QUATERNION SPARSE DISCRIMINANT ANALYSIS FOR COLOR FACE RECOGNITION

Xiaolin Xiao, Yicong Zhou

Department of Computer and Information Science, University of Macau, Macau, China
shellyxiaolin@gmail.com, yicongzhou@umac.mo

ABSTRACT

To reduce feature dimensions while obtaining robust classification, in this paper, we propose quaternion sparse discriminant analysis (QSDA) for color face recognition. QSDA is formulated as a quaternion sparse regression-type model. It employs the quaternion algebra to provide an elegant and holistic way to represent color face images. The succeeding operations are directly applied to two-dimensional quaternion matrices, and hence QSDA is computationally efficient and well preserves the spatial structure of color face images. Benefited from sparsity constraints, QSDA is robust for classification. An alternating minimization algorithm is designed to solve QSDA. Experimental results demonstrate the effectiveness of QSDA for color face recognition, especially for partially occluded color face images.

Index Terms**—** QSDA, feature extraction, color face recognition, partial occlusion

1. INTRODUCTION

Linear discriminant analysis (LDA) is a supervised learning method for data classification as well as feature dimensionality reduction [1]. It can be used for face recognition. However, its performance is limited when applied to recognizing color face images. Color information is an important cue for face recognition [2, 3]. When dealing with color face images, LDA either processes different color channels separately or concatenates them into a larger matrix. This way, the cross-channel correlation of color face images is ignored.

To preserve the correlation among different color channels of face images, quaternion discriminant analysis (QDA) was proposed by extending LDA from real space to quaternion space. The quaternion algebra provides an elegant and holistic way to represent color face images by encoding the cross-channel correlation into quaternion numbers.

Nevertheless, QDA still suffers the following problems: (1) QDA converts the input color face images into high-dimensional (HD) quaternion vectors. It is computationally expensive to calculate the between-class and within-class variance matrices in the HD space. Moreover, the vectorization of color face images results in structural loss. (2) Real-world face recognition may be affected by outliers, e.g., partial occlusions. QDA is sensitive to outliers because it uses l_2 norm as measurement [4, 5].

To address these problems, in this paper, we propose quaternion sparse discriminant analysis (QSDA) for color face recognition. The advantages of QSDA are:

1) Using quaternion representation, QSDA can well preserve the cross-channel correlation of color face images;
2) QSDA directly copes with quaternion matrices rather than transforming them into HD quaternion vectors. Hence it can preserve the spatial structure of color face images and avoid extensive computation;
3) Benefited from the sparsity constraints, QSDA is robust to classify color face images with partial occlusions.

2. PRELIMINARIES

In this section, we review the quaternion algebra and quaternion discriminant analysis (QDA). The related notations are listed in Table 1.

Table 1: Summary of Notations.

Notation	Description
a, $\mathbf{a}$, $\mathbf{A}$	scalars, vectors, and matrices in real space ($\mathbb{R}$) and complex space ($\mathbb{C}$)
$\dot{a}$, $\dot{\mathbf{a}}$, $\dot{\mathbf{A}}$	scalars, vectors, and matrices in quaternion space ($\mathbb{H}$)
$(\cdot)^T$	transpose
$\bar{(\cdot)}$	conjugate
$(\cdot)^*$	transpose conjugate
$Tr(\cdot)$	trace of matrix
$Re(\cdot)$	real part of variable

2.1. Quaternions

Quaternions are a four-dimensional vector space over the field of real numbers [6]. A quaternion ($\dot{q} \in \mathbb{H}$) has one real part

This work was supported in part by the Macau Science and Technology Development Fund under Grant FDCT/016/2015/A1 and by the Research Committee at University of Macau under Grant MYRG2016-00123-FST.

978-1-5386-1737-3/18/$31.00 ©2018 IEEE

and three imaginary parts, represented by

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

where q_0, q_1, q_2, $q_3 \in \mathbb{R}$ and $\{1, i, j, k\}$ are the bases of the four dimensions whose product obey $i^2 = j^2 = k^2 = ijk = -1$. The conjugate and module of $\dot{q}$ are defined as $\bar{\dot{q}} = q_0 - q_1 i - q_2 j - q_3 k$ and $|\dot{q}| = \sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}$, respectively.

Multiplication is **non-commutative** for quaternions, i.e., $\dot{p}\dot{q} \neq \dot{q}\dot{p}$ in general, making it complicated to process quaternions [7]. To provide an effective approach for quaternion analysis, the complex adjoint form is adopted by transforming quaternions into equivalent complex pairs [6]. This transformation is given in Definition 1 and 2.

Definition 1. 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 form [6] of $\dot{\mathbf{Q}}$ is expressed by

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

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 2. Let $\dot{\mathbf{Q}} = \mathbf{Q}_a + \mathbf{Q}_b j$, $\dot{\mathbf{Q}} \in \mathbb{H}^{m\times n}$. The complex adjoint form [6] of $\dot{\mathbf{Q}}$ is formulated as

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

where $\chi_{\dot{\mathbf{Q}}} \in \mathbb{C}^{2m\times 2n}$. $\dot{\mathbf{Q}}$ and $\chi_{\dot{\mathbf{Q}}}$ are isomorphic [6].

We also define the norms of quaternion vectors and matrices to formulate our objective function in quaternion space.

Definition 3. Let $\dot{\mathbf{q}} = (\dot{q}_s) \in \mathbb{H}^m$, where $s = 1, ..., m$ is a position indicator. The l_1 norm of $\dot{\mathbf{q}}$ is defined as $\|\dot{\mathbf{q}}\|_1 = \sum_{s=1}^{m} |\dot{q}_s|$; let $\dot{\mathbf{Q}} = (\dot{q}_{s,t}) \in \mathbb{H}^{m\times n}$, where $s = 1, ..., m$ and $t = 1, ..., n$ are the position indicators. The F norm of $\dot{\mathbf{Q}}$ is defined as $\|\dot{\mathbf{Q}}\|_F = \sqrt{\sum_{s=1}^{m}\sum_{t=1}^{n} |\dot{q}_{s,t}|^2} = \sqrt{Tr(\dot{\mathbf{Q}}^*\dot{\mathbf{Q}})}$.

Properties. Let $\dot{\mathbf{P}}, \dot{\mathbf{Q}} \in \mathbb{H}^{m\times m}$. A list of facts follows [6]:

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. $2\|\dot{\mathbf{Q}}\|_F^2 = 2Tr(\dot{\mathbf{Q}}^*\dot{\mathbf{Q}}) = \|\chi_{\dot{\mathbf{Q}}}\|_F^2 = Tr(\chi_{\dot{\mathbf{Q}}^*}\chi_{\dot{\mathbf{Q}}})$

2.2. Quaternion Discriminant Analysis (QDA)

QDA seeks a set of quaternion projection bases, such that after projecting onto these bases, the distances of the projected samples from the same class are as close as possible, while the distances of the projected samples from different classes are as far as possible. QDA is effective in classifying color face images [8].

Let $\dot{\mathbf{x}}_j^i$ represent the jth quaternion sample from the ith class. Each sample is a one-dimensional quaternion vector by concatenating the rows (or columns) of quaternion image matrix. The mean quaternion sample of the ith class is denoted by $\bar{\dot{\mathbf{x}}}^i = \frac{1}{h_i}\sum_{j=1}^{h_i} \dot{\mathbf{x}}_j^i$, where h_i is the number of samples in the ith class. The mean quaternion sample of all input samples is represented by $\bar{\dot{\mathbf{x}}} = \frac{1}{c}\sum_{i=1}^{c} \bar{\dot{\mathbf{x}}}^i$, where c is the total number of classes.

Then we construct $\dot{\mathbf{S}}_b = \sum_{i=1}^{c} h_i(\bar{\dot{\mathbf{x}}}^i - \bar{\dot{\mathbf{x}}})(\bar{\dot{\mathbf{x}}}^i - \bar{\dot{\mathbf{x}}})^*$ and $\dot{\mathbf{S}}_w = \sum_{i=1}^{c}\sum_{j=1}^{h_i}(\dot{\mathbf{x}}_j^i - \bar{\dot{\mathbf{x}}}^i)(\dot{\mathbf{x}}_j^i - \bar{\dot{\mathbf{x}}}^i)^*$ to represent the between-class variance and within-class variance of input samples, respectively. Let the columns of $\dot{\mathbf{V}} = [\dot{\mathbf{v}}_1, ..., \dot{\mathbf{v}}_k]$ be the quaternion projection bases of QDA. QDA seeks optimal bases that maximizing $\dot{\mathbf{V}}^*\dot{\mathbf{S}}_b\dot{\mathbf{V}}$ while minimizing $\dot{\mathbf{V}}^*\dot{\mathbf{S}}_w\dot{\mathbf{V}}$. Thus, the objective of QDA is formulated as

$$\max_{\dot{\mathbf{V}}}(\dot{\mathbf{V}}^*\dot{\mathbf{S}}_b\dot{\mathbf{V}} - \mu\dot{\mathbf{V}}^*\dot{\mathbf{S}}_w\dot{\mathbf{V}}) = \max_{\dot{\mathbf{V}}}\{\dot{\mathbf{V}}^*(\dot{\mathbf{S}}_b - \mu\dot{\mathbf{S}}_w)\dot{\mathbf{V}}\}. \quad (1)$$

The solution of Eq. (1) equals to the leading eigenvectors of $\dot{\mathbf{S}}_b - \mu\dot{\mathbf{S}}_w$, where μ is a parameter to control the relative importance of $\dot{\mathbf{V}}^*\dot{\mathbf{S}}_b\dot{\mathbf{V}}$ and $\dot{\mathbf{V}}^*\dot{\mathbf{S}}_w\dot{\mathbf{V}}$.

Note that, in real space, LDA is optimized by maximizing $\frac{\mathbf{V}^T\mathbf{S}_b\mathbf{V}}{\mathbf{V}^T\mathbf{S}_w\mathbf{V}}$. Calculating the derivation of $\frac{\mathbf{V}^T\mathbf{S}_b\mathbf{V}}{\mathbf{V}^T\mathbf{S}_w\mathbf{V}}$ and setting it to zero, the solution of LDA reduces to the leading eigenvectors of $(\mathbf{S}_w)^{-1}\mathbf{S}_b$ [1]. However, the quaternion derivation is complicated [9] and the solution of LDA cannot be directly extended to the solution of QDA by maximizing $\frac{\dot{\mathbf{V}}^*\dot{\mathbf{S}}_b\dot{\mathbf{V}}}{\dot{\mathbf{V}}^*\dot{\mathbf{S}}_w\dot{\mathbf{V}}}$. Instead, QDA optimizes Eq. (1), which is an approximation of maximizing $\frac{\dot{\mathbf{V}}^*\dot{\mathbf{S}}_b\dot{\mathbf{V}}}{\dot{\mathbf{V}}^*\dot{\mathbf{S}}_w\dot{\mathbf{V}}}$ and facilitates the subsequent optimization by avoiding quaternion division.

3. QUATERNION SPARSE DISCRIMINANT ANALYSIS

In this section, we introduce our novel quaternion sparse regression-type model for quaternion sparse discriminant analysis (QSDA) and design an alternating minimization algorithm to solve it.

3.1. Quaternion Sparse Discriminant Analysis (QSDA)

Motivated by the sparse regression-type models in [10, 11], this section proposes a quaternion sparse regression-type model for QSDA. To protect the spatial structure of samples and to reduce the computational cost, QSDA operates directly on quaternion matrices.

Suppose there are quaternion samples from c classes and the ith class has h_i samples. Let $\dot{\mathbf{X}}_j^i \in \mathbb{H}^{m\times n}$ be the jth quaternion sample from the ith class, $\bar{\dot{\mathbf{X}}}^i = \frac{1}{h_i}\sum_{j=1}^{h_i} \dot{\mathbf{X}}_j^i$ and $\bar{\dot{\mathbf{X}}} = \frac{1}{c}\sum_{i=1}^{c} \bar{\dot{\mathbf{X}}}^i$ represent the mean quaternion sample of the

ith class and the mean quaternion sample of all input quaternion samples, respectively. Then, the between-class scatter and within-class scatter are defined as $\dot{\mathbf{S}}_b = \sum_{i=1}^{c} h_i(\bar{\dot{\mathbf{X}}}^i - \bar{\dot{\mathbf{X}}})(\bar{\dot{\mathbf{X}}}^i - \bar{\dot{\mathbf{X}}})^*$ and $\dot{\mathbf{S}}_w = \sum_{i=1}^{c}\sum_{j=1}^{h_i}(\dot{\mathbf{X}}_j^i - \bar{\dot{\mathbf{X}}}^i)(\dot{\mathbf{X}}_j^i - \bar{\dot{\mathbf{X}}}^i)^*$, respectively.

QSDA achieves its quaternion sparse bases, denoted by $\dot{\mathbf{V}}_s = [\dot{\mathbf{v}}_{s1}, ..., \dot{\mathbf{v}}_{sk}]$, as follows. Let the quaternion eigen-decomposition of $\dot{\mathbf{S}}_b - \mu\dot{\mathbf{S}}_w$ be $\dot{\mathbf{R}}\mathbf{\Sigma}\dot{\mathbf{R}}^*$ and $\dot{\mathbf{\Omega}} = \dot{\mathbf{R}}\sqrt{|\,\mathbf{\Sigma}\,|}\dot{\mathbf{R}}^*$. Find optimal $\dot{\mathbf{A}} = [\dot{\mathbf{a}}_1, ..., \dot{\mathbf{a}}_k] \in \mathbb{H}^{m\times k}$ and $\dot{\mathbf{B}} = [\dot{\mathbf{b}}_1, ..., \dot{\mathbf{b}}_k] \in \mathbb{H}^{m\times k}$ that, for any $\lambda_2 \geq 0$ and $\lambda_{1,j} \geq 0$, $j = 1, ..., k$, satisfy:

$$\min_{\dot{\mathbf{A}},\dot{\mathbf{B}}}(\|\dot{\mathbf{R}}^{-*}\dot{\mathbf{\Omega}} - \dot{\mathbf{A}}\dot{\mathbf{B}}^*\dot{\mathbf{\Omega}}\|_F^2 + \lambda_2\|\dot{\mathbf{R}}\dot{\mathbf{B}}\|_F^2 + \sum_{j=1}^{k}\lambda_{1,j}\|\dot{\mathbf{b}}_j\|_1)$$
$$\text{subject to } \dot{\mathbf{A}}^*\dot{\mathbf{A}} = \mathbf{I}_k, \tag{2}$$

Then, $\dot{\mathbf{v}}_{sj} = \frac{\dot{\mathbf{b}}_j}{\|\dot{\mathbf{b}}_j\|_2}, j = 1, ..., k$.

Eq. (2) gives the criterion of QSDA. The sparsity of the jth projection bases $\dot{\mathbf{v}}_{sj}$ is controlled by the corresponding l_1 norm penalty ($\lambda_{1,j}\|\dot{\mathbf{b}}_j\|_1$).

3.2. Solution of QSDA

Due to the non-commutativity of quaternion multiplication, solving Eq. (2) directly in quaternion space is difficult. We adopt the complex form of Eq. (2) to obtain its equivalent complex solution and convert this solution back to the quaternion space.

The quternion F-norm terms in Eq. (2) can be transformed into complex space using the F norms of their complex adjoint forms as follows. (Property 5. in Section 2.1.)

$$\begin{aligned} &2(\|\dot{\mathbf{R}}^{-*}\dot{\mathbf{\Omega}} - \dot{\mathbf{A}}\dot{\mathbf{B}}^*\dot{\mathbf{\Omega}}\|_F^2 + \lambda_2\|\dot{\mathbf{R}}\dot{\mathbf{B}}\|_F^2) \\ &= \|\chi_{\dot{\mathbf{R}}^{-*}}\chi_{\dot{\mathbf{\Omega}}} - \chi_{\dot{\mathbf{A}}}\chi_{\dot{\mathbf{B}}^*}\chi_{\dot{\mathbf{\Omega}}}\|_F^2 + \lambda_2\|\chi_{\dot{\mathbf{R}}}\chi_{\dot{\mathbf{B}}}\|_F^2. \end{aligned} \tag{3}$$

Let
$$\gamma = \chi_{\dot{\mathbf{R}}},\ \alpha = \chi_{\dot{\mathbf{A}}},\ \beta = \chi_{\dot{\mathbf{B}}},\ \phi = \chi_{\dot{\mathbf{\Omega}}}.$$

Eq. (3) can be rewritten as

$$\begin{aligned} &\|\gamma^{-*}\phi - \alpha\beta^*\phi\|_F^2 + \lambda_2\|\beta\|_F^2 \\ =&Tr[\phi^*(\gamma^{-*} - \alpha\beta^*)^*(\gamma^{-*} - \alpha\beta^*)\phi] + \lambda_2 Tr(\beta^*\gamma^*\gamma\beta) \\ =&Tr[(\gamma^{-1} - \beta\alpha^*)(\gamma^{-*} - \alpha\beta^*)(\phi\phi^*)] + \lambda_2 Tr(\beta^*\gamma^*\gamma\beta). \end{aligned} \tag{4}$$

Setting $\varphi = \phi\phi^*$, Eq. (4) can be further reduced to

$$Tr(\gamma^{-*}\varphi\gamma^{-1}) - Tr(\alpha^*\gamma^{-*}\varphi\beta) + Tr[\beta^*(\varphi + \lambda_2\gamma^*\gamma)\beta]. \tag{5}$$

We then define the complex form of quaternion l_1 norm.

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 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 $\xi(\mathbf{q}) = [\mathbf{q}_a^T; \mathbf{q}_b^T]$, where $\xi(\mathbf{q}) \in \mathbb{C}^{2\times m}$. The l_1 norm of $\dot{\mathbf{q}}$ can be represented by

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

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 $\|\mathbf{M}\|_{2,1} = \sum\limits_{j=1}^{m}\|\mathbf{M}(:,j)\|_2$.

According to the construction of the complex adjoint form (Definition 2), to recover a matrix in complex adjoint form, we only need to calculate the first half columns and then infer the other half columns from the previous ones.

Finally, Eq. (2) can be rewritten into a complex form as

$$\begin{aligned} &\min_{\alpha,\beta}\{Tr(\gamma^{-*}\varphi\gamma^{-1}) - 2Tr(\alpha^*\gamma^{-*}\varphi\beta) + \\ &\qquad Tr[\beta^*(\varphi + \lambda_2\gamma^*\gamma)\beta] + 2\sum_{j=1}^{k}\lambda_{1,j}\|\xi(\beta_j)\|_{2,1}\} \\ =&\min_{\alpha,\beta}\{\sum_{j=1}^{k}[\beta_j^*(\varphi+\lambda_2\gamma^*\gamma)\beta_j - 2Re(\alpha_j^*\gamma^{-*}\varphi\beta_j) + \lambda_{1,j}\|\xi(\beta_j)\|_{2,1}]\} \\ &\text{subject to } \alpha^*\alpha = \mathbf{I}_{2k}. \end{aligned} \tag{6}$$

We develop an alternating minimization algorithm to solve Eq.(6).

1. Fixing α, find optimal β. Eq.(6) can be solved by optimizing

$$\min_{\beta_j}[\beta_j^*(\varphi+\lambda_2\gamma^*\gamma)\beta_j - 2Re(\alpha_j^*\gamma^{-*}\varphi\beta_j) + \lambda_{1,j}\|\xi(\beta_j)\|_{2,1}]. \tag{7}$$

for $j = 1, ..., k$.

Due to the $l_{2,1}$-norm-based regularization (group lasso problem), Eq. (7) has no closed-form solution. We thus devise an algorithm to solve it under the framework of the complex ADMM algorithm [12].

Assume $\mathbf{Z} = \xi(\beta_j)$ and $vec(\mathbf{Z})$ be the vectorization of $\mathbf{Z}$. The augmented Lagrangian function of Eq. (7) is written as

$$\begin{aligned} L(\beta_j, \mathbf{Z}, \mathbf{y}) =& \beta_j^*(\varphi+\lambda_2\gamma^*\gamma)\beta_j - 2Re(\alpha_j^*\gamma^{-*}\varphi\beta_j) + \lambda_{1,j}\|\mathbf{Z}\|_{2,1} \\ &+ \mathbf{y}^*[\beta_j - vec(\mathbf{Z})] + \frac{\rho}{2}[\beta_j - vec(\mathbf{Z})]_2^2, \end{aligned} \tag{8}$$

where $\mathbf{y}$ is the Lagrangian multiplier and $\rho > 0$ is the penalty parameter. To solve $L(\beta_j, \mathbf{Z}, \mathbf{y})$, we iteratively optimize $\beta_j, \mathbf{Z}, \mathbf{y}$ one by one. Particularly, given the result of the τth iteration, the $(\tau + 1)$th iteration to optimize $L(\beta_j, \mathbf{Z}, \mathbf{y})$ is expressed by:

- Update $\beta_j^{\tau+1}$ by minimizing L w.r.t β_j. Setting the derivation of L w.r.t β_j to zero, we obtain

$$\beta_j^{\tau+1} = [\varphi + (\lambda_2 + \rho^\tau\gamma^*\gamma)]^{-1}[\varphi\gamma^{-1}\alpha_j + \rho^\tau vec(\mathbf{Z}^\tau) - \mathbf{y}^\tau]. \tag{9}$$

- Update $\mathbf{Z}^{\tau+1}$ by minimizing L w.r.t $\mathbf{Z}$. The optimization of L equals to

$$\begin{aligned} &\min_{\mathbf{Z}}\{\frac{\rho^\tau}{2}\|\beta_j^{\tau+1} - vec(\mathbf{Z})\|_2^2 \\ &\qquad + \mathbf{y}^*[\beta_j^{\tau+1} - vec(\mathbf{Z})] + \lambda_{1,j}\|\mathbf{Z}\|_{2,1}\} \\ =&\min_{\mathbf{Z}}\{\|vec(\mathbf{Z}) - (\beta_j^{\tau+1} + \frac{\mathbf{y}^\tau}{\rho^\tau})\|_2^2 + \frac{\lambda_{1,j}}{\rho^\tau}\|\mathbf{Z}\|_{2,1}\} \\ =&\min_{\mathbf{Z}}\{\|\mathbf{Z} - vec^{-1}(\beta_j^{\tau+1} + \frac{\mathbf{y}^\tau}{\rho^\tau})\|_F^2 + \frac{\lambda_{1,j}}{\rho^\tau}\|\mathbf{Z}\|_{2,1}\}, \end{aligned} \tag{10}$$

where $vec^{-1}(\cdot)$ is the inverse of $vec(\cdot)$. Eq. (10) can be solved using Lemma 1, which is derived according to the optimization of the group lasso problem in [13, 12].

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

$$\min_{\mathbf{Z}} \{\|\mathbf{Z} - \mathbf{R}\|_F^2 + \sigma\|\mathbf{Z}\|_{2,1}\}.$$

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

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

- Update $\mathbf{y}^{\tau+1}$ by

$$\mathbf{y}^{\tau+1} = \mathbf{y}^{\tau} + \rho[\beta_j^{\tau+1} - vec(\mathbf{Z}^{\tau+1})]. \tag{11}$$

We obtain the optimal β_j (the solution of Eq. (7)) when the above algorithm converges, and then we can also obtain the optimal β_{j+k} from β_j. Finally, β is optimized after Eq. (7) is solved for $k = 1, ..., k$.

2. Fixing β, find optimal α. Given β, the minimization of Eq. (6) is equivalent to find

$$\max_{\alpha} Re[Tr(\alpha^* \gamma^{-*} \varphi \beta)]. \tag{12}$$

where α can be solved using Lemma 2 by setting $\eta = \gamma^{-*}\varphi\beta$.

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

$$\max_{\alpha} Re[Tr(\alpha^* \eta)] \tag{13}$$
$$\text{subject to } \ \alpha^*\alpha = I_k.$$

Suppose the SVD of η is $\eta = U_\eta \Sigma_\eta V_\eta^*$, then $\hat{\alpha} = U_\eta V_\eta^*$. The proof of Lemma 2 is in [14] (Theorem 4).

The solution of Eq. (6) is achieved when the alternating minimization algorithm converges. Afterwards, we transfer this complex-valued solution into the quaternion-valued solution using the operator in Definition 5 [15].

Definition 5. Let $\mathbf{c} = [c_1, ..., c_n, c_{m+1}, ..., c_{2m}]^T$, $\mathbf{c} \in \mathbb{C}^{2m}$. Define an operator $\gamma(\cdot)$ as

$$\gamma(\mathbf{c}) = [c_1, c_2, ..., c_n]^T + [c_{m+1}, c_{m+2}, ..., c_{2m}]^T j,$$

where $\gamma(\mathbf{c}) \in \mathbb{H}^m$ is in the Cayley-Dickson form.

Then, the solution of Eq. (2), $\dot{\mathbf{V}}_s = [\dot{\mathbf{v}}_{s1}, ..., \dot{\mathbf{v}}_{sk}]$, can be recovered from the optimal columns of β as $\dot{\mathbf{v}}_{sj} = \gamma(\frac{\beta_j}{\|\beta_j\|_2})$, $j = 1, ..., k$. Finally, the detail procedures of QSDA are summarized in Algorithm 1.

4. COLOR FACE RECOGNITION

We test the effectiveness of QSDA using three challenging color face databases, namely, AR database [16], Bosphorus database [16], and EURECOM Kinect database [17]. These three databases are composed of non-occluded and partially occluded color face images. All images are cropped and resized to 32*32. Examples of the cropped face images are given in Fig. 1.

Algorithm 1: QSDA

Input : Training set $\{\dot{\mathbf{X}}_j^i\}$, sparsity constraints, and stopping criterion.
Output: Optimal sparse bases $[\dot{\mathbf{v}}_{s1}, ..., \dot{\mathbf{v}}_{sk}]$.

```
Calculate Ṡ_b and Ṡ_w. Derive Ṙ and Ω̇ from Ṡ_b and Ṡ_w.
Rewrite QSDA to its complex form (Eq.(6)). Given γ, φ, find the optimal α, β.
Initialize α such that α*α = I_2k.
repeat
    (1) Fixing α, find optimal β.
    for each β_j, j = 1, ..., k, do
        Find optimal β_j under the framework of the complex ADMM algorithm.
    end
    for each β_{j+k}, j = 1, ..., k, do
        Infer optimal β_{j+k} from β_j.
    end
    (2) Fixing β, find optimal α. Let γ^{-*}φβ = UDV*, update α using α̂ = UV*.
until β converges;
for j = 1, ..., k, do
    Recover v̇_sj from β_j using v̇_sj = γ(β_j / ||β_j||_2), where γ(·) is an operator in Definition 5.
end
Output [v̇_s1, ..., v̇_sk].
```

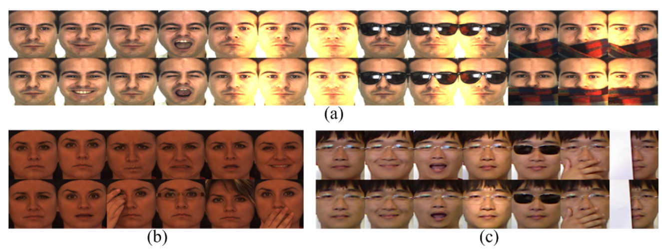

Fig. 1: Color face images of one person from (a) AR database; (b) Bosphorus database; (c) EURECOM Kinect database.

4.1. Recognition accuracy and robustness

The proposed QSDA is compared with ten state-of-the-art algorithms, namely, four PCA-based algorithms (PCA, QPCA, MPCA, and MSPCA), four discriminant analysis algorithms (LDA, QDA, TDA, and STDA), and two kernel discriminant analysis algorithms (KFDA and IKFDA).

For the parameters of QSDA, μ is selected from $10^{-4}, 10^{-3}, ..., 10^4$; $\lambda_2 = 0.001$; $\lambda_{1,j}$ $j = 1, ..., k$ are adaptively tuned so that the number of nonzero elements in each sparse bases are kept as $2, 4, ..., 32$. For the parameters of the competing algorithms, we test all recommended values and record their best results. To provide a fair comparison, (1) the recognition accuracies of all algorithms are tested with varying feature dimensions, and we select their best performance; (2) all experiments use the nearest neighbor classifier with l_1 norm distance.

In the first three experiments, we compare the recognition accuracies of different algorithms on the non-occluded color face images. For AR and EURECOM, we use the non-

Table 2: Recognition rates on non-occluded color face images.

Database	Train/Test /Person	PCA [18]	QPCA [19]	MPCA [20]	MSPCA [21]	LDA [1]	QDA [8]	TDA [22]	STDA [11]	KFDA [23]	IKFDA [24]	QSDA
AR	7/7/100	0.7486	0.8142	0.7614	0.8357	0.3857	0.4571	0.8271	0.8443	0.7971	***0.8642***	**0.8971**
Bosphorus	N/E/105	0.6711	0.8543	0.8631	0.8874	0.5188	0.7351	***0.908***	**0.9367**	0.8083	0.8145	0.8830
EURECOM	4/4/52	0.7212	0.75	0.7548	0.7933	0.4904	0.5144	0.7837	0.8029	0.8413	**0.8701**	***0.8462***

1. N: neural color face images; E: color face images with expression variants.
2. Bold type indicates the best performance; italic bold type number represents the second best performance.

Table 3: Recognition rates on partially occluded color face images.

Database	Train/Test /Person	PCA [18]	QPCA [19]	MPCA [20]	MSPCA [21]	LDA [1]	QDA [8]	TDA [22]	STDA [11]	KFDA [23]	IKFDA [24]	QSDA
AR	14/12/100	0.275	0.4575	0.5283	0.6667	0.2417	0.4250	0.68	***0.7275***	0.5291	0.5342	**0.8208***
Bosphorus	N/O/105	0.6326	0.7218	0.7375	0.7927	0.4514	0.6614	0.7428	***0.8058***	0.7321	0.7304	**0.8242**
EURECOM	8/6/52	0.3782	0.432	0.4281	0.5289	0.205	0.2243	0.5160	***0.5801***	0.4551	0.4711	**0.6987***

1. N: neural color face images; O: partially occluded color face images.
2. Bold type indicates the best performance; italic bold type number represents the second best performance; double asterisks ($*$) designates more than 10% improvement between the algorithm with the best recognition rate and the algorithm with the second best algorithm.

occluded color face images in session one to train the projection bases, and project the corresponding non-occluded color face images in session two onto those bases for testing. For Bosphorus, the neural color face images are used for training while the face images with different expressions are adopted for testing. The performance of all competing algorithms is reported in Table 2. QSDA obtains the best or comparable performance among the peer algorithms.

To evaluate the robustness of different algorithms against partial occlusions, in the following three experiments, we train the projection bases of different algorithms using non-occluded color face images and test their performance by recognizing partially occluded color face images. The results are given in Table 3. QSDA has the best recognition accuracies on all databases and shows superiority over the peer algorithms.

4.2. Interpretation of the sparse projection bases

To analyze the semantic interpretation of the sparse projection bases of QSDA, Fig. 2 visualizes the nonzero elements, represented by the non-black lines, in the first five bases. In this example, we set $\mu = 10^{-3}$ and $\lambda_2 = 0.001$. $\lambda_{1,j}, j = 1, ..., k$ are adjusted so that each sparse projection base has exactly four nonzero elements. When projecting color face images onto these sparse bases, only non-black regions are maintained. These non-black regions emphasize the chin, forehead, eyes, mouth, and cheek, individually, which coincide the discriminative regions reported in [25, 26].

Fig. 2: Visualization of nonzero elements in the first five sparse projection bases of QSDA.

5. CONCLUSION

In this paper, we developed QSDA to extract features and robustly classify patterns on color face images. A quaternion sparse regression-type model was proposed for QSDA. We designed an alternating minimization algorithm to solve QSDA in the complex space. Experimental results demonstrated that QSDA outperforms its competitors when applied to color face recognition, especially on partially occluded color face images.

6. REFERENCES

[1] Peter N. Belhumeur, João P Hespanha, and David J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, 1997.

[2] Luis Torres, Jean-Yves Reutter, and Luis Lorente, "The importance of the color information in face recognition," in *IEEE Int. Conf. Image Process.* IEEE, 1999, vol. 3, pp. 627–631.

[3] Andrew W Yip and Pawan Sinha, "Contribution of color to face recognition," *Perception*, vol. 31, no. 8, pp. 995–1003, 2002.

[4] Andrew Y Ng, "Feature selection, L1 vs. L2 regularization, and rotational invariance," in *Proc. Int. Conf. Mach. Intell.* ACM, 2004, p. 78.

[5] Tao Li, Mengyuan Li, Quanxue Gao, and Deyan Xie, "F-norm distance metric based robust 2DPCA and face recognition," *Neural Networks*, vol. 94, pp. 204–211, 2017.

[6] Fuzhen Zhang, "Quaternions and matrices of quaternions," *Linear Alg. Appl.*, vol. 251, pp. 21–57, 1997.

[7] William Rowan Hamilton, "On quaternions; or on a new system of imaginaries in algebra," *Philos. Mag. Ser. 3*, vol. 25, no. 163, pp. 10–13, 1844.

[8] Yong Xu, "Quaternion-based discriminant analysis method for color face recognition," *PloS one*, vol. 7, no. 8, pp. e43493, 2012.

[9] Dongpo Xu and Danilo P Mandic, "The theory of quaternion matrix derivatives.," *IEEE Trans. Signal Process.*, vol. 63, no. 6, pp. 1543–1556, 2015.

[10] Zhihua Qiao, Lan Zhou, and Jianhua Z Huang, "Effective linear discriminant analysis for high dimensional, low sample size data," in *Proc. World Cong. Eng.*, 2008, vol. 2, pp. 2–4.

[11] Zhihui Lai, Yong Xu, Jian Yang, Jinhui Tang, and David Zhang, "Sparse tensor discriminant analysis," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3904–3915, 2013.

[12] Lu Li, Xingyu Wang, and Guoqiang Wang, "Alternating direction method of multipliers for separable convex optimization of real functions in complex variables," *Math. Probl. Eng.*, vol. 2015, 2015.

[13] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[14] Hui Zou, Trevor Hastie, and Robert Tibshirani, "Sparse principal component analysis," *J. Comput. Graph. Stat.*, vol. 15, no. 2, pp. 265–286, 2006.

[15] Nicolas Le Bihan and Jérôme Mars, "Singular value decomposition of quaternion matrices: a new tool for vector-sensor signal processing," *Signal Prog.*, vol. 84, no. 7, pp. 1177–1199, 2004.

[16] Aleix M Martinez, "The AR face database," *CVC technical report*, 1998.

[17] Rui Min, Neslihan Kose, and Jean-Luc Dugelay, "KinectFaceDB: A kinect database for face recognition," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 44, no. 11, pp. 1534–1548, Nov 2014.

[18] Matthew A Turk and Alex P Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* IEEE, 1991, pp. 586–591.

[19] Nicolas Le Bihan and Stephen J Sangwine, "Quaternion principal component analysis of color images," in *IEEE Int. Conf. Image Process.* IEEE, 2003, vol. 1, pp. I–809.

[20] Haiping Lu, Konstantinos N Plataniotis, and Anastasios N Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 18–39, 2008.

[21] Zhihui Lai, Yong Xu, Qingcai Chen, Jian Yang, and David Zhang, "Multilinear sparse principal component analysis," *IEEE Trans. Neural Netw. Learn.*, vol. 25, no. 10, pp. 1942–1950, 2014.

[22] Shuicheng Yan, Dong Xu, Qiang Yang, Lei Zhang, Xiaoou Tang, and Hong-Jiang Zhang, "Multilinear discriminant analysis for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 212–220, 2007.

[23] Sebastian Mika, Gunnar Ratsch, Jason Weston, Bernhard Scholkopf, and Klaus-Robert Mullers, "Fisher discriminant analysis with kernels," in *Proc. Signal Prog. Workshop.* IEEE, 1999, pp. 41–48.

[24] Zizhu Fan, Yong Xu, Ming Ni, Xiaozhao Fang, and David Zhang, "Individualized learning for improving kernel fisher discriminant analysis," *Pattern Recog.*, vol. 58, pp. 100–109, 2016.

[25] Timothy C Faltemier, Kevin W Bowyer, and Patrick J Flynn, "A region ensemble for 3-D face recognition," *IEEE Trans. Inf. Forensic Secur.*, vol. 3, no. 1, pp. 62–73, 2008.

[26] Omar Ocegueda, Shishir K Shah, and Ioannis A Kakadiaris, "Which parts of the face give out your identity?," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* IEEE, 2011, pp. 641–648.