TWO-DIMENSIONAL QUATERNION SPARSE PRINCIPLE COMPONENT ANALYSIS

Xiaolin Xiao, Yicong Zhou

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

ABSTRACT

Motivated by the facts that, (1), the spatial structure of images and the correlation among color channels are important for color face recognition, and (2), natural face images may be occluded, in this work, we propose two-dimensional quaternion sparse principle component analysis (2DQSPCA) to extract features for color face recognition. 2DQSPCA inherently takes the advantage of 2DPCA in preserving the structure of two-dimensional data, as well as the strength of quaternions in representing color images holistically. Benefited from the sparsity constraints, 2DQSPCA is robust for occlusions. Experiments demonstrate the superior performance of 2DQSPCA on color face recognition, especially with occlusions.

Index Terms— 2DPCA, quaternion, sparse, color face recognition

1. INTRODUCTION

Principal component analysis (PCA) [1] and its variants are unsupervised learning approaches for feature extraction, dimensionality reduction and pattern classification [2–4].

In the application of face recognition, PCA was adopted to represent gray-scale face images in [2]. It converts two-dimensional (2D) face images to high-dimensional (HD) vectors prior to feature extraction. To avoid the intensive computation of HD data while preserving the spatial structure of face images, 2DPCA algorithms were proposed [3–5].

Considering the color face images, PCA and 2DPCA process different color channels independently. They fail to consider the correlation among color channels. However, the correlation among color channels is important for color face recognition [6]. To address this problem, quaternion PCA (QPCA) was introduced [7]. The quaternions encode the relationship of color channels holistically, providing a nice way to process color face images. Nevertheless, QPCA converts color face images into quaternion vectors and fails to preserve the spatial structure of images. In addition, these PCA models use l_2 norm as the measurement of errors. Hence they are susceptible to outliers, e.g., occlusions [5, 8].

In this paper, we propose two-dimensional quaternion sparse PCA (2DQSPCA). Compared with previous works, the strengths of 2DQSPCA are:

1) Taking the advantages of 2DPCA and quaternion representation, 2DQSPCA preserves both the spatial structure and the color channel correlation of color face images, as well as relieving the burden of HD data processing;
2) Exploiting the sparsity constraints, 2DQSPCA is robust for occlusions.

2. QUATERNIONS

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, $\mathbf{a}$, $\mathbf{A}$. In contrast, scalars, vectors, and matrices in quaternion space ($\mathbb{H}$) are denoted by $\dot{a}$, $\dot{\mathbf{a}}$, $\dot{\mathbf{A}}$, etc. Throughout this paper, T, $\bar{\cdot}$, and $*$ denote the transpose, conjugate, and transpose conjugate of complex/quaternion variables, respectively.

Quaternions are a four-dimensional number system ($\mathbb{H}$) [9]. A quaternion number ($\dot{q} \in \mathbb{H}$) has one real part and three imaginary parts. It can be represented as, $\dot{q} = q_0 + q_1 i + q_2 j + q_3 k$, where $q_0, q_1, q_2, q_3 \in \mathbb{R}$, and $\{i, j, k\}$ are the ordered bases of the imaginary parts, satisfying:

$$i^2 = j^2 = k^2 = ijk = -1.$$
$$ij = -ji = k, \;\; jk = -kj = i, \;\; ki = -ik = j.$$

The norm of $\dot{q}$ is defined as $|\dot{q}| = \sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}$, and the conjugate of $\dot{q}$ follows $\bar{\dot{q}} = q_0 - q_1 i - q_2 j - q_3 k$. Note that the multiplication of two quaternions is **noncommutative**, i.e., $\dot{p}\dot{q} \neq \dot{q}\dot{p}$ in general. This makes it quite difficult for many quaternion operations. As pointed in [10], one effective approach to deal with quaternions is to convert them into the equivalent complex forms. The following definitions provide useful tools for analyzing quaternions.

Definition 1. Let $\dot{\mathbf{q}} = (\dot{q}_s) \in \mathbb{H}^m$, where $s = 1, ..., m$ is a position indicator. The l_1 and l_2 norms of $\dot{\mathbf{q}}$ are defined as $\|\dot{\mathbf{q}}\|_1 = \sum_{s=1}^{m} |\dot{q}_s|$, and $\|\dot{\mathbf{q}}\|_2 = (\sum_{s=1}^{m} |\dot{q}_s|^2)^{\frac{1}{2}}$, respectively.

Definition 2. Let $\dot{\mathbf{Q}} = (\dot{q}_{s,t}) \in \mathbb{H}^{m \times n}$, where $s = 1, ..., m$ and $t = 1, ..., n$ are the row and column indicators, respectively. The F norm of $\dot{\mathbf{Q}}$ is defined as $\|\dot{\mathbf{Q}}\|_F = (\sum_{s=1}^{m}\sum_{t=1}^{n} |\dot{q}_{s,t}|^2)^{\frac{1}{2}}$.

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 Grants MYRG2016-00123-FST.

Definition 3. Let $\dot{\mathbf{Q}} = \mathbf{Q}_0 + \mathbf{Q}_1 i + \mathbf{Q}_2 j + \mathbf{Q}_3 k$, $\dot{\mathbf{Q}} \in \mathbb{H}^{m\times n}$, $\mathbf{Q}_0, \mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3 \in \mathbb{R}^{m\times n}$. The Cayley-Dickson construction represents $\dot{\mathbf{Q}}$ using an ordered pair of complex matrices [10]:

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

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 4. Let $\dot{\mathbf{Q}} = \mathbf{Q}_a + \mathbf{Q}_b j$, $\dot{\mathbf{Q}} \in \mathbb{H}^{m\times n}$. The complex adjoint form of $\dot{\mathbf{Q}}$ is formulated as [10, 11]:

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

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

Let $\dot{\mathbf{P}}, \dot{\mathbf{Q}} \in \mathbb{H}^{m\times m}$. The following facts come from the definition of complex adjoint form [10]:

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

Besides, the complex adjoint form is commutative for multiplication and has been widely used for quaternion matrix analysis [10, 12].

3. 2DQSPCA

Inspired by the regression-type model of sparse PCA [13], this section proposes 2DQSPCA as a quaternion regression optimization model.

3.1. 2DQSPCA

Let $\{\dot{\mathbf{X}}_i \in \mathbb{H}^{m\times n}\}_{i=1}^h$ be a collection of quaternion samples. The objective of 2DQSPCA is to find a set of orthonormal and sparse bases, denoted by the columns of $\dot{\mathbf{V}} = [\dot{\mathbf{v}}_1, ..., \dot{\mathbf{v}}_k] \in \mathbb{H}^{m\times k}$, such that, when projecting samples onto these bases, the projected samples $\{\dot{\mathbf{V}}^*\dot{\mathbf{X}}_i\}_{i=1}^h$ have the largest scatter.

2DQSPCA obtains the optimal $[\dot{\mathbf{v}}_1, ..., \dot{\mathbf{v}}_k]$ as follows: Find $\dot{\mathbf{A}} = [\dot{\mathbf{a}}_1, ..., \dot{\mathbf{a}}_k] \in \mathbb{H}^{m\times k}$ and $\dot{\mathbf{B}} = [\dot{\mathbf{b}}_1, ..., \dot{\mathbf{b}}_k] \in \mathbb{H}^{m\times k}$, satisfying:

$$(\hat{\dot{\mathbf{A}}}, \hat{\dot{\mathbf{B}}}) = \arg\min_{\dot{\mathbf{A}},\dot{\mathbf{B}}} \Big(\sum_{i=1}^h \|\dot{\mathbf{X}}_i - \dot{\mathbf{A}}\dot{\mathbf{B}}^*\dot{\mathbf{X}}_i\|_F^2 + \lambda_2 \sum_{j=1}^k \|\dot{\mathbf{b}}_j\|_2^2 + \sum_{j=1}^k \lambda_{1,j}\|\dot{\mathbf{b}}_j\|_1\Big)$$
$$\text{subject to } \dot{\mathbf{A}}^*\dot{\mathbf{A}} = \mathbf{I}_k. \tag{1}$$

Then, $\hat{\dot{\mathbf{v}}}_j = \frac{\hat{\dot{\mathbf{b}}}_j}{\|\hat{\dot{\mathbf{b}}}_j\|}$, $j = 1, 2, ..., k$.

Eq. (1) is the proposed quaternion regression model for 2DQSPCA. We have the following statements:

1) If we restrict $\dot{\mathbf{A}} = \dot{\mathbf{B}}$ and discard the penalty terms, Eq. (1) reduces to minimize $\sum_{i=1}^h \|\dot{\mathbf{X}}_i - \dot{\mathbf{B}}\dot{\mathbf{B}}^*\dot{\mathbf{X}}_i\|_F^2$ subject to $\dot{\mathbf{B}}^*\dot{\mathbf{B}} = \mathbf{I}_k$. As pointed in [13, 14], with the F norm measurement, the optimal $\dot{\mathbf{B}}$ minimizing $\sum_{i=1}^h \|\dot{\mathbf{X}}_i - \dot{\mathbf{B}}\dot{\mathbf{B}}^*\dot{\mathbf{X}}_i\|_F^2$ equals to the one that maximizing the scatter of $\{\dot{\mathbf{B}}^*\dot{\mathbf{X}}_i\}_{i=1}^h$. Hence Eq. (1) can be regarded as the regression model of quaternion PCA.
2) Theorem 3 in [13] proved that by relaxing the constraint $\dot{\mathbf{A}} = \dot{\mathbf{B}}$ and adding l_2 norm penalties on the columns of $\dot{\mathbf{B}}$, the optimal columns of $\dot{\mathbf{B}}$ are still proportional to the bases that maximize the scatter of the projected samples. We adopt a fixed coefficient $\lambda_2 > 0$ for all columns since they will not affect the solution of Eq. (1). In fact, the l_2 terms are used to avoid the colinearity problem [15].
3) Imposing l_1 norm penalties on the columns of $\dot{\mathbf{B}}$, we find a sparse approximation of the bases that maximize the scatter of the projected samples. In our model, we exploit different coefficients $\lambda_{1,j} > 0$, $j = 1, ..., k$, for the l_1 norm penalties to provide flexible control on sparsity.

3.2. Solution of 2DQSPCA

Due to the non-commutativity of quaternion multiplication, the solution of 2DQSPCA is slightly complicated. Inspired by the isomorphism between quaternions and their complex adjoint forms, we rewrite the objective function of 2DQSPCA to a complex form, and solve it in the complex space. Finally, we convert the solution back to the quaternion space to obtain the final solution of 2DQSPCA.

Considering the model of 2DQSPCA (Eq. (1)), the F and l_2 norm measurements can be easily transformed into complex space using the complex adjoint form. And the conversion of quaternion l_1 norm is given in Definition 5.

Definition 5. Let $\dot{\mathbf{q}} = \mathbf{q}_a + \mathbf{q}_b j$, $\dot{\mathbf{q}} \in \mathbb{H}^m$, and $\mathbf{q} \in \mathbb{C}^{2m}$ be the first column of of $\chi_{\dot{\mathbf{q}}}$, i.e., $\mathbf{q} = \chi_{\dot{\mathbf{q}}}(:, 1) = [\mathbf{q}_a; -\overline{\mathbf{q}_b}]$. We define an operator $\xi(\cdot)$:

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

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

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

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

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

Hence Eq. (1) has an equivalent complex form:

$$(\hat{\alpha}, \hat{\beta}) = \arg\min_{\alpha,\beta} \ \{Tr\varphi - 2Tr(\alpha^*\varphi\beta) + Tr[\beta^*(\varphi + \lambda_2\mathbf{I})\beta] + 2\sum_{j=1}^k \lambda_{1,j}\|\xi(\beta_j)\|_{2,1}\} \tag{2}$$
$$= \arg\min_{\alpha,\beta} \ \{\sum_{j=1}^k [\beta_j^*(\varphi + \lambda_2\mathbf{I})\beta_j - 2Re(\alpha_j^*\varphi\beta_j) + \lambda_{1,j}\|\xi(\beta_j)\|_{2,1}]\} \tag{3}$$
$$\text{subject to } \alpha^*\alpha = \mathbf{I}_{2k}.$$

where $Re(\cdot)$ denotes the real part of a complex number.

We develop an alternating minimization algorithm to solve Eqs. (2)/(3).

1. Fixing α, find optimal β. Given α, Eq. (3) reduces to individually solve k sub-problems:

$$\hat{\beta}_j = \arg\min_{\beta_j} [\beta_j^*(\varphi+\lambda_2\mathbf{I})\beta_j - 2Re(\alpha_j^*\varphi\beta_j) + \lambda_{1,j}\|\xi(\beta_j)\|_{2,1}] \quad (4)$$

Due to the sum-of-norms regularization ($l_{2,1}$ norm peanlties), there is no closed-form expression for $\hat{\beta}_j$. Therefore, we obtain the solution using the complex ADMM algorithm [16]. Let $\mathbf{T} = \xi(\beta_j)$, and $vec(\mathbf{T})$ represents the vectorization of $\mathbf{T}$ by concatenating the transpose of each row into a column vector. The augmented Lagrangian function of Eq. (4) is:

$$\begin{aligned} L(\beta_j, \mathbf{T}, \mathbf{y}) =& \beta_j^*(\varphi + \lambda_2\mathbf{I})\beta_j - 2Re(\alpha_j^*\varphi\beta_j) + \lambda_{1,j}\|\mathbf{T}\|_{2,1} \\ &+ \mathbf{y}^*[\beta_j - vec(\mathbf{T})] + \frac{\rho}{2}[\beta_j - vec(\mathbf{T})]_2^2, \quad (5) \end{aligned}$$

where $\mathbf{y}$ is the Lagrangian multiplier and $\rho > 0$ is the penalty parameter. The complex ADMM algorithm for optimizing $L(\beta_j, \mathbf{T}, \mathbf{y})$ is composed of iterations:

$$\beta_j^{(\tau+1)} = \arg\min_{\beta_j} L(\beta_j, \mathbf{T}^\tau, \mathbf{y}^\tau), \quad (6)$$

$$\mathbf{T}^{(\tau+1)} = \arg\min_{\mathbf{T}} L(\beta_j^{(\tau+1)}, \mathbf{T}, \mathbf{y}^\tau), \quad (7)$$

$$\mathbf{y}^{(\tau+1)} = \mathbf{y}^{(\tau)} + \rho[\beta_j^{(\tau+1)} - vec(\mathbf{T})^{(\tau+1)}]. \quad (8)$$

Step 1 of complex ADMM: Fixing $(\mathbf{T}^{(\tau)}, \mathbf{y}^{(\tau)})$, update $\beta_j^{(\tau+1)}$ by minimizing L w.r.t β_j. This is implemented by setting the complex gradient of L w.r.t β_j to zero [17].

$$\beta_j^{(\tau+1)} = [\varphi + (\lambda_2 + \rho)\mathbf{I}]^{-1}[\varphi\alpha_j + \rho vec(\mathbf{T}^{(\tau)}) - \mathbf{y}^{(\tau)}] \quad (9)$$

Step 2 of complex ADMM: Fixing $(\beta_j^{(\tau+1)}, \mathbf{y}^{(\tau)})$, update $\mathbf{T}^{(\tau+1)}$ by minimizing L w.r.t $\mathbf{T}$. Considering $\mathbf{T}$, the optimal L is equivalent to find:

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

where $vec^{-1}(\cdot)$ is the inverse of $vec(\cdot)$. This way, L is converted to the problem that utilizes the sum-of-norms regularization on $\mathbf{T}$. It can be solved following Lemma 1 [18].

Lemma 1. If a problem considering $\mathbf{T}$ is to find:

$$\hat{\mathbf{T}} = \arg\min_{\mathbf{T}}[(\mathbf{T} - \mathbf{S})_F^2 + \gamma\|\mathbf{T}\|_{2,1}] \quad (11)$$

Then, the optimal $\mathbf{T}$ achieves at:

$$\hat{\mathbf{T}}(:, i) = \begin{cases} \frac{\|\mathbf{S}(:,i)\|_2 - \gamma}{\|\mathbf{S}(:,i)\|_2}\mathbf{S}(:, i), & \|\mathbf{S}(:, i)\|_2 > \gamma \\ \mathbf{0}, & \text{otherwise} \end{cases} \quad (12)$$

Step 3 of complex ADMM: Fixing $(\beta_j^{(\tau+1)}, \mathbf{T}^{(\tau+1)})$, update $\mathbf{y}^{(\tau+1)}$ using Eq.(8).

We obtain $\hat{\beta}_j$ when the complex ADMM algorithm converges. The same operations are performed independently for $j = 1, ..., k$. Then, due to the redundant structure of complex adjoint form, we can infer $\hat{\beta}_{j+k}$ from $\hat{\beta}_j$, $j = 1, ..., k$.

2. Fixing β, find optimal α. The minimizing of Eq. (2) w.r.t α equals to finding α that maximizes $Re[Tr(\alpha^*\varphi\beta)]$. The optimal α can be obtained following Lemma 2, and the proof of Lemma 2 is given by Theorem 4 in [13].

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

$$\begin{aligned} \hat{\alpha} = \arg\max_{\alpha} Re[Tr(\alpha^*\eta)] \quad (13) \\ \text{subject to } \alpha^*\alpha = I_{2k} \end{aligned}$$

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

The alternating minimization algorithm for solving Eqs. (2)/(3) starts at any $\alpha^*\alpha = \mathbf{I}_{2k}$. The stopping condition is the convergence of β. Afterwards, we can recover the quaternion-valued solution from the complex-valued solution using the operator defined in Definition 6 [12].

Definition 6. Let $\mathbf{c} = [c_1, c_2, ..., c_n, c_{m+1}, c_{m+2}, ..., c_{2m}]^T$, $\mathbf{c} \in \mathbb{C}^{2m}$. Define an operator $\gamma(\cdot)$:

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

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

Then, the optimal bases $[\dot{\mathbf{v}}_1, ..., \dot{\mathbf{v}}_k]$ for 2DQSPCA can be recovered from the columns of $\hat{\beta}$ using: $\hat{\dot{\mathbf{v}}}_j = \gamma(\frac{\hat{\beta}_j}{\|\hat{\beta}_j\|})$, $j = 1, ..., k$. Finally, the detailed implementation of 2DQSPCA is summarized in Algorithm 1.

Algorithm 1: 2DQSPCA

Input : Training set $\{\dot{\mathbf{X}}_i\}_{i=1}^h$, sparsity constraints, and stopping criterion.

Output: Optimal sparse bases $[\dot{\mathbf{v}}_1, ..., \dot{\mathbf{v}}_k]$.

Rewrite the model of 2DQSPCA to its complex form (Eqs. (2)/(3)). Convert the objective to finding optimal α, β that satisfy Eqs. (2)/(3).

Initialize α such that $\alpha^*\alpha = I_{2k}$.

repeat

  (1) Fixing α, find optimal β.

  for each β_j, $j = 1, ..., k$, **do**

    Find optimal $\hat{\beta}_j$ that satisfies the sparsity constraint using complex ADMM.

  end

  for each β_{j+k}, $j = 1, ..., k$, **do**

    Infer $\hat{\beta}_{j+k}$ from $\hat{\beta}_j$.

  end

  (2) Fixing β, find optimal α. Let $\varphi\beta = UDV^*$, update α using $\hat{\alpha} = UV^*$, where $\varphi = \sum_{i=1}^{h} \chi_{\dot{\mathbf{X}}_i}\chi_{\dot{\mathbf{X}}_i^*}$.

until β converges;

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

  Recover $\hat{\dot{\mathbf{v}}}_j$ from $\hat{\beta}_j$ using $\hat{\dot{\mathbf{v}}}_j = \gamma(\frac{\hat{\beta}_j}{\|\hat{\beta}_j\|})$, where $\gamma(\cdot)$ is a recover operator defined in Definition 6.

end

Table 1: Recognition rates of different algorithms on AR database.

Test	Person/Train/Test	PCA [1]	SPCA [13]	2DPCA [3]	MPCA [19]	MSPCA [8]	OSPCA [20]	G2DPCA [21]	QPCA [7]	2DQSPCA
Clean images	100/7/7	0.7486	0.7514	0.79	0.7614	0.8357	0.7571	0.7943	0.8142	**0.8914***
Realistic occlusion	100/14/12	0.275	0.4225	0.7617	0.5283	0.6667	0.4	0.7675	0.4575	**0.845****
Artificial occlusion1	100/7/7	0.1371	0.2485	0.31	0.2843	0.5429	0.322	0.5429	0.1671	**0.6614****
Artificial occlusion2	100/7/7	0.2842	0.3371	0.5128	0.3886	0.6029	0.4772	0.5986	0.3943	**0.7957****
Artificial occlusion3	100/7/7	0.5057	0.5357	0.58	0.5229	0.7	0.5445	0.6400	0.59	**0.7729****
Artificial occlusion4	100/7/7	0.4671	0.5142	0.5957	0.5257	0.7214	0.56	0.6814	0.58	**0.8014** **

Bold type indicates the best performance under the same experimental setting; one asterisk ($*$) and two asterisks ($**$) designate more than 5% improvement and more than 10% improvement between the current algorithm and the best peer algorithm, respectively.

4. APPLICATION

4.1. Color face recognition

To verify the effectiveness of 2DQSPCA on color face recognition, AR face database [22, 23] is employed. Optimal bases are learned from the training set and the testing samples are projected onto the bases afterwards. Finally, we use the nearest neighbor classifier with l_1 norm distance for classification.

We provide six tests. The first one is on the clean faces with seven images per person for training and the rest seven images per person for testing; the second one exploits 14 clean images for training, and the rest images with natural occlusions (sunglasses, or scarf) for testing. For the last four tests, each one adopts the first seven clean images for training, and other seven clean images with manually imposed occlusions for testing. The occlusions are black blocks, white blacks, color squares, and random meaningful images (we adopt seven random images in total), respectively. Each occlusion is scaled to 1/4 of the size of face images, and is imposed at a random location. Examples of the AR database and the four kinds of occlusions are given in Fig. 1.

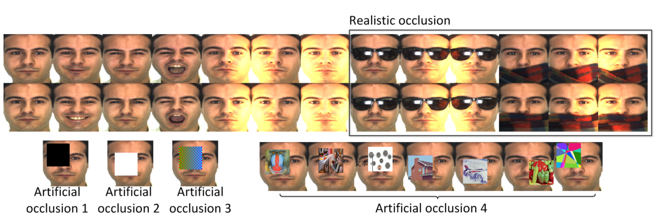

Fig. 1: Examples of AR database and the occlusions.

2DQSPCA is compared with nine state-of-the-art peer algorithms. For all competing algorithms, we obtain their recognition rates under all settings of dimensions; considering the sparse methods, we further test them under multiple sparsity constraints, and in each test, the sparsity constraints of different bases are fixed for simplicity. Generally, 2DQSPCA achieves better performance than its competitors, and impressive improvements are obtained when the testing images are occluded. Finally, the best recognition rates of all compared algorithms are reported in TABLE 1.

4.2. Analysis on the properties of 2DQSPCA

In this subsection, we investigate the properties of 2DQSPCA, including the ability of dimensionality reduction and the sparsity for the obtained bases. Fig. 2 illustrates the relationship among the recognition rate, dimensions and cardinalities (the number of nonzero elements) of bases for the experiment on clean AR faces. Considering dimensionality reduction, 2DQSPCA achieves stable good performance within half ($16*32$) of original dimensions ($32*32$). In the viewpoint of the sparsity, less than 10 nonzero elements are enough for the bases of 2DQSPCA, yielding bases with high sparsity.

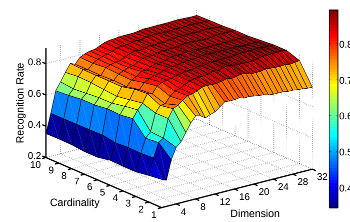

Fig. 2: Recognition rates versus dimensionality and cardinality (number of nonzero elements) on clean AR face images.

Besides, as we can see from TABLE 1, 2DQSPCA shows superior performance compared with SPCA and QPCA. This is because 2DQPCA preserves the spatial structure and the color channel correlation of color face images, and it also conducts sparse feature selection for robust recognition.

5. CONCLUSION

In this paper, we proposed 2DQSPCA to extract features for color face recognition. 2DQSPCA has advantages in processing color face images since it well preserves the spatial structure and the color channel correlation of images. Further, 2DQSPCA obtains benefit from the sparsity constraints, making it robust for occlusions. Experiments on the AR face database demonstrate that 2DQSPCA has good recognition performance for color face recognition, especially for recognizing face images with occlusions.

6. REFERENCES

[1] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, no. 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 Recog.* IEEE, 1991, pp. 586–591.

[3] J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, 2004.

[4] 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.

[5] X. Li, Y. Pang, and Y. Yuan, "L1-norm-based 2DPCA," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 4, pp. 1170–1175, 2010.

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

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

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

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

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

[11] P. Baldi and Z. Lu, "Complex-valued autoencoders," *Neural Netw.*, vol. 33, pp. 136–147, 2012.

[12] N. Le Bihan and J. 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.

[13] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Comput. Graph. Stat.*, vol. 15, no. 2, pp. 265–286, 2006.

[14] K. R. Gabriel, "Least squares approximation of matrices by additive and multiplicative models," *J. R. Stat. Soc. Ser. B-Stat.*, pp. 186–196, 1978.

[15] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Stat. Soc. Ser. B-Stat.*, vol. 67, no. 2, pp. 301–320, 2005.

[16] 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, 2015.

[17] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, 2007.

[18] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. 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.

[19] 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, 2008.

[20] K. Benidis, Y. Sun, P. Babu, and D. P. Palomar, "Orthogonal sparse PCA and covariance estimation via procrustes reformulation," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6211–6226, 2016.

[21] J. Wang, "Generalized 2-D principal component analysis by lp-norm for image analysis," *IEEE Trans. Cybern.*, vol. 46, no. 3, pp. 792–803, 2016.

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

[23] A. M. Martínez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, 2001.