IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 4, 2019, accepted October 15, 2019, date of publication October 22, 2019, date of current version November 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2948937

From Grayscale to Color: Quaternion Linear Regression for Color Face Recognition

CUIMING ZOU[1], KIT IAN KOU[2], LI DONG[3], XIANWEI ZHENG[4], AND YUAN YAN TANG[2], (Life Fellow, IEEE)

[1]School of Information Science and Engineering, Chengdu University, Chengdu 610106, China
[2]Faculty of Science and Technology, University of Macau, Macau 999078, China
[3]Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China
[4]School of Mathematics and Big Data, Foshan University, Foshan 528000, China

Corresponding author: Kit Ian Kou (kikou@umac.mo)

This work was supported in part by the Science and Technology Development Fund, Macau, under Grant SAR FDCT/085/2018/A2, in part by the National Natural Science Foundation of China under Grant 61806027, Grant 61702057, Grant 61901237, Grant 61901116, and Grant 61672114, and in part by the Department of Education of Guangdong Province, China, under Grant 2018KTSCX245.

ABSTRACT Linear regression has shown an effective tool for face recognition in recent years. Most existing linear regression based methods are devised for grayscale image based face recognition and fail to exploit the color information of color face images. To extend linear regression for color images, we propose a novel color face recognition method by formulating the color face recognition problem as a quaternion linear regression model. The proposed quaternion linear regression classification (QLRC) algorithm models each color facial image as a quaternion signal and codes multiple channels of each query color image in a *holistic* manner. Thus, the correlation among distinct channels of each color image is well preserved and leveraged by QLRC to further improve the recognition performance. To further improve QLRC, we propose a quaternion collaborative representation optimized classifier (QCROC) which integrates QLRC and quaternion collaborative representation based classifier into a unified framework. The experiments on benchmark datasets demonstrate the efficacy of the proposed approaches for color face recognition.

INDEX TERMS Quaternion representation, linear regression, color image, face recognition.

I. INTRODUCTION

Face recognition (FR) has been a hot research topic in pattern recognition for decades due to its wide applications in reality [1]–[5]. In the past two decades, a variety of FR methods have been proposed in the literatures [3], [4], [6], [7]. The nearest neighbor (NN) classifier may be the most popular FR method owing to its simplicity and efficacy. However, NN only uses one gallery image from each class to represent the query image and fails to take into account the structure and distribution of data. Deep learning based classifiers have shown impressive performance in many visual tasks including face recognition [8]–[10]. However, they often rely heavily on a large number of training samples, which are often limited and the acquisition of labeled data is generally very laborious in reality and difficult in some cases [11]. For instance, it is very difficult and costly to obtain the facial images and labels of terrorists for terrorist recognition. Their performance may be severely hampered in the presence of limited training data. Consequently, there is a need to develop effective FR methods when the training data is limited [12].

Recent research suggest that facial images of a subject under varying illumination and expression approximately lie in a low-dimensional subspace [13]. By taking advantage of such subspace structure of face images, many representation based classifiers (RC) have been developed and shown impressive performance for FR and other classification tasks in recent years [14]–[17]. For instance, inspired by the success of compressed sensing, Wright *et al.* [3] devised a sparse representation based classifier by coding the query image as a sparse linear combination of gallery images from all classes. To improve the efficiency, the collaborative representation based classifier (CRC) [7] was proposed by formulating the FR problem as a regularized least squares problem. The collaborative representation optimized classifier (CROC) [15] integrates the idea of CRC and nearest subspace classifier [6] to further improve the recognition performance. Naseem *et al.* [6] put forward a linear regression

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

classification (LRC) method for face recognition. To be specific, LRC belongs to the category of nearest subspace classifiers and tries to compute the minimal distance to the subspace spanned by all gallery images from each class. However, LRC is originally designed for *grayscale* image based face recognition and can only be applied on each color channel of color facial images individually and separately. Accordingly, the correlation information among distinct color channel may be severely discarded during coding the query image by LRC. In our previous work [12], we extend SRC and CRC to the quaternion field and apply them into face recognition. Quaternion has shown to be a promising mathematical framework in a variety of color image processing tasks, including but not limited to color image denoising [18], super-resolution [19] and classification [20]–[23], color image quality assessment [24] and remote sensing [25]. Thus, the color information is of great importance and can be leveraged to further improve the performance for face recognition.

A. PAPER CONTRIBUTION

In this work, we extend the linear regression classification algorithm to the quaternion field and apply it for color face recognition. In summary, we make the following contributions:

1. We propose a quaternion linear regression classification (QLRC) approach for color face recognition. The QLRC approach codes multiple color channels of the query color image in a *holistic* manner by formulating the problem as a quaternion linear regression model. In contrast, LRC and most of previous RC methods code different color channels of the color image *individually*.
2. To further improve QLRC, we develop a quaternion collaborative representation optimized classifier (QCROC) by integrating QLRC and quaternion collaborative representation (QCR) into a unified framework. We show that QCROC can take advantage of both merits of QLRC and QCR and outperforms them for face recognition.
3. We devise an efficient algorithm to compute the closed-form solution of QLRC and QCROC. This makes us get rid of computing the highly sophisticated gradient of quaternion matrix function due to the noncommutativity of quaternion multiplication [26].

B. PAPER ORGANIZATION

The remainder of the paper is structured as follows. In Section II, we introduce some basic facts of quaternion algebra. In Section III, we depict the proposed method for color face recognition. Section IV presents the experimental results. Eventually, Section V concludes the paper.

II. PRELIMINARIES

A. QUATERNION ALGEBRA

In this subsection, we introduce some basic facts of quaternion algebra. To enhance the readability, we first illustrate the notations used in this paper. To be specific, scalars, vectors and matrices are denoted by italic letters (e.g., x), boldface lower-case letters (e.g., $\mathbf{x}$), and boldface upper-case letters (e.g., $\mathbf{X}$), respectively. A quaternion is denoted by italic letters with a dot on the above (e.g., $\dot{q}$). Similarly, a quaternion vector and a quaternion matrix are denoted by $\dot{\mathbf{x}}$ and $\dot{\mathbf{X}}$, respectively.

The quaternion space denoted by $\mathbb{H}$ was first proposed by Hamilton [27] as a generalization of the traditional complex space $\mathbb{C}$ with three imagery units i, j and k. A quaternion $\dot{q} \in \mathbb{H}$ is defined as

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

where $q_i \in \mathbb{R}$ $(i = 0, 1, 2, 3)$ and the three imagery units obey the quaternion rules

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

$\dot{q}$ is said to be a pure quaternion if $q_0 = 0$. The most important characteristic of quaternion algebra is that **the product of quaternions are noncommutative**, i.e, $\dot{p}\dot{q} \neq \dot{q}\dot{p}$. The conjugate $\bar{\dot{q}}$ of $\dot{q}$ is defined as

$$\bar{\dot{q}} = q_0 - q_1\mathrm{i} - q_2\mathrm{j} - q_3\mathrm{k}. \tag{3}$$

The modulus $|\dot{q}|$ is defined as $|\dot{q}| = \sqrt{\dot{q}\bar{\dot{q}}}$. Analogous to vector of real numbers, a quaternion vector $\dot{\mathbf{v}} \in \mathbb{H}^n$ is denoted as $\dot{\mathbf{v}} = [\dot{v}_1, \cdots, \dot{v}_n]$ where $\dot{v}_i$ is the i-th coordinate of $\dot{\mathbf{v}}$. The inner product of two quaternion vectors $\dot{\mathbf{u}}$ and $\dot{\mathbf{v}}$ is defined by

$$\langle \dot{\mathbf{u}}, \dot{\mathbf{v}} \rangle = \dot{\mathbf{u}}^H \dot{\mathbf{v}} = \sum_{i=1}^{n} \bar{\dot{u}}_i \dot{v}_i \in \mathbb{H},$$

where $\dot{\mathbf{u}}^H = [\bar{\dot{u}}_1, \cdots, \bar{\dot{u}}_n]^T$ denotes the conjugate transpose of $\dot{\mathbf{u}}$. Then, the ℓ_2 norm of $\dot{\mathbf{v}}$ can be written as $\|\dot{\mathbf{v}}\|_2 = \sqrt{\langle \dot{\mathbf{v}}, \dot{\mathbf{v}} \rangle}$. Similarly, a quaternion matrix $\dot{\mathbf{M}} \in \mathbb{H}^{m \times n}$ is a matrix of quaternion $\dot{\mathbf{M}} = (\dot{M}_{ij})$ where $\dot{M}_{ij} \in \mathbb{H}$. To derive efficient quaternion algorithms, we define two quaternion operators $\mathcal{P}$ and $\mathcal{Q}$ [12]. For $\dot{\mathbf{M}} = \mathbf{M}_0 + \mathbf{M}_1\mathrm{i} + \mathbf{M}_2\mathrm{j} + \mathbf{M}_3\mathrm{k} \in \mathbb{H}^{m \times n}$, the operator $\mathcal{P} : \mathbb{H}^{m \times n} \to \mathbb{R}^{4m \times 4n}$ is defined as

$$\mathcal{P}(\dot{\mathbf{M}}) := \begin{bmatrix} \mathbf{M}_0 & -\mathbf{M}_1 & -\mathbf{M}_2 & -\mathbf{M}_3 \\ \mathbf{M}_1 & \mathbf{M}_0 & -\mathbf{M}_3 & \mathbf{M}_2 \\ \mathbf{M}_2 & \mathbf{M}_3 & \mathbf{M}_0 & -\mathbf{M}_1 \\ \mathbf{M}_3 & -\mathbf{M}_2 & \mathbf{M}_1 & \mathbf{M}_0 \end{bmatrix} \in \mathbb{R}^{4m \times 4n}. \tag{4}$$

For any quaternion vector $\dot{\mathbf{v}} = \mathbf{v}_0 + \mathbf{v}_1\mathrm{i} + \mathbf{v}_2\mathrm{j} + \mathbf{v}_3\mathrm{k} \in \mathbb{H}^n$, define the operator $\mathcal{Q} : \mathbb{H}^n \to \mathbb{R}^{4n}$ such that

$$\mathcal{Q}(\dot{\mathbf{v}}) := \left[\mathbf{v}_0^T, \mathbf{v}_1^T, \mathbf{v}_2^T, \mathbf{v}_3^T\right]^T \in \mathbb{R}^{4n}. \tag{5}$$

The inverse of $\mathcal{Q}$ is defined as $\mathcal{Q}^{-1} : \mathbb{R}^{4n} \to \mathbb{H}^n$, which transforms a real valued vector into a quaternion valued vector.

B. LINEAR REGRESSION CLASSIFICATION

In this subsection, we briefly review the real valued linear regression classification (LRC) [6] algorithm, which is originally designed for grayscale image based face recognition.

For each grayscale facial image $\mathbf{I} \in \mathbb{R}^{M \times N}$, we stack their pixels in a column-wise manner and obtain a vector $\mathbf{v} \in \mathbb{R}^d$ where $d = M \times N$ is the data dimension. Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ be the matrix of n vectorized gallery facial images of K subjects for training. Here $\mathbf{x}_i \in \mathbb{R}^d$

denotes the i-th vectorized training sample. For each class k, let $\mathbf{X}_k \in \mathbb{R}^{d\times n_k}$ be the matrix composed of training samples from class k. Given the training data matrix $\mathbf{X}$ and their labels, the goal is to correctly recognize any new vectorized facial image $\mathbf{y} \in \mathbb{R}^d$. Firstly, LRC computes the class-specific coefficient vector $\mathbf{c}_k$ for each class using real valued linear regression

$$\min_{\mathbf{c}\in\mathbb{R}^{n_k}} \|\mathbf{y} - \mathbf{X}_k\mathbf{c}\|_2^2, \tag{6}$$

where the optimal solution is

$$\mathbf{c}_k^{LR} = \left(\mathbf{X}_k^T\mathbf{X}_k\right)^{-1}\mathbf{X}_k^T\mathbf{y} \tag{7}$$

by setting the gradient of objective function (6) with respect to $\mathbf{c}$ as the zero vector. Secondly, we reconstruct the test sample using $\mathbf{X}_k$ and $\mathbf{c}_k^{LR}$ for each class, i.e.,

$$\hat{\mathbf{y}}_k = \mathbf{X}_k\mathbf{c}_k^{LR} = \mathbf{X}_k\left(\mathbf{X}_k^T\mathbf{X}_k\right)^{-1}\mathbf{X}_k^T\mathbf{y}, \tag{8}$$

for $k = 1, \cdots, K$. Then we calculate the distance between the test sample $\mathbf{y}$ and its reconstructed version $\hat{\mathbf{y}}_k$, i.e.,

$$d_k(\mathbf{y}) = \|\mathbf{y} - \hat{\mathbf{y}}_k\|_2, \quad k = 1, \cdots, K. \tag{9}$$

Finally, the test sample $\mathbf{y}$ is assigned to the class leading to the minimal distance.

III. PROPOSED METHOD

A. QUATERNION LINEAR REGRESSION CLASSIFICATION

Quaternion has shown to be a promising mathematical framework for color image processing [18], [28]. In this work, we model each RGB color image as a pure quaternion matrix $\dot{\mathbf{I}} \in \mathbb{H}^{M\times N}$ where M and N denote the number of rows and columns of $\dot{\mathbf{I}}$, respectively. It can be written as

$$\dot{\mathbf{I}} = \mathbf{0} + \mathbf{I}_r\mathrm{i} + \mathbf{I}_g\mathrm{j} + \mathbf{I}_b\mathrm{k},$$

where $\mathbf{I}_c \in \mathbb{R}^{M\times N}$ stands for the c channel of $\dot{\mathbf{I}}$ and $c \in \{r, g, b\}$ indicates the red, green and blue channel, respectively. For ease of formulation, it is common to stack the pixels in each channel $\mathbf{I}_c$ into a vector denoted as $\mathbf{y}_c$ for the c channel. Then each color image $\dot{\mathbf{I}} \in \mathbb{H}^{M\times N}$ is transformed into a quaternion vector $\dot{\mathbf{v}} \in \mathbb{H}^d$ ($d = M \times N$), which can be expressed as

$$\dot{\mathbf{v}} = \mathbf{0} + \mathbf{v}_r\mathrm{i} + \mathbf{v}_g\mathrm{j} + \mathbf{v}_b\mathrm{k}.$$

Consider a face recognition problem with K subjects corresponding to K classes. Assume that we are given n_k gallery color facial images $\dot{\mathbf{x}}_1^k, \cdots, \dot{\mathbf{x}}_{n_k}^k$ from the k-th class where $k = 1, \cdots, K$. Here $\dot{\mathbf{x}}_i^k \in \mathbb{H}^d$ denotes the i-th vectorized color image from class k. For ease of statement, these images from class k are arranged as columns of a quaternion matrix $\dot{\mathbf{X}}_k = [\dot{\mathbf{x}}_1^k, \dot{\mathbf{x}}_2^k, \cdots, \dot{\mathbf{x}}_{n_k}^k] \in \mathbb{H}^{d\times n_k}$. Let $\dot{\mathbf{X}} = [\dot{\mathbf{X}}_1, \dot{\mathbf{X}}_2, \cdots, \dot{\mathbf{X}}_K]$ be the quaternion matrix of all gallery color images. The goal is to correctly recognize any new query color facial image $\dot{\mathbf{y}} \in \mathbb{H}^d$.

With the notations above, we depict the proposed QLRC (Quaternion Linear Regression Classification) method for color face recognition as follows. The columns of $\dot{\mathbf{X}}$ are normalized to have unit quaternion ℓ_2 norm to reduce the effect of the scale of data. Firstly, we attempt to represent the query color image $\dot{\mathbf{y}}$ as a quaternion linear combination of the gallery images from each class. To this end, we consider the following quaternion linear regression (QLR) model

$$\min_{\dot{\mathbf{c}}\in\mathbb{H}^{n_k}} \left\|\dot{\mathbf{y}} - \dot{\mathbf{X}}_k\dot{\mathbf{c}}\right\|_2^2, \quad k = 1, 2, \cdots, K. \tag{10}$$

To avoid computing the complicated gradient of quaternion function, we transform the QLR model (10) into the following equivalent form

$$\min_{\dot{\mathbf{c}}\in\mathbb{H}^{n_k}} \left\|\mathcal{Q}(\dot{\mathbf{y}}) - \mathcal{P}(\dot{\mathbf{X}}_k)\mathcal{Q}(\dot{\mathbf{c}})\right\|_2^2, \quad k = 1, 2, \cdots, K. \tag{11}$$

The equivalence between Eq. (10) and Eq. (11) is formally phrased in the following theorem and proved in the appendix.

Theorem 1: The problems (10) and (11) are equivalent and their optimal solutions are identical.

The closed-form solution of Eq. (11) can be obtained by setting the gradient of the objective function with respect to $\mathcal{Q}(\dot{\mathbf{c}})$ to zero and we have

$$\dot{\mathbf{c}}_k^{QLR} = \mathcal{Q}^{-1}\left(\left(\mathcal{P}\left(\dot{\mathbf{X}}_k\right)^T\mathcal{P}\left(\dot{\mathbf{X}}_k\right)\right)^{-1}\mathcal{P}\left(\dot{\mathbf{X}}_k\right)^T\mathcal{Q}\left(\dot{\mathbf{y}}\right)\right). \tag{12}$$

Secondly, we compute the class-specific reconstruction of the query color image

$$\hat{\dot{\mathbf{y}}}_k = \dot{\mathbf{X}}_k\dot{\mathbf{c}}_k^{QLR}, \quad k = 1, \cdots, K.$$

Then the distance between the query image and the class-specific reconstruction can be calculated as

$$d_k(\dot{\mathbf{y}}) = \left\|\dot{\mathbf{y}} - \hat{\dot{\mathbf{y}}}_k\right\|_2, \quad k = 1, \cdots, K.$$

Algorithm 1 Quaternion Linear Regression Classification (QLRC)

Input: Quaternion training color image matrix $\dot{\mathbf{X}}_k \in \mathbb{H}^{d\times n_k}$, $k = 1, \cdots, K$ and a test color image vector $\dot{\mathbf{y}} \in \mathbb{H}^{d\times 1}$.

Output: identity($\dot{\mathbf{y}}$).

1: Normalize the columns of $\dot{\mathbf{X}}_k$, $k = 1, \cdots, K$ to have unit quaternion ℓ_2 norm.

2: Compute the quaternion linear regression vectors $\left\{\dot{\mathbf{c}}_k^{QLR}, k = 1, \cdots, K\right\}$ using Eq. (12).

3: Compute the class-specific reconstruction

$$\hat{\dot{\mathbf{y}}}_k = \dot{\mathbf{X}}_k\dot{\mathbf{c}}_k^{QLR}, \quad k = 1, \cdots, K.$$

3: Calculate the distance between the test image and the class-specific reconstruction

$$d_k(\dot{\mathbf{y}}) = \left\|\dot{\mathbf{y}} - \hat{\dot{\mathbf{y}}}_k\right\|_2, \quad k = 1, \cdots, K.$$

4: Output identity($\dot{\mathbf{y}}$) = $arg\min_k d_k(\dot{\mathbf{y}})$.

Eventually, the query color image $\dot{\mathbf{y}}$ is assigned to the class minimizing the distance

$$\text{identity}(\dot{\mathbf{y}}) = arg\min_k d_k(\dot{\mathbf{y}}).$$

Algorithm 1 summarizes the complete procedure of the proposed QLRC method for color face recognition.

B. QUATERNION COLLABORATIVE REPRESENTATION OPTIMIZER CLASSIFICATION (QCROC)

In this subsection, we aim to further improve QLRC by integrating it and quaternion collaborative representation (QCR) into a unified framework. Unlike QLRC coding the test sample using training samples in each class independently, QCR codes it using training samples from all classes in a collaborative way. Formally, the model of QCR is written as

$$\min_{\dot{\mathbf{c}}\in\mathbb{H}^n} \left\|\dot{\mathbf{y}} - \dot{\mathbf{X}}\dot{\mathbf{c}}\right\|_2^2 + \lambda\|\dot{\mathbf{c}}\|_2^2, \tag{13}$$

where λ is a positive regularization parameter. Analogous to the equivalence between Eq. (10) and Eq. (11), we can also rewritten Eq. (13) in the following equivalent form

$$\min_{\dot{\mathbf{c}}\in\mathbb{H}^{n_k}} \left\|\boldsymbol{\mathcal{Q}}(\dot{\mathbf{y}}) - \boldsymbol{\mathcal{P}}(\dot{\mathbf{X}})\boldsymbol{\mathcal{Q}}(\dot{\mathbf{c}})\right\|_2^2 + \lambda\|\boldsymbol{\mathcal{Q}}(\dot{\mathbf{c}})\|_2^2, \tag{14}$$

where we used the fact that $\|\dot{\mathbf{c}}\|_2^2 = \|\boldsymbol{\mathcal{Q}}(\dot{\mathbf{c}})\|_2^2$. By setting the derivative with respect to $\boldsymbol{\mathcal{Q}}(\dot{\mathbf{c}})$ to zero, we have the closed-form solution

$$\dot{\mathbf{c}}^{QCR} = \boldsymbol{\mathcal{Q}}^{-1}\left(\left(\boldsymbol{\mathcal{P}}\left(\dot{\mathbf{X}}\right)^T\boldsymbol{\mathcal{P}}\left(\dot{\mathbf{X}}\right) + \lambda\mathbf{I}\right)^{-1}\boldsymbol{\mathcal{P}}\left(\dot{\mathbf{X}}\right)^T\boldsymbol{\mathcal{Q}}\left(\dot{\mathbf{y}}\right)\right). \tag{15}$$

To take advantage of both merits of QLRC and QCR, we compute the class-specific residual

$$r_k(\dot{\mathbf{y}}) = \gamma\left\|\dot{\mathbf{y}} - \dot{\mathbf{X}}_k\dot{\mathbf{c}}_k^{QCR}\right\|_2^2 + (1-\gamma)\left\|\dot{\mathbf{y}} - \dot{\mathbf{X}}_k\dot{\mathbf{c}}_k^{QLR}\right\|_2^2, \tag{16}$$

where $\gamma \in [0, 1]$ is parameter striking a balance between the weights of QLRC and QCRC. Eventually, the test sample $\dot{\mathbf{y}}$ is assigned to the class minimizing the residual, i.e., $\text{identity}(\dot{\mathbf{y}}) = arg\min_k r_k(\dot{\mathbf{y}})$. The complete procedure of QCROC is formally presented in Algorithm 2.

Algorithm 2 Quaternion Collaborative Representation Optimized Classification (QCROC)

Input: Quaternion training data matrix $\dot{\mathbf{X}} = [\dot{\mathbf{X}}_1, \cdots, \dot{\mathbf{X}}_K] \in \mathbb{H}^{d\times n}$, a test quaternion vector $\dot{\mathbf{y}} \in \mathbb{H}^{d\times 1}$, and the parameters γ and λ.

Output: $\text{identity}(\dot{\mathbf{y}})$.

1: Normalize the columns of $\dot{\mathbf{X}}$ to have unit quaternion ℓ_2 norm.
2: Compute the quaternion linear regression vectors $\left\{\dot{\mathbf{c}}_k^{QLR}, k = 1, \cdots, K\right\}$ using Eq. (12).
3: Compute the quaternion collaborative representation vector $\{\dot{\mathbf{c}}^{QCR}\}$ using Eq. (15).
4: Calculate the class-specific residual using Eq. (16).
5: Predict $\text{identity}(\dot{\mathbf{y}}) = arg\min_k r_k(\dot{\mathbf{y}})$.

IV. EXPERIMENTS

In this section, we conduct experiments on real-world data to verify the efficacy of the proposed QLRC and QCROC for color face recognition. To begin with, we illustrate the used datasets and the competing methods. We test the proposed methods on three real-life benchmark color face image datasets and they are the SCface (Surveillance Cameras Face) dataset [29], the AR dataset [30], and the Caltech Faces dataset, respectively. The details of the datasets will be introduced in the subsequent subsections. We compare our proposed QLRC and QCROC algorithms with the nearest neighbor (NN) algorithm [31], the quaternion nearest neighbor (QNN), LRC [6], CRC [7], QCRC [12], SRC [3], QSRC [12] and CROC [15]. The QCROC method has two parameters λ and γ. The λ is tuned by 5-fold cross validation in each experiment. As for the parameter γ, we set it as $\gamma = 0.1$ in all experiments. For other competing algorithms, we use the source codes from the authors and their parameters are tuned to by 5-fold cross validation in each experiment for fair comparison.

FIGURE 1. Sample images from the SCface dataset.

A. EXPERIMENTS ON THE SCFACE DATABASE

The SCface (Surveillance Cameras Face) dataset [29] contains 4160 facial images of 130 subjects captured under uncontrolled environment using five video surveillance cameras of various visual qualities. For this reason, the images have different resolutions. In particular, many of the images have very low resolution, making the recognition become more challenging. Fig. 1 shows several sample images of a subject with varying resolutions from the SCface dataset.

1) EXPERIMENT 1–EFFECT OF THE PARAMETER γ

In the first experiment, we compare the performance of CROC and QCROC for color face recognition on the SCface dataset with varying value of the parameter γ to validate the benefits of quaternion representation. We randomly select half of images per subject to compose the training set and the rest images are used for testing. For efficiency, we resize each image to 8×6 pixels. For evaluation, we randomly select half images per subject for training and the remaining images are utilized for testing. To obtain more reliable results, we repeat each experiment with 10 random runs and record the average recognition rates of each algorithm. Fig. 2 shows the average recognition rates over 10 random runs as a function of the parameter γ, which varies from 0 to 1. Note that QCROC

TABLE 1. Recognition rates (%) using varying feature dimension d on the SCface dataset. Best results are marked bold and the second best results are marked with underline.

Dimension m	NN	QNN	LRC	QLRC	CRC	QCRC	SRC	QSRC	CROC	QCROC
48	58.27	60.96	72.60	74.54	63.83	68.68	68.75	80.00	77.79	**81.44**
80	61.15	62.69	75.10	75.19	75.19	81.35	75.58	81.25	81.06	**85.29**
192	63.65	63.85	75.67	77.50	80.58	85.58	77.88	83.75	83.46	**86.25**
300	64.52	65.96	77.69	79.04	81.63	84.81	77.50	83.75	83.65	**86.92**

TABLE 2. Recognition rates (%) using varying percent p_c(%) of training samples per class on the SCface dataset over 10 random runs. Best results are marked bold and the second best results are marked with underline.

p_c(%)	Methods	NN	QNN	LRC	QLRC	CRC	QCRC	SRC	QSRC	CROC	QCROC
30	Mean	49.15	48.72	57.42	57.87	68.53	69.22	62.47	66.56	70.40	**70.75**
	Min	46.86	47.50	55.13	55.51	66.60	67.31	61.03	64.23	68.27	**68.72**
	Max	51.15	50.64	59.17	59.81	70.06	71.35	63.85	68.46	**72.63**	72.44
40	Mean	57.16	57.21	68.47	70.15	75.39	75.64	71.06	75.52	79.47	**80.11**
	Min	55.77	56.54	67.31	68.08	73.54	74.54	68.77	73.85	76.85	**78.69**
	Max	59.23	58.08	69.69	72.23	77.15	76.62	73.15	77.46	80.85	**81.69**
50	Mean	62.33	63.68	75.81	76.96	78.52	80.25	76.10	79.54	84.64	**85.14**
	Min	60.96	62.40	74.04	75.38	77.12	77.69	74.33	77.79	82.60	**83.56**
	Max	64.90	65.87	77.12	78.75	79.81	82.02	77.69	80.87	86.06	**86.15**
60	Mean	65.35	65.21	79.20	80.57	80.51	81.41	78.81	81.27	86.48	**86.87**
	Min	62.31	62.75	76.37	78.57	78.35	79.78	77.14	79.34	**85.05**	83.74
	Max	67.80	67.03	80.77	82.86	82.09	84.62	80.33	82.53	89.12	**89.23**
70	Mean	68.60	69.78	82.42	84.31	83.32	83.35	81.52	84.14	87.75	**89.00**
	Min	67.23	67.08	80.46	82.77	81.08	80.62	79.23	81.85	84.92	**86.62**
	Max	70.31	71.85	83.38	87.23	84.46	86.15	83.54	87.08	90.15	**90.46**

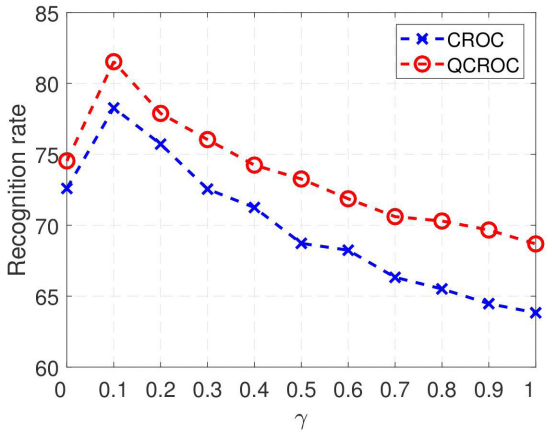


FIGURE 2. Average recognition rates of CROC and QCROC as a function of the parameter γ over 10 random runs on the SCface dataset.

outperforms CROC with better recognition performance with varying values of the parameter γ. Recall that when $\gamma = 0$, QCROC and CROC reduce to QLRC and LRC, respectively. When $\gamma = 1$, QCROC and CROC reduce to QCRC and CRC, respectively. Thus, the results also verify the superiority of QLRC over LRC and QCRC over CRC for face recognition on the SCface dataset, respectively.

2) EXPERIMENT 2–EFFECT OF FEATURE DIMENSION

In the second experiment, we analyze the effect of the data dimension to the recognition performance of competing algorithms. To that end, we downsample and resize each facial image to have 8×6, 10×8, 16×12 and 20×15 pixels, respectively. Then the data dimension is 48, 80, 192 and 300, respectively. Table 1 displays the recognition results of various classifiers with varying data dimension on the SCface dataset. As the data dimension increases, the recognition rates of each method also grows due to more useful information.

3) EXPERIMENT 3–EFFECT OF TRAINING SET SIZE

In the third experiment, we perform the competing RC methods with different sizes of training data set. To this end, we randomly choose $p_c = 30, 40, 50, 60, 70$ percent of images of each subject to form the training data set while the remaining images are utilized to evaluate the performance of the proposed approaches. The data dimension is fixed as 80 and analogous results can also be obtained using other data dimensions. To make the results more convincing, we randomly resample the training and testing data with the same size 10 times and record the mean, minimal and maximal recognition rates over the 10 random runs. The results are listed in Table 2. From the results, we make the following conclusions:

- Quaternion representation based classifiers (QRC) generally perform better than the corresponding RC methods. This indicates that the color information of color images is beneficial for face recognition.
- Note that QCROC outperforms other competing classifiers with distinct number of training samples. This indicates the robustness of QCROC towards different choices of training set.

TABLE 3. Recognition rates (%) using varying feature dimension d on the AR dataset. Best results are marked bold and the second best results are marked with underline.

Data dimension m	NN	QNN	LRC	QLRC	CRC	QCRC	SRC	QSRC	CROC	QCROC
48	41.85	43.35	72.62	80.77	68.00	78.62	68.08	77.00	78.85	**85.62**
80	43.85	45.56	77.08	81.23	85.25	88.54	76.69	86.00	87.69	**91.38**
192	44.08	48.88	80.38	86.00	95.01	95.85	89.23	93.69	92.77	**97.31**
300	46.62	49.93	83.69	87.15	96.19	96.92	91.38	93.85	95.00	**97.54**

FIGURE 3. Sample images from the AR dataset.

B. EXPERIMENTS ON THE AR DATABASE

The AR dataset [30] has more than 4000 color facial images of 126 persons. For each person, 26 images are captured in two separate sessions with distinct lighting condition, expressions and occlusions. We utilize a commonly used subset with facial images of 100 persons [3] for the experiments. Several sample images of a person from the AR dataset are shown in Fig. 3.

1) EXPERIMENT 1–EFFECT OF THE PARAMETER γ

In the first experiment, we investigate the performance of CROC and QCROC with varying value of the parameter γ for face recognition on the AR dataset. Similarly, we select half of images (i.e., 13 images) of each subject uniformly at random for training and use the remaining images for testing. Fig. 4 shows the average recognition results over 10 random runs as a regularization path when the parameter γ varies from 0 to 1. The results align with those on the SCface dataset and further corroborate the efficacy of QCROC.

2) EXPERIMENT 2–EFFECT OF FEATURE DIMENSION

The experimental settings are analogous to those of the experiments using the SC dataset. Likewise, half images per subject are randomly chosen for training and the rest are used for testing. Table 3 reports the recognition rates as the data dimension varies from 48 to 300. Note that QRC significantly improves the corresponding real valued RC approaches with low data dimension. For instance, when the data dimension is 48, the recognition rates of QLRC, QCRC, QSRC and QCROC are 80.77%, 78.62%, 77.00%, 85.62% while those of LRC, CRC, SRC and CROC are 72.62%, 68.00%, 68.08%, 78.85%, respectively. The improvements in terms of recognition rates exceeds 5%. This is attributed to fact that when data dimension is low, the color information of images has a more prominent contribution to the recognition.

3) EXPERIMENT 3–EFFECT OF TRAINING SET SIZE

In this experiment, we vary the size of the training set from the 30% to 70% of images per subject and test each competing algorithm on the remaining test images. The data dimension is fixed as 80. Table 4 displays the recognition results of various

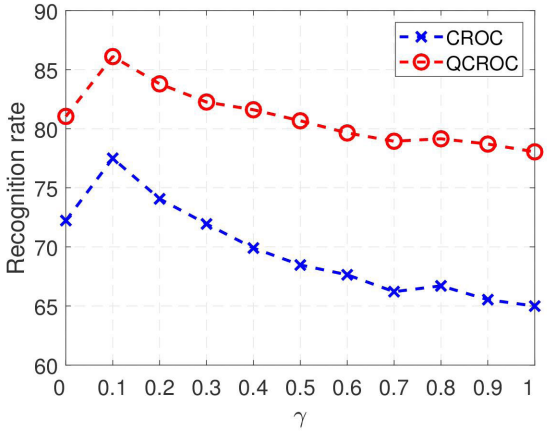


FIGURE 4. Average recognition rates of CROC and QCROC as a function of the parameter γ over 10 random runs on the AR dataset.

RC approaches with using different training set sizes over 10 random runs.

- The performance of NN and QNN are inferior to that of other RC methods in terms of recognition rates in the experiments. Because they fail to take into consideration of the distribution and subspace structure of data while other RC methods do.
- QCROC improves other competing RC algorithms with notable performance gains in most cases at no extra computational cost. This owes to the fact that QCROC can fully harness the color information of color images and subspace structural information of data.

4) EXPERIMENT 4–COMPARISON WITH MCF and 2d-MCF

In this experiment, we compare QCROC with the MCF (Multi-Colour Fusion) and the 2D-MCF methods [2], [32] for color face recognition. We follow the experimental settings of the 2D-MCF method [2], [32]. Specifically, we choose 10 images per subject of the first 50 subjects from the AR database for evaluation. The 10 images include the 2 neutral images, 6 images with expression variations, two images occluded by sunglasses and scarf, respectively. We randomly select n_c images among the 10 images per subject for training and the rest are used for testing. To get reliable results, we repeat each method with 10 random runs and report the average recognition rates in Table 7. Note that QCROC achieves comparable results with other methods. It is worth pointing out that 2D-MCF uses 27 color components from 9 different colour spaces while QCROC only uses the RGB color space.

TABLE 4. Recognition rates (%) using varying percent p_c(%) of training samples per class on the AR dataset over 10 random runs. Best results are marked bold and the second best results are marked with underline.

$p_c(\%)$	Methods	NN	QNN	LRC	QLRC	CRC	QCRC	SRC	QSRC	CROC	QCROC
30	Mean	30.44	32.69	50.98	57.43	74.19	80.07	62.28	71.99	70.45	**83.88**
	Min	28.58	30.84	48.63	55.89	71.79	78.95	60.42	69.89	68.74	**82.53**
	Max	31.53	33.58	55.26	59.47	75.95	81.89	63.68	73.26	72.74	**86.32**
40	Mean	37.02	39.76	66.53	73.78	80.40	85.80	70.95	80.89	81.55	**88.65**
	Min	35.75	38.00	63.56	70.00	77.94	84.50	68.81	79.50	78.94	**86.94**
	Max	38.00	42.25	70.31	77.94	82.00	87.69	72.31	82.13	83.69	**90.31**
50	Mean	42.39	45.61	77.94	83.98	84.34	89.80	77.10	86.48	88.14	**91.17**
	Min	40.92	43.69	76.31	82.23	82.85	87.77	75.77	84.69	85.92	**90.00**
	Max	43.38	47.15	80.38	86.23	86.31	90.69	79.23	88.23	89.31	**92.62**
60	Mean	45.34	49.00	82.31	88.25	86.26	91.45	80.62	88.55	89.75	**92.34**
	Min	43.91	46.09	80.18	86.82	84.36	89.45	78.82	87.09	88.09	**91.00**
	Max	46.27	51.09	85.00	89.09	87.82	93.64	81.55	90.00	90.91	**93.36**
70	Mean	49.61	53.46	87.36	92.89	88.31	92.68	84.39	91.70	92.36	**93.59**
	Min	47.38	51.63	84.38	90.88	85.63	91.38	82.38	90.63	91.25	**92.13**
	Max	52.25	57.13	89.13	93.88	89.88	93.63	86.00	93.13	93.00	**94.75**

TABLE 5. Recognition rates (%) using varying feature dimension d on the Caltech Faces dataset. Best results are marked bold and the second best results are marked with underline.

Data dimension m	NN	QNN	LRC	QLRC	CRC	QCRC	SRC	QSRC	CROC	QCROC
48	67.82	69.83	72.10	74.83	88.56	91.64	87.39	91.35	89.77	**93.51**
80	69.17	72.39	77.13	77.90	89.89	92.79	91.21	92.53	91.90	**93.88**
192	74.63	75.89	79.08	81.29	92.44	94.60	92.79	93.59	92.64	**94.89**
300	75.32	77.24	80.46	82.76	92.79	94.83	94.63	93.48	93.30	**95.06**

FIGURE 5. Sample images from the Caltech Faces dataset.

C. EXPERIMENTS ON THE CALTECH FACES DATABASE

The Caltech Faces dataset [33] contains 450 color facial images of 27 subjects with varying illumination, expression, and background. Fig. 5 shows several sample images from this dataset. Before performing face recognition, we first adopt the Viola-Jones face detector [34] to detect and extract the face regions from the original images.

1) EXPERIMENT 1–EFFECT OF THE PARAMETER γ

In the first experiment, we evaluate the effect of the parameter γ to the performance of QCROC in comparison with CROC. To this end, we $n_c = 3$ images per subject uniformly at random for training and use the rest ones for testing and evaluation. Fig. 6 shows the average recognition rates of CROC and QCROC as a function of γ.

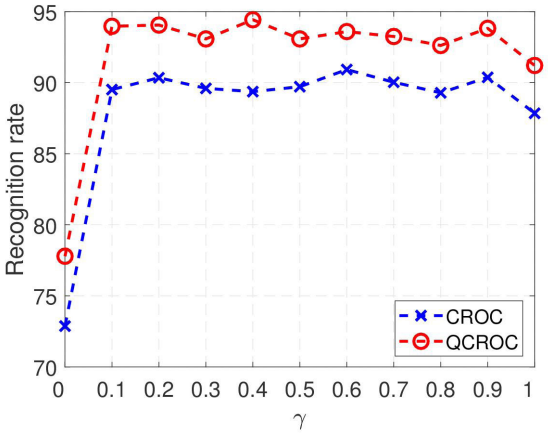


FIGURE 6. Average recognition rates of CROC and QCROC as a function of the parameter γ over 10 random runs on the Caltech Faces dataset.

2) EXPERIMENT 2–EFFECT OF FEATURE DIMENSION

Table 5 displays the recognition rates of the competing classifiers with varying data dimension. The results suggest that QCROC works well on this dataset and achieves decent recognition rates even when the data dimension is very low. This is because the color information can be well harnessed by QCROC to compensate the insufficiency of data dimension.

3) EXPERIMENT 3–EFFECT OF TRAINING SET SIZE

In this experiment, we vary the number of randomly selected training samples per subject from 3 to 7 and run each classifier on the remaining samples with data dimension as 48. Table 6 shows the recognition results of various RC

TABLE 6. Recognition rates (%) using varying number n_c of training samples per class on the Caltech Faces dataset over 10 random runs. Best results are marked bold and the second best results are marked with underline.

n_c	Methods	NN	QNN	LRC	QLRC	CRC	QCRC	SRC	QSRC	CROC	QCROC
3	Mean	67.30	70.03	73.02	77.10	87.99	92.01	87.67	92.79	90.26	**92.96**
	Min	61.78	66.09	68.97	74.43	84.77	89.08	81.61	**89.66**	85.63	86.21
	Max	72.41	73.56	78.45	82.47	91.09	94.54	91.67	**96.55**	93.10	**96.55**
4	Mean	72.37	74.19	81.58	83.77	90.03	95.41	90.88	**95.56**	93.16	**95.56**
	Min	69.60	69.60	76.60	78.12	86.02	93.31	86.32	93.01	89.36	**94.22**
	Max	75.99	78.72	86.02	86.63	92.10	96.96	94.53	**98.18**	94.83	97.57
5	Mean	75.45	77.03	83.00	87.90	91.74	**97.00**	92.35	96.87	95.42	96.58
	Min	71.94	73.55	79.68	82.58	90.32	**94.84**	90.65	94.52	93.55	93.87
	Max	77.42	80.32	86.77	91.94	93.55	**98.71**	94.19	99.35	98.06	98.06
6	Mean	77.56	79.69	87.11	91.44	93.75	97.73	94.50	97.66	95.67	**98.04**
	Min	74.23	77.66	85.22	87.63	91.07	95.53	93.47	**96.56**	93.13	95.53
	Max	80.76	81.79	88.66	94.16	95.88	**99.66**	96.56	98.28	97.59	98.97
7	Mean	79.15	81.62	90.07	92.98	95.44	97.94	94.82	98.71	96.51	**99.01**
	Min	73.16	77.94	86.03	91.18	94.12	96.69	92.65	97.79	94.85	**98.16**
	Max	81.99	86.03	93.38	94.85	97.79	98.90	97.43	99.26	99.26	**100.00**

TABLE 7. Average recognition rates (%) over 10 random runs of different methods with varying number n_c of gallery images on the 50 subjects of the AR database. Best results are marked bold and the second best results are marked with underline.

n_c	2	3	4	5	6	7
MCF	80.92	86.37	86.60	88.36	89.45	90.23
2D-MCF	**88.25**	**91.80**	91.53	**93.84**	95.30	94.33
CROC	73.68	83.03	87.63	90.60	90.35	91.67
QCROC	76.50	86.14	**92.00**	93.24	**95.70**	**96.47**

TABLE 8. Recognition rates (%) of QCROC using different color spaces with varying dimension on the AR database. Best results are marked bold and the second best results are marked with underline.

Dimension m	RGB	YIQ	YCbCr
48	**85.62**	67.30	73.07
80	**91.38**	83.23	86.38
192	**97.31**	93.38	96.46
300	**97.54**	95.38	97.15

algorithms over 10 random runs. The results are consistent with those on the SCface and AR datasets, which further verify the efficacy of QCROC for color face recognition.

D. EXPERIMENTS WITH DIFFERENT COLOR SPACES

In this subsection, we investigate the effect of the color spaces to the performance of the proposed QCROC method for color face recognition. Table 8 reports the recognition rates of QCROC using different color spaces with varying dimension on the AR database. It can be seen from the results that the RGB space yields the best results in terms of recognition accuracy in most cases. The reason may be attributed to the fact that the intensity values in different channels of the RGB space are balanced and close while those in the YIQ and YCbCr space are unbalanced. For instance, we experimentally observe that the intensity values in the first channel of YIQ are generally much larger than those in the other two channels. This may degrade the performance of QCROC in extracting the correlation among different channels.

V. CONCLUSION AND DISCUSSION

We have proposed a quaternion linear regression based classifier referred to as QLRC for face recognition using color facial images. QLRC allows us to directly recognize the color facial images in a holistic manner instead of processing each color channel independently. We have further improved QLRC by integrating it and the quaternion collaborative representation into a unified framework and developed the QCROC method. The experiments have demonstrated that QCROC provides significant improvement over other competing representation based classifiers for face recognition.

The merit of QCROC is that it can take full advantage of multiple color channels of color images for face recognition by coding these channels in a holistic manner with quaternions. This gives rise to the substantial improvement of QCROC over other competing algorithms in the paper. Meanwhile, the demerit of QCROC is that it generally requires more computation resources than the corresponding real valued based methods. This is expected because quaternion based methods compute the structural correlation between multiple color channels while previous real valued based methods overlook such correlation. An interesting topic for future work is to devise more efficient algorithms for QCROC.

APPENDIX

The appendix is to present the proof of Theorem 1.

Proof: We prove the result by showing that the objective functions of Eq. (11) and Eq. (14) have the same value for every quaternion variable $\dot{\mathbf{c}} \in \mathbb{H}^{n_k}$.

For a quaternion matrix $\dot{\mathbf{M}} = \mathbf{M}_0 + \mathbf{M}_1\mathrm{i} + \mathbf{M}_2\mathrm{j} + \mathbf{M}_3\mathrm{k} \in \mathbb{H}^{m\times n}$ and a quaternion vector $\dot{\mathbf{v}} = \mathbf{v}_0 + \mathbf{v}_1\mathrm{i} + \mathbf{v}_2\mathrm{j} + \mathbf{v}_3\mathrm{k} \in \mathbb{H}^n$,

we have

$$\begin{aligned}\dot{\mathbf{M}}\dot{\mathbf{v}} &= (\mathbf{M}_0 + \mathbf{M}_1\mathrm{i} + \mathbf{M}_2\mathrm{j} + \mathbf{M}_3\mathrm{k})\,(\mathbf{v}_0 + \mathbf{v}_1\mathrm{i} + \mathbf{v}_2\mathrm{j} + \mathbf{v}_3\mathrm{k}) \\ &= (\mathbf{M}_0\mathbf{v}_0 - \mathbf{M}_1\mathbf{v}_1 - \mathbf{M}_2\mathbf{v}_2 - \mathbf{M}_3\mathbf{v}_3) \\ &\quad + (\mathbf{M}_1\mathbf{v}_0 + \mathbf{M}_0\mathbf{v}_1 - \mathbf{M}_3\mathbf{v}_2 + \mathbf{M}_2\mathbf{v}_3)\mathrm{i} \\ &\quad + (\mathbf{M}_2\mathbf{v}_0 + \mathbf{M}_3\mathbf{v}_1 + \mathbf{M}_0\mathbf{v}_2 - \mathbf{M}_1\mathbf{v}_3)\mathrm{j} \\ &\quad + (\mathbf{M}_3\mathbf{v}_0 - \mathbf{M}_2\mathbf{v}_1 + \mathbf{M}_1\mathbf{v}_2 + \mathbf{M}_0\mathbf{v}_3)\mathrm{k}. \qquad (17)\end{aligned}$$

Recall the definitions of the two quaternion operator $\mathcal{P}$ and $\mathcal{Q}$.

$$\mathcal{P}\left(\dot{\mathbf{M}}\right) := \begin{bmatrix} \mathbf{M}_0 & -\mathbf{M}_1 & -\mathbf{M}_2 & -\mathbf{M}_3 \\ \mathbf{M}_1 & \mathbf{M}_0 & -\mathbf{M}_3 & \mathbf{M}_2 \\ \mathbf{M}_2 & \mathbf{M}_3 & \mathbf{M}_0 & -\mathbf{M}_1 \\ \mathbf{M}_3 & -\mathbf{M}_2 & \mathbf{M}_1 & \mathbf{M}_0 \end{bmatrix} \in \mathbb{R}^{4m\times 4n}. \qquad (18)$$

$$\mathcal{Q}\left(\dot{\mathbf{v}}\right) := \nu\left[\mathbf{v}_0^T, \mathbf{v}_1^T, \mathbf{v}_2^T, \mathbf{v}_3^T\right]^T \in \mathbb{R}^{4n}. \qquad (19)$$

Combining Eqs. (17-19), we have

$$\mathcal{Q}\left(\dot{\mathbf{M}}\dot{\mathbf{v}}\right) = \mathcal{P}\left(\dot{\mathbf{M}}\right)\mathcal{Q}(\dot{\mathbf{v}}). \qquad (20)$$

It follows that

$$\begin{aligned}\mathcal{Q}(\dot{\mathbf{y}} - \dot{\mathbf{X}}_k\dot{\mathbf{c}}) &= \mathcal{Q}(\dot{\mathbf{y}}) - \mathcal{Q}(\dot{\mathbf{X}}_k\dot{\mathbf{c}}) \qquad (21)\\ &= \mathcal{Q}(\dot{\mathbf{y}}) - \mathcal{P}(\dot{\mathbf{X}}_k)\mathcal{Q}(\dot{\mathbf{c}}), \qquad (22)\end{aligned}$$

where the first equality is from the linearity of $\mathcal{Q}$ and the second one results from Eq. (20). On the other hand, according to the definition of the operator $\mathcal{Q}$, there holds $\|\dot{\mathbf{v}}\|_2^2 = \|Q(\dot{\mathbf{v}})\|_2^2$, $\forall\dot{\mathbf{v}} \in \mathbb{H}^n$. Then we have

$$\begin{aligned}\left\|\dot{\mathbf{y}} - \dot{\mathbf{X}}_k\dot{\mathbf{c}}\right\|_2^2 &= \left\|\mathcal{Q}(\dot{\mathbf{y}} - \dot{\mathbf{X}}_k\dot{\mathbf{c}})\right\|_2^2 \qquad (23)\\ &= \left\|\mathcal{Q}(\dot{\mathbf{y}}) - \mathcal{P}(\dot{\mathbf{X}}_k)\mathcal{Q}(\dot{\mathbf{c}})\right\|_2^2. \qquad (24)\end{aligned}$$

This proves the equivalence of Eq. (11) and Eq. (14) and completes the proof. ■

ACKNOWLEDGMENT

The authors would like to thank the editor, the associate editor and anonymous reviewers for their constructive comments and suggestions that have greatly improved this paper.

REFERENCES

[1] Z. Liu, Y. Qiu, Y. Peng, J. Pu, and X. Zhang, "Quaternion based maximum margin criterion method for color face recognition," *Neural Process. Lett.*, vol. 45, no. 3, pp. 913–923, Jun. 2017.

[2] M. Alrjebi, N. Pathirage, W. Liu, and L. Li, "Face recognition against occlusions via colour fusion using 2D-MCF model and SRC," *Pattern Recognit. Lett.*, vol. 95, pp. 14–21, Aug. 2017.

[3] J. Wright, A. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 2, pp. 210–227, Feb. 2009.

[4] Y. Wang, Y. Y. Tang, L. Li, and H. Chen, "Modal regression based atomic representation for robust face recognition and reconstruction," *IEEE Trans. Cybern.*, to be published.

[5] R. Lan, Y. Zhou, Z. Liu, and X. Luo, "Prior knowledge-based probabilistic collaborative representation for visual recognition," *IEEE Trans. Cybern.*, to be published.

[6] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2106–2112, Nov. 2010.

[7] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 471–478.

[8] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 815–823.

[9] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2892–2900.

[10] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2015, vol. 1, no. 3, p. 6.

[11] Y. Gao, J. Ma, and A. L. Yuille, "Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2545–2560, May 2017.

[12] 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.

[13] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 3, pp. 218–233, Feb. 2003.

[14] S. Shekhar, V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Joint sparse representation for robust multimodal biometrics recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 113–126, Jan. 2014.

[15] Y. Chi and F. Porikli, "Classification and boosting with multiple collaborative representations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1519–1531, Aug. 2014.

[16] S. Cai, L. Zhang, W. Zuo, and X. Feng, "A probabilistic collaborative representation based approach for pattern classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2950–2959.

[17] H. Chen and Y. Wang, "Kernel-based sparse regression with the correntropy-induced loss," *Appl. Comput. Harmon. Anal.*, vol. 44, no. 1, pp. 144–164, 2018.

[18] Y. Xu, L. Yu, H. Xu, H. Zhang, and T. Nguyen, "Vector sparse representation of color image using quaternion matrix analysis," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1315–1329, Apr. 2015.

[19] L. Liu, S. Li, and C. L. P. Chen, "Quaternion locality-constrained coding for color face hallucination," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1474–1485, May 2017.

[20] L. Yu, Y. Xu, H. Xu, and H. Zhang, "Quaternion-based sparse representation of color image," in *Proc. Int. Conf. Multimedia Expo*, Jul. 2013, pp. 1–7.

[21] X. Zhu, Y. Xu, H. Xu, and C. Chen, "Quaternion convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 631–647.

[22] Q. Yin, J. Wang, X. Luo, J. Zhai, S. K. Jha, and Y.-Q. Shi, "Quaternion convolutional neural network for color image classification and forensics," *IEEE Access*, vol. 7, pp. 20293–20301, 2019.

[23] R. Lan, H. Lu, Y. Zhou, Z. Liu, and X. Luo, "An LBP encoding scheme jointly using quaternionic representation and angular information," *Neural Comput. Appl.*, to be published.

[24] A. Kolaman and O. Pecht, "Quaternion structural similarity: A new quality index for color images," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1526–1536, Apr. 2012.

[25] F. Shang and A. Hirose, "Quaternion neural-network-based PolSAR land classification in Poincare-sphere-parameter space," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5693–5703, Sep. 2014.

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

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

[28] Q. Barthélemy, A. Larue, and J. I. Mars, "Color sparse representations for image processing: Review, models, and prospects," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3978–3989, Nov. 2015.

[29] M. Grgic, K. Delac, and S. Grgic, "SCface—Surveillance cameras face database," *Multimedia Tools Appl.*, vol. 51, no. 3, pp. 863–879, Feb. 2011.

[30] A. Martinez and R. Benavente, "The AR face database," Comput. Vis. Center, Ohio State Univ., Columbus, OH, USA, Tech. Rep. 24, Jun. 1998.

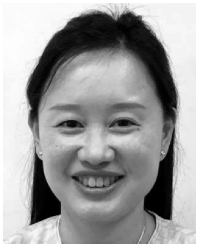
[31] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.

[32] M. Alrjebi, W. Liu, and L. Li, "Two directional multiple colour fusion for face recognition," in *Proc. Int. Conf. Comput., Commun., Syst.*, Nov. 2015, pp. 171–177.

[33] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2003, pp. 264–271.

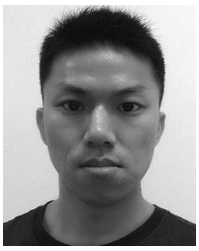
[34] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

CUIMING ZOU received the B.Sc. and M.Sc. degrees from Hubei University, Wuhan, China, and the Ph.D. degree from the University of Macau, Macau, China. She is currently with the School of Information Science and Engineering, Chengdu University, Chengdu, China. Her current research interests include pattern recognition and computer vision.



KIT IAN KOU received the M.Sc. and Ph.D. degrees in mathematics from the University of Macau, Macao, China. She is a Life Member of the Clare Hall, University of Cambridge, U.K. She is currently an Associated Professor with the Department of Mathematics, Faculty of Science and Technology, University of Macau. Her current research interests include pattern recognition, Quaternion analysis in image processing, Fourier analysis, and sampling theory. She received the Third Prize in the Natural Science Award of 2018 Macao Science and Technology Awards.

LI DONG received the B.Eng. degree from Chongqing University, in 2012, and the M.S. and Ph.D. degrees from the University of Macau, in 2014 and 2018, respectively. He is currently an Assistant Professor with the Department of Computer Science, Faculty of Electrical Engineering and Computer Science, Ningbo University. His research interests include statistical image modeling and processing, multimedia security and forensic, and machine learning.



XIANWEI ZHENG received the B.Sc. degree in mathematics from Hanshan Normal University, Chaozhou, China, in 2009, the M.Sc. degree in applied mathematics from Shantou University, Shantou, China, in 2012, the Ph.D. degree in mathematics from Shantou University, and the Ph.D. degree in computer science from the University of Macau, Macau, China, in 2018. In 2018, he joined the School of Mathematics and Big Data, Foshan University, Foshan, China, as a Young Research Fellow. His current research interests include graph signal processing, image processing, machine learning, wavelet analysis, and frame theory.

YUAN YAN TANG (F'04–LF'16) received the B.Sc. degree in electrical and computer engineering from Chongqing University, Chongqing, China, the M.Eng. degree in electrical engineering from the Beijing Institute of Posts and Telecommunications, Beijing, China, and the Ph.D. degree in computer science from Concordia University, Montreal, QC, Canada. He is currently a Chair Professor of the Faculty of Science and Technology, University of Macau, Macau, China, and a Professor/Adjunct Professor/Honorary Professor with several institutes, including Chongqing University, Concordia University, and Hong Kong Baptist University, Hong Kong. He has published over 400 academic articles and has authored/coauthored over 25 monographs/books/bookchapters. His current research interests include wavelets, pattern recognition, and image processing.

Dr. Tang is a Fellow of the IAPR. He has served as a General Chair, a Program Chair, and a Committee Member for several international conferences. He is the Founder and an Editor-in-Chief of *International Journal on Wavelets, Multiresolution, and Information Processing*, and an Associate Editor of several international journals. He is the Founder and the Chair of the Pattern Recognition Committee in the IEEE International Conference on Systems, Man, and Cybernetics. He is the Founder and a General Chair of the series International Conferences on Wavelets Analysis and Pattern Recognition. He is the Founder and the Chair of the Macau Branch of International Associate of Pattern Recognition (IAPR).