysis of color images," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1, Sep. 2003, p. I-809.
[26] X. Xiao and Y. Zhou, "Two-dimensional quaternion sparse principle component analysis," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Apr. 2018, pp. 1528–1532.
[27] R. Lan, Y. Zhou, and Y. Y. Tang, "Quaternionic weber local descriptor of color images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 2, pp. 261–274, Feb. 2017.
[28] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 18–39, Jan. 2008.
[29] Z. Lai *et al.*, "Multilinear sparse principal component analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 10, pp. 1942–1950, Oct. 2014.
[30] P. R. Girard *et al.*, "Differential geometry revisited by biquaternion Clifford algebra," in *Proc. Int. Conf. Curves Surf.* Springer, 2014, pp. 216–242.
[31] Y. Xu *et al.*, "Vector sparse representation of color image using quaternion matrix analysis," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1315–1329, Apr. 2015.
[32] Z.-G. Jia, S.-T. Ling, and M.-X. Zhao, "Color two-dimensional principal component analysis for face recognition based on quaternion model," in *Proc. Int. Conf. Intell. Comput.* Springer, 2017, pp. 177–189.
[33] J. Yang and C. Liu, "Horizontal and vertical 2DPCA-based discriminant analysis for face verification on a large-scale database," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 4, pp. 781–792, Dec. 2007.
[34] W. R. Hamilton, "On quaternions; or on a new system of imaginaries in algebra," *Philos. Mag.*, vol. 25, no. 163, pp. 10–13, 1844.
[35] F. Zhang, "Quaternions and matrices of quaternions," *Linear Algebra Appl.*, vol. 251, pp. 21–57, Jan. 1997.
[36] R. D. Schafer, "On the algebras formed by the Cayley-Dickson process," *Amer. J. Math.*, vol. 76, no. 2, pp. 435–446, 1954.
[37] N. Le Bihan and J. Mars, "Singular value decomposition of quaternion matrices: A new tool for vector-sensor signal processing," *Signal Process.*, vol. 84, no. 7, pp. 1177–1199, Jul. 2004.
[38] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Statist. Soc. B, (Stat. Methodol.)*, vol. 67, no. 2, pp. 301–320, 2005.
[39] C. Zou, K. 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.
[40] L. Li, X. Wang, and G. Wang, "Alternating direction method of multipliers for separable convex optimization of real functions in complex variables," *Math. Probl. Eng.*, vol. 2015, Nov. 2015, Art. no. 104531.
[41] S. Boyd *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
[42] M. T. Hanna, N. P. A. Seif, and W. A. E. M. Ahmed, "Hermite-Gaussian-like eigenvectors of the discrete Fourier transform matrix based on the direct utilization of the orthogonal projection matrices on its eigenspaces," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2815–2819, Jul. 2006.
[43] N. Vakhania and G. Chelidze, "Quaternion Gaussian random variables," *Theory Probab. Appl.*, vol. 54, no. 2, pp. 363–369, 2010.
[44] Z. Su *et al.*, "Sparse envelope model: Efficient estimation and response variable selection in multivariate linear regression," *Biometrika*, vol. 103, no. 3, pp. 579–593, 2016.
[45] A. M. Martínez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.
[46] R. Min, N. Kose, and J.-L. Dugelay, "KinectFaceDB: A Kinect database for face recognition," *IEEE Trans. Syst., Man, Cybern. A, Syst.*, vol. 44, no. 11, pp. 1534–1548, Nov. 2014.
[47] P. J. Phillips *et al.*, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.
[48] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proc. Int. Conf. Autom. Face Gesture Recognit.*, 2002, pp. 53–58.
[49] A. Nefian. (2013). *Georgia Tech Face Database*. [Online]. Available: http://www.anefian.com/research/face_reco.htm
[50] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
[51] J. R. Beveridge *et al.*, "Parametric and nonparametric methods for the statistical evaluation of human ID algorithms," in *Proc. Empirical Eval. Comput. Vis. Syst.*, 2001. [Online]. Available: http://www.cs.colostate.edu/pubserv/pubs/Beveridge-draper-publications-beveridge_eecv01.pdf
[52] Y. Xu, "Quaternion-based discriminant analysis method for color face recognition," *PLoS ONE*, vol. 7, no. 8, p. e43493, 2012.
[53] T.-H. Chan *et al.*, "PCANet: A simple deep learning baseline for image classification?" *IEEE Trans. Image Process*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015.
[54] R. Gross *et al.*, "Multi-PIE," *Image Vis. Comput.*, vol. 28, no. 5, pp. 807–813, 2010.
[55] R. Jenatton, J.-Y. Audibert, and F. Bach, "Structured variable selection with sparsity-inducing norms," *J. Mach. Learn. Res.*, vol. 12, pp. 2777–2824, Feb. 2011.
[56] S. S. Haykin, *Adaptive Filter Theory*. New Delhi, India: Pearson Education, 2008.
[57] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.

Xiaolin Xiao received the B.E. degree in software engineering from Wuhan University, Wuhan, China, in 2013. She is currently pursuing the Ph.D. degree with the Department of Computer and Information Science, University of Macau, Macau, China.

Her current research interests include superpixel segmentation, saliency detection, and color image processing and understanding.

Yicong Zhou (M'07–SM'14) received the B.S. degree from Hunan University, Changsha, China, in 1992, and the M.S. and Ph.D. degrees from Tufts University, Medford, MA, USA, in 2008 and 2010, respectively, all in electrical engineering.

He is currently an Associate Professor and Director of the Vision and Image Processing Laboratory in the Department of Computer and Information Science at University of Macau, Macau, China. His research interests include chaotic systems, multimedia security, computer vision, and machine learning.

Dr. Zhou was a recipient of the Third Prize of Macau Natural Science Award in 2014. He serves as an Associate Editor for *Neurocomputing*, *Journal of Visual Communication and Image Representation*, and *Signal Processing: Image Communication*. He is a Co-Chair of Technical Committee on Cognitive Computing in the IEEE Systems, Man, and Cybernetics Society. Dr. Zhou is a senior member of the International Society for Optical Engineering (SPIE).