Image Classification Based on Sparse Representation in the Quaternion Wavelet Domain

LONG H. NGO1, NIKOLAY M. SIRAKOV2, MARIE LUONG1, EMMANUEL VIENNET1, AND THUONG LE-TIEN3, (Member, IEEE)

1Laboratoire de Traitement et de Transport de l’Information (L2TI), UR 3043, Université Sorbonne Paris Nord, 93430 Villetaneuse, France
2Department of Mathematics, Texas A&M University-Commerce, Commerce, TX 75428, USA
3Department of Telecommunications, University of Technology, National University of Ho Chi Minh City, Ho Chi Minh City 700000, Vietnam

Corresponding author: Long H. Ngo (longngo155@gmail.com)

ABSTRACT In this study, we propose a novel sparse representation learning method in the Quaternion Wavelet (QW) domain for multi-class image classification. The proposed method takes advantages from: i) the QW decomposition, which promotes sparsity and provides structural information about the image data while allowing approximate shift-invariance, to extract meaningful features from low-frequency QW sub-bands, ii) the dimensionality reduction method using Principal Component Analysis (PCA) for reducing the complexity of the problem, and iii) the sparse representation of the generated QW features to efficiently learn and capture the meaningful and compact information of this data. After the QW decomposition, the features extracted from low-frequency image sub-bands information are projected, by the PCA, into a new feature space with lower dimensionality. The features extracted from the training samples are used to construct a dictionary, while the features of the test samples are sparsely coded for the classification step. The sparse coding problem is formulated in a QW Least Absolute Shrinkage and Selection Operator (QWLasso) model applying quaternion $l_1$ minimization. A novel Quaternion Fast Iterative Shrinkage-Thresholding Algorithm (QFISTA) is developed to solve the QWLasso model. The experiments conducted on various public image datasets validated that the proposed method possesses higher accuracy, sparsity, and robustness in comparison with several contemporary methods in the field including Neural Networks.

INDEX TERMS Face/object classification, quaternion algebra, quaternion wavelet, sparse representation, wavelet.

I. INTRODUCTION

Image classification is an important problem in computer vision and finds applications covering multiple domains such as healthcare, security, and manufacturing [1]–[4]. Image classification is a challenging task due to the complex nature of the images with specific properties and intricate structures as well as possible noises, varying illuminations, occlusion, outliers and complex backgrounds.

Recent years have witnessed a high attention to deep neural networks (NNs) for learning feature representation of data while benefiting from the huge amounts of available data as well as the great power of computers [5], [6]. However, these networks suffer from many drawbacks such as their data-hungriness [5], [6]. Mainly, NNs are capable of exhibiting their potentials when they are trained with very large datasets which are not always available.

Among successful learning methods developed in the literature, Sparse Representation (SR) has shown to be one of the most efficient approaches producing compact as well as simple representation of the signal through a small number of meaningful features [7]. SR has achieved state-of-the-art performance in signal and image processing [8], particularly in image classification [9]–[12]. Also, it was determined that SR is the mechanism in the primary visual cortex to achieve concise description of images in terms of features and considered as a main principle to efficiently represent complex data [13], [14]. Furthermore, SR is used to enforce the learning process [15] and to solve optimization problems with non-convex objective function [16]. Hence, SR-based
classification (SRC) [9]–[12] methods have proven to be efficient and robust to noise, occlusion, and corruption.

Although SRC methods [10]–[12] have achieved promising performances, efforts have been made to improve their accuracy and robustness especially for large scale systems. One approach is to combine neural network (NN) with SR to take advantages from their strengths to achieve the goal [17], [18]. Using unsupervised SR coding inside of supervised learning method improves the performance of deep NNs [19]. However, due to the limitation of NNs, it is necessary to develop an efficient approach for learning the optimal representations of data and providing high classification performance when large training datasets are not available. Another approach is the continuation of advances in SRC. Usually, the aforementioned SRC method is applied to the spatial domain. A recent approach was proposed by Ngo et al. [20] and performed in the sparsity-promoting wavelet domain to enhance the sparsity level of the codes and their discrimination ability. Despite its successful applications to classification [21], the discrete wavelet transform (DWT) is restricted by its lack of shift-invariance [22].

To overcome this drawback, we developed a novel SRC method using the quaternion wavelet transform (QWT) [22], [23], which is based on the 2D analytic signal, 2D Hilbert Transform (HT), and quaternion algebra $\mathbb{H}$ [24]. The QWT is proven to be shift-invariant and conveys richer geometric information than DWT [22], [23]. Moreover, it can be computed using a dual-tree filter bank with linear computational complexity [22], [23]. Based on its properties, the QWT has been successfully applied to a number of research fields such as disparity estimation [23], image denoising [25], face recognition [26] and texture classification [27].

To further enhance the classification performance, we design a new SRC approach by exploiting sparsity coding in the Quaternion Wavelet (QW) domain to take advantage of the useful properties of the QWT and the SR. To the best of our knowledge, there is no other SRC approach implemented in the QW domain. In [27], QWT is applied for image texture classification. The same transform is also implemented to face recognition in [26]. In [28], [29], we have identified two main SR-based methods but they are performed in the quaternion space, where the three channels of color images are modelled as a quaternion. In [28], Xu et al. proposed a Quaternion Sparse Representation (QSR) model, with $l_0$ minimization, for color image restoration and developed a Quaternion Orthogonal Matching Pursuit (QOMP) algorithm to determine the sparse coefficients. Further, Zou et al. [29] proposed an $l_1$-norm QSR model, resulting in the quaternion Lasso (QLasso) model, which computes the sparse vector using the Alternating Direction Method of Multipliers (ADMM) algorithm [30].

Unlike the above methods, the newly proposed approach constructs the dictionary in the 4D space of the algebra of quaternions (AQ), using features described by QW coefficients in four low-frequency (LL) wavelet sub-bands. In fact, the QWT decomposition of an input image yields one low-frequency QW sub-band and three high-frequency QW sub-bands, where each quaternion sub-band is defined by four wavelet coefficients sub-bands. Here, we only need features described by QW coefficients in LL sub-bands as they constitute the main component of the image. The proposed method benefits from the SR of the QW coefficients features to learn and capture the meaningful information of the visual data. Specifically, we formulate the problem of finding the SR in the QW domain, by a novel QW Least Absolute Shrinkage and Selection Operator (QWLasso) model with quaternion $l_1$ minimization. To solve this QWLasso, we develop the novel Quaternion Fast Iterative Shrinkage-Thresholding Algorithm (QFISTA), which is based on the real-valued FISTA method [31]. The QFISTA maps the quaternion dictionary to a 4D real-valued matrix, composed by specific low-frequency wavelet sub-band coefficients. In addition, we develop an upper bound for the QWLasso model and use it as an approximation to design the iterative scheme applied to find the sparse coefficients of the QW features. The higher separability of the quaternion vector gives an advantage over the minimization model in the field of real numbers. The advantage of classifying in the 4D space of quaternions is that formulating the minimization model in quaternions provides additional information coming from the direction of the 4D vector composed by the quaternion components. Hence, the proposed method improves not only the sparsity level of the sparse codes for classification, but also the robustness, resulting in an increase of the classification accuracy.

In our new method, training and test samples are transformed into the QW domain and principal component analysis (PCA) is used to reduce the dimension of the generated features and computational cost. The proposed method is performed in the QW domain in two steps. In the first one, the dictionary is constructed from the QW coefficients in the low frequency wavelet sub-bands of the training samples. In the second step, the features extracted from the test samples are sparsely coded over the dictionary. Then, the class of the test samples is identified using a reconstruction residuals task.

The main contributions of this study could be summarized:

i) We develop a novel SRC method in the 4D AQ over the wavelet domain;

ii) In the QW domain, we build a dictionary of quaternions constructed by low-frequency wavelet sub-bands;

iii) We define the new QWLasso model to find the sparse codes in the AQ;

iv) We develop the advanced QFISTA algorithm to solve the novel QWLasso and calculate the quaternion sparse coefficients.

The experimental results validate the main advantage of the proposed method, which outperforms, in terms of classification accuracy, the state-of-the-art methods including NNs (Tables 2 and 3 with 99.6% accuracy reported). The proposed method is referred to as SRCQW, which stands for SRC in the QW domain.

The rest of the paper is organized as follows: Section II reviews related works; Section III presents basic quaternion...
concepts; Two subsection of section IV describe: A. 2D QWT as four 1D functions; B. The implementation of the QWT; Sections V develops the novel SRCQW method; Section VI presents its computational complexity; Section VII validates the SRCQW on commonly used datasets and compares the obtained results with several contemporary methods; Section VIII presents the conclusion and discussions on the contributions, advantages, and bottlenecks.

II. RELATED WORKS
In the conventional SRC method [9], a test sample is a linear combination of a few basis vectors taken from an overcomplete dictionary whose atoms are the training samples. The sparse coefficients of the linear combination model are first computed via a sparsity-constrained optimisation problem over the dictionary. Then, the reconstruction residual of each class is calculated and the test sample is assigned to the class with the minimum residual. However, the weakness of the method relates to the case of complex and large datasets whose dictionaries are formed by all training samples of each class. To overcome this drawback, a number of methods have been developed to learn a compact dictionary [10]–[12], [32]. One approach is to learn a discriminative dictionary of small size from a selective dataset instead of the entire dataset for example, the Discriminative K-SVD (D-KSVD) [10] and the Label Consistent K-SVD (LC-KSVD) [11], which are both based on the K-SVD model [33], defined the objective function with the help of classification and reconstruction errors terms [10] or a new constraint called discriminative sparse code error to impose discriminability on the sparse codes [11]. However, this approach is non-convex and a global optimal solution cannot be guaranteed. Another approach learns class-specific dictionaries [12], [32], [34]. In [12], the Fisher Discrimination Dictionary Learning (FDDL) method utilizes a Fisher discrimination criterion to learn a structured dictionary whose sub-dictionaries have specific class labels. Based on the discriminative FDDL method, Vu et al. [32] developed the low-rank shared dictionary learning (LRSDL) method, characterized by a shared dictionary and class-specific dictionaries. For these methods, learning a dictionary instead of using all the training samples as a dictionary, provides a better classification compared to the classical SRC method. However, these methods are not resistant to training samples degradation by high-amplitude noise or occlusions.

A recent approach, namely Sparse Representation Wavelet based Classification (SRWC) [20] was developed in the wavelet domain, taking advantage of the wavelet-promoted sparsity and leading to better discrimination. To improve the image classification accuracy, the SRWC exploits the sparse representation of the features described by the complementary information from the low-frequency sub-band of the wavelet coefficients. As a consequence, the SRWC method exhibited higher accuracy compared to the contemporary SR-based methods in [9], [11], [12]. In [20], an overcomplete dictionary is constructed by mapping the training samples into the wavelet domain. To determine the membership of the query test sample, its sparse vector is first computed through an $l_1$-norm minimization problem. Then it is utilized for the recognition of its label via a reconstruction residual task.

Inspired by the work in [20] and taking advantage of the SR of sparsity-promoting wavelets in the AQ, the present study develops a novel SRC method in the QW domain.

III. BASIC CONCEPTS OF THE ALGEBRA OF QUATERNIONS
The quaternions, which are proposed by W. R. Hamilton in 1843 [35], are numbers having one real and three imaginary parts. The AQ, denoted by $\mathbb{H}$, could be regarded as a 4D Clifford algebra $Cl_{0,2}$ [36]. In this paper, we denote scalar variables, vectors and matrices by lowercase letters (e.g., $a \in \mathbb{R}$), bold types (e.g., $\mathbf{a} \in \mathbb{R}^M$) and bold capital letters (e.g., $\mathbf{A} \in \mathbb{R}^{M \times N}$), respectively. In the AQ, a dot (above the variable) denotes quaternion variable, e.g., $\hat{a} \in \mathbb{H}$. Accordingly, vectors and matrices with quaternion entries are indicated as $\hat{\mathbf{a}} \in \mathbb{H}^M$ and $\hat{\mathbf{A}} \in \mathbb{H}^{M \times N}$, respectively.

$\mathbb{H} = \{ \hat{a} = q_0 + iq_1 + jq_2 + kq_3 : q_0, q_1, q_2, q_3 \in \mathbb{R} \}$ is the set of quaternions composing the AQ [36]. The three imaginary numbers satisfy the following properties:

\[
i^2 = j^2 = k^2 = ijk = -1,
ij = -ji = k, \quad ik = -ki = -j, \quad jk = -kj = i,
\]

and its norm is: $\|\hat{a}\| = \sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}$.

IV. QUATERNION WAVELET TRANSFORM
The quaternion wavelets considered in this study are based on Bulow quaternion analytic signal [24] and a dual-tree QWT introduced in [22], [23].

A. QUATERNION ANALYTIC SIGNAL AND QW
The quaternion analytic signal associated with a real 2D signal $s(x,y)$ is defined by its partial $(s_{\mathcal{H}_x}, s_{\mathcal{H}_y})$ and total $(s_{\mathcal{H}_0})$ HTs as follows [24]:

\[
s_{\mathcal{H}_i}(x, y) = s(x, y) + is_{\mathcal{H}_x}(x, y) + js_{\mathcal{H}_y}(x, y) + ks_{\mathcal{H}_0}(x, y),
\]

where $s_{\mathcal{H}_0} = s(x, y) \ast \frac{\delta(x)}{\pi x}, s_{\mathcal{H}_x} = s(x, y) \ast \frac{\delta(x)}{\pi y}, s_{\mathcal{H}_y} = s(x, y) \ast \frac{1}{\pi x y}$. The symbol $\ast$ denotes the 2D convolution operation, while $\delta(x)$ and $\delta(y)$ are impulse functions along y-axis and x-axis, respectively.

In the dual-tree QWT, each quaternion wavelet is composed of four quadrature components (a real wavelet and its 2D HTs), which are organised as a quaternion: a real DWT wavelet and three imaginary wavelets obtained by 1D HT along either or both coordinates.

Denote by $\phi(t)$ and $\psi(t)$ the scaling and the wavelet functions of the 1D DWT, respectively. The 2D DWT is computed as the separable tensor products of 1D DWTs over each coordinate: the scaling function $\phi(x) \phi(y)$ and three wavelet functions $\psi(x) \psi(y)$, $\psi(x) \phi(y)$ and $\phi(x) \psi(y)$ oriented...
in the diagonal, vertical and horizontal directions, respectively [22], [23]. Mathematically, the 2D QWT is defined by 1D functions as follows:

\[
\begin{align*}
\psi_q^D &= \psi_g(x) \psi_g(y) + i \psi_f(x) \psi_g(y) + j \psi_g(x) \psi_f(y) + k \psi_f(x) \psi_f(y) \\
\psi_q^V &= \psi_g(x) \phi_g(y) + i \psi_f(x) \phi_g(y) + j \psi_g(x) \phi_f(y) + k \psi_f(x) \phi_f(y) \\
\psi_q^H &= \phi_g(x) \psi_g(y) + i \phi_f(x) \psi_g(y) + j \phi_g(x) \psi_f(y) + k \phi_f(x) \psi_f(y) \\
\phi_q^I &= \phi_g(x) \phi_g(y) + i \phi_f(x) \phi_g(y) + j \phi_g(x) \phi_f(y) + k \phi_f(x) \phi_f(y),
\end{align*}
\]

where $\psi_q^D$, $\psi_q^V$, $\psi_q^H$ are quaternion wavelets oriented in the diagonal, vertical and horizontal directions, respectively. The subscripts {g,f} refer to a real-valued filter and its HT counterpart, respectively. The quaternion scaling function $\phi_q^I$ in (3d) corresponds to the QWT low-frequency coefficients.

\section*{B. QWT IMPLEMENTATION}

The QWT is realized by using the dual-tree algorithm [23]. The decomposition of an image by QWT is performed using 2D DWT to derive a low-frequency (LL) sub-band and three high-frequency (LH, HL, and HH) sub-bands containing image details information in horizontal, vertical, and diagonal directions, respectively [22], [23]. More precisely, the QWT can be implemented using a combination of four filters $G_L$, $G_H$, $F_L$, and $F_H$, where $G_L(n)$ and $G_H(n)$ are respectively low-pass and high-pass wavelet filters, whereas low-pass filter $F_L(n)$ and high-pass filter $F_H(n)$ correspond to the HT of $G_L(n)$ and $G_H(n)$, respectively [22], [23], [25]. Using orthogonal Daubechies wavelets 'db8' (wavelet having 8 vanishing moments) [37], the coefficients of these filters are illustrated in Figure 1. From these filter banks, four different wavelet coefficients are generated in the AQ to obtain the QWT coefficients. The structure of the QWT decomposition is presented in Figure 2, where an image is decomposed into 16 wavelet sub-bands, which construct four quaternion wavelet sub-bands: one low-frequency (LLq) (4a) and three high-frequency (LH_q, HL_q, HH_q) sub-bands (4b-4d). Each QW sub-band is represented by 4 DWT [22] sub-bands.

\[
\begin{align*}
LL_q &= L_q L_q + i L_q L_q + j L_q L_q + k L_q L_q \\
LH_q &= L_q H_q + i L_q H_q + j L_q H_q + k L_q H_q \\
HL_q &= H_q L_q + i H_q L_q + j H_q L_q + k H_q L_q \\
HH_q &= H_q H_q + i H_q H_q + j H_q H_q + k H_q H_q.
\end{align*}
\]

Hence, $LL_q$ represents the low-frequency quaternion sub-bands carrying the QWT low-frequency information. For illustration, the QWT decomposition on Barbara image is depicted in Figure 3. The real part of the chart in Figure 2 is represented by the four real valued wavelet sub-bands in Figure 3a, while the three imaginary parts of Figure 2 are depicted in Figures 3b-3d. It can be noted that the low-frequency sub-band contains significant information of an image. As stated in [21] and validated in [20], only using low-frequency sub-band is sufficient to increase the accuracy of classification in the wavelet domain. We adopt this proposition and implement only the low-frequency sub-band $LL_q$ in the further developments.

\section*{V. THE PROPOSED SRCQW METHOD}

In this section, we introduce a SRC method in the QW domain. First, training and test samples will be mapped into the 4D space, using the QWT, by which the features extracted
from low-frequency sub-bands information are projected into a new feature space using PCA. Then, the dictionary construction (shown in sub-section A) and the classification (described in sub-section B) are performed. The dictionary is constructed by the extracted features of the training samples, while the classification step is fed up with the features of the test samples. Assuming the dictionary is generated, the classification step first estimates the quaternion SR of a query image by solving the novel QWLasso model through applying the novel QFISTA method (quaternion sparse coding stage described in section V-B2); then the class-dependent residual is computed to identify the label of the query image (label assignment stage shown in section V-B3). An overview of the novel SRCQW method is illustrated in Figure 4, while the training procedure is summarized in Algorithm 1 and the main classification procedure is summarized in Algorithm 2.

**A. QUATERNION DICTIONARY CONSTRUCTION**

Consider a classification problem with \( K \) classes. Let \( N_k \) be the number of training samples from the \( k \)-th class for \( 1 \leq k \leq K \). Our aim is to construct a quaternion dictionary matrix defined by quaternion vectors called quaternion atoms. In this dictionary, the entries are quaternions composed of low-frequency sub-bands. The QWT is first applied on every training sample to obtain the low-frequency sub-band (given with Eq.4a) as a descriptor (atom). Next, in the \( k \)-th class, \( N_k \) atoms are arranged as vector columns of quaternions to generate the dictionary for class \( k \):

\[
\hat{D}^k = [\hat{d}^k_1, \hat{d}^k_2, \ldots, \hat{d}^k_{N_k}] \in \mathbb{H}^{M^* \times N_k},
\]

where quaternion atoms \( \hat{d}^k_{n_k} \in \mathbb{H}^{M^*} \), \( n_k = 1, \ldots, N_k \). Then, the PCA [38] is applied on \( \hat{D}^k \) to reduce the dimension of the quaternion atoms by compressing them onto a lower-dimensional feature space with reduced dimension \( M \ll M^* \). Now, we compose the dictionary for all classes by concatenating the training dictionaries \( \hat{D}^k, k = 1, \ldots, K \):

\[
\hat{D} = [\hat{D}^1, \hat{D}^2, \ldots, \hat{D}^K] = [\hat{d}^1_1, \ldots, \hat{d}^1_{N_1}, \hat{d}^2_1, \ldots, \hat{d}^2_{N_2}, \ldots, \hat{d}^K_1, \ldots, \hat{d}^K_{N_K}].
\]

The total number of atoms in the dictionary \( \hat{D} \) is \( N = \sum_{k=1}^{K} N_k \).

Employing Eq. 27, we present \( \hat{D} \) as follows:

\[
\hat{D} = \hat{D}^0 + \hat{D}^1 i + \hat{D}^2 j + \hat{D}^3 k \in \mathbb{H}^{M \times N},
\]

where \( \hat{D}^e = [\hat{D}^{e,1}, \hat{D}^{e,2}, \ldots, \hat{D}^{e,K}] = [\hat{d}^{e,1}_1, \ldots, \hat{d}^{e,1}_{N_1}, \hat{d}^{e,2}_1, \ldots, \hat{d}^{e,2}_{N_2}, \ldots, \hat{d}^{e,K}_1, \ldots, \hat{d}^{e,K}_{N_K}] \in \mathbb{H}^{M \times N}, e = 0, 1, 2, 3 \). Each training dictionary \( \hat{D}^e \) presents information from a single low-frequency sub-band only. For example, \( \hat{D}^0 \) is constructed by the real vectors \( L_f L_q \) from Eq. 4a. It follows from Eq. 25 in the appendix that a quaternion atom can be represented as a quaternion vector of real-value vectors, with \( e = 0, 1, 2, 3 \):

\[
\hat{d}^{e,k}_{n_k} = d^{e,k}_{n_k} + d^{e,k,1}_{n_k} i + d^{e,k,2}_{n_k} j + d^{e,k,3}_{n_k} k, \quad d^{e,k}_{n_k} \in \mathbb{R}^M.
\]

From Eq. 7 and Eq. 4a, it can be seen that \( d^{e,k}_{n_k} \) represents the real part \( L_q \), while the vectors \( d^{e,k,1}_{n_k}, d^{e,k,2}_{n_k}, \) and \( d^{e,k,3}_{n_k} \) represent the three imaginary parts \( i L_q, j L_q, \) and \( k L_q \), respectively. An example image for each vector is presented with every upper left image in Figure 3.

**B. CLASSIFICATION IN THE QW DOMAIN**

Given a test image, the classification problem is to identify its class label. To this end, we first estimate its sparse code in the QW domain by formulating the QWLasso model and proposing the QFISTA method to resolve this problem. Then

---

**Algorithm 1 Quaternion Dictionary Construction**

**Input:** \( N \) training images from the \( K \) classes.

1. Apply QWT on every image to obtain the low-frequency sub-band (Eq.4a) as a descriptor (atom).
2. In the \( k \)-th class, \( N_k \) quaternion atoms are arranged to generate the dictionary for class \( k \):

\[
\hat{D}^k = [\hat{d}^k_1, \hat{d}^k_2, \ldots, \hat{d}^k_{N_k}] \in \mathbb{H}^{M^* \times N_k}.
\]

3. Apply PCA on \( \hat{D}^k \) to reduce its dimension to \( M \ll M^* \).

**Output:** \( \hat{D} = [\hat{D}^1, \hat{D}^2, \ldots, \hat{D}^K] \).
Algorithm 2 SRCQW

**Input:** a test quaternion vector $\hat{y} \in \mathbb{H}^M$, the quaternion dictionary matrix $\hat{D} \in \mathbb{H}^{M \times N}$ for $K$ classes in QW domain and the parameter $\lambda$ from Eq. 9.

1. Normalize the columns of $\hat{D}$ to have unit $l_2$-norm.
2. Compute the quaternion sparse vector via the QWLasso model: $\hat{x} = \arg \min_{x \in \mathbb{H}^N} \left\{ \left\| \hat{y} - \hat{D} \hat{x} \right\|_2^2 + \lambda \| \hat{x} \|_1 \right\}$ (Go to Algorithm 3).
3. Compute the residuals: $r_k(\hat{y}) = \left\| \hat{y} - \hat{D} \hat{x}_k(\hat{x}) \right\|_2^2$ for $k = 1, \ldots, K$.

**Output:** identity $\hat{y} = \arg \min_k r_k(\hat{y})$.

1) SPARSE REPRESENTATION QUATERNION WAVELET MODEL

In [29], the authors extended the SRC and developed the QLasso model to work in the AQ, where every quaternion represents the three color channels, at every image pixel, and preserves the correlation information among the channels. In contrast, the proposed SRCQW method uses four low-frequency sub-bands in the AQ. In [29], to solve the QLasso model and calculate the sparse vector, the authors applied the ADMM approach, while in the present study, we develop the novel optimization algorithm, QFISTA, which we apply to solve the QWlasso model defined with Eq. 9. Another difference between the QLasso [29] and the present QWLasso is that the former is formulated in 3D sub-space of the AQ, while the latter is formulated in the entire 4D space of the AQ. The higher dimension is expected to provide higher accuracy of classification.

Assume we are given an unknown test image $\hat{y}$. By applying QWT we obtain the low-frequency quaternion $LL_{QW}$. After reducing its dimension with the PCA and following Eq. 7, we determine its form as an approximation quaternion $\hat{y} = \hat{y}_0 + \hat{y}_1 i + \hat{y}_2 j + \hat{y}_3 k \in \mathbb{H}^M$, where $\hat{y}_e \in \mathbb{H}^M$, $e = 0, 1, 2, 3$. Accordingly, the SR of $\hat{y}$ is $\hat{x} = \hat{x}_0 + \hat{x}_1 i + \hat{x}_2 j + \hat{x}_3 k \in \mathbb{H}^N$, $\hat{x}_e \in \mathbb{R}^N$. Then, we formulate the SR quaternion wavelet model to estimate vector $\hat{x}$ of the quaternion sparse vector $\hat{x}$:

$$\hat{x} = \arg \min_{x \in \mathbb{H}^N} \| \hat{x} \|_1, \text{ s.t. } \hat{y} = \hat{D} \hat{x},$$

where $\hat{D} \in \mathbb{H}^{M \times N}$ is the entire quaternion dictionary built up by the $N$ quaternion atoms. The symbol $\| \hat{x} \|_1 := \sum_i |\hat{x}_i|$ denotes the $l_1$-norm of the quaternion vector, which is naturally based on the $l_1$-norm of the real-valued vector. Following [9], we use Eq. 8 to formulate the new QWLasso model as follows:

$$\hat{x} = \arg \min_{x \in \mathbb{H}^N} \left\{ \left\| \hat{y} - \hat{D} \hat{x} \right\|_2^2 + \lambda \| \hat{x} \|_1 \right\},$$

where $\lambda > 0$ is the regularization parameter, which controls the sparsity of $\hat{x}$ and provides a trade-off between the sparsity penalty and the fidelity term. To solve problem (9), we propose a novel gradient-based algorithm called QFISTA.

2) QUATERNION SPARSE CODING STAGE

Paper [29] demonstrated that calculations in the AQs are considerably more complex than those in real number system. This greatly increases the complexity of solving the optimization problem in Eq. 9 with the novel SRCQW method. Therefore, to simplify the solution of Eq. 9 (also point 2 in Algorithm 2), we map the quaternion dictionary of low-frequency sub-bands to a real-valued dictionary. To develop the mapping, we apply Eqs. 25 and 27 and rewrite equation $\hat{y} = \hat{D} \hat{p}$ in the form:

$$\hat{y}^0 + \hat{y}_1 i + \hat{y}_2 j + \hat{y}_3 k = (\hat{D}_0 + \hat{D}_1 i + \hat{D}_2 j + \hat{D}_3 k)(\hat{x}_0 + \hat{x}_1 i + \hat{x}_2 j + \hat{x}_3 k).$$

(10)

Then, using Eq. 29, the above Eq. 10 can be rewritten as:

$$\hat{y}^0 + \hat{y}_1 i + \hat{y}_2 j + \hat{y}_3 k = \hat{D}_0 \hat{x}_0^0 - \hat{D}_1 \hat{x}_1^0 - \hat{D}_2 \hat{x}_2^0 - \hat{D}_3 \hat{x}_3^0 + \hat{D}_1 \hat{x}_0^1 \hat{i} + \hat{D}_2 \hat{x}_1^1 \hat{j} + \hat{D}_3 \hat{x}_2^1 \hat{k}.$$  

(11)

By considering the coefficients of $i, j$ and $k$ in both sides of Eq. 11, we transform it to the following one:

$$\begin{pmatrix} \hat{y}_0^0 \\ \hat{y}_1^0 \\ \hat{y}_2^0 \\ \hat{y}_3^0 \\ \hat{y}_0^i \\ \hat{y}_1^i \\ \hat{y}_2^i \\ \hat{y}_3^i \\ \hat{y}_0^j \\ \hat{y}_1^j \\ \hat{y}_2^j \\ \hat{y}_3^j \\ \hat{y}_0^k \\ \hat{y}_1^k \\ \hat{y}_2^k \\ \hat{y}_3^k \end{pmatrix} = \begin{pmatrix} D_0^0 & D_1^0 & D_2^0 & D_3^0 \\ D_0^1 & D_1^1 & D_2^1 & D_3^1 \\ D_0^2 & D_1^2 & D_2^2 & D_3^2 \\ D_0^3 & D_1^3 & D_2^3 & D_3^3 \end{pmatrix} \begin{pmatrix} x_0^0 \\ x_1^0 \\ x_2^0 \\ x_3^0 \\ x_0^i \\ x_1^i \\ x_2^i \\ x_3^i \\ x_0^j \\ x_1^j \\ x_2^j \\ x_3^j \\ x_0^k \\ x_1^k \\ x_2^k \\ x_3^k \end{pmatrix}.$$  

(12)

Since $D^e (e = 0, 1, 2, 3)$ are real-valued matrices we solve Eq. 12 in the field of real numbers. To develop the solution to Eq. 9, we use the operators $P$ and $Q$, which naturally arise from Eq. 12. Consider $\hat{D} = D^0 + D^1 i + D^2 j + D^3 k \in \mathbb{H}^{M \times N}$, $D^e \in \mathbb{R}^{M \times N}$, which is a quaternion matrix. There exists a unique operator $P : \mathbb{H}^{M \times N} \rightarrow \mathbb{R}^{4M \times 4N}$:

$$P(\hat{D}) := \begin{pmatrix} D_0^0 & D_1^0 & -D_2^0 & -D_3^0 \\ D_1^1 & D_2^1 & -D_3^1 & -D_0^1 \\ D_0^2 & D_1^2 & D_2^2 & D_3^2 \\ D_1^3 & D_2^3 & D_3^3 & D_0^3 \end{pmatrix} \in \mathbb{R}^{4M \times 4N}.$$  

(13)

The existence and uniqueness of $P$ follow from Eqs. 10-12. Accordingly, for any quaternion vector $\hat{x} = x^0 + x^1 i + x^2 j + x^3 k, \hat{x} \in \mathbb{H}^N, x^e \in \mathbb{R}^N$, $e = 0, 1, 2, 3$, we define the operator $Q : \mathbb{H}^N \rightarrow \mathbb{R}^{4N}$ such that:

$$Q(\hat{x}) := \begin{pmatrix} (x^0)^T \\ (x^1)^T \\ (x^2)^T \\ (x^3)^T \end{pmatrix} \in \mathbb{R}^{4N}.$$  

(14)

Let $Q^{-1}$ be the inverse of $Q$, i.e. $Q^{-1}(Q(\hat{x})) = \hat{x}$. As proven in [29], the operators $P$ and $Q$ possess the properties:

1) $P$ and $Q$ are linear: $\| Q(\hat{x}) \|_2 = \| \hat{x} \|_2, \forall \hat{x} \in \mathbb{H}^N$.

2) $Q(P(\hat{D})Q(\hat{x})) = \| P(\hat{D})Q(\hat{x}) \|_2, \forall \hat{D} \in \mathbb{H}^{M \times N}, \forall \hat{x} \in \mathbb{H}^N$. 

VOLUME 10, 2022
Next, we use the operator $qmat(Q(\hat{s})) : \mathbb{R}^{4N \times 1} \rightarrow \mathbb{R}^{N \times 4}$ [29]:

$$qmat(Q(\hat{s})) := [x^0, x^1, x^2, x^3] \in \mathbb{R}^{N \times 4}. \quad (15)$$

Now we apply the operators $P, Q$, and $qmat$ on the QWlasso model in Eq. 9 and map it from the AQ to the field of real numbers:

$$Q(\hat{s}) = \arg\min_{Q(s) \in \mathbb{R}^{4N}} \left\{ \|Q(\hat{y}) - P(\hat{D})Q(s)\|_2^2 + \lambda\|qmat(Q(s))\|_{1,2} \right\}$$

(16)

Note that in Eq. 16 we apply the $l_{1,2}$-norm because the quaternion $\hat{s}$ in the $l_1$ term of Eq. 9 is mapped to a matrix in Eq. 16. For the sake of simplicity, we denote $u := Q(\hat{y}) \in \mathbb{R}^{4M}, B := P(\hat{D}) \in \mathbb{R}^{4M \times 4N}, s := Q(\hat{s}) \in \mathbb{R}^{4N}$, and rewrite Eq. 16 in the following concise form:

$$\hat{s} = \arg\min_{s \in \mathbb{R}^{4N}} \left\{ \|u - Bs\|_2^2 + \lambda\|qmat(s)\|_{1,2} \right\} \quad (17)$$

Then, we split Eq. 17 into two parts and denote them: $f(s) = \|u - Bs\|_2^2$ and $g(s) = \lambda\|qmat(s)\|_{1,2}$. Hence, Eq. 17 becomes: $\hat{s} = \arg\min_{s \in \mathbb{R}^{4N}} [f(s) + g(s)]$. Note that $f(s)$ is smooth since its second derivative exists. It implies that in the vicinity of $s$, $\exists w \in \mathbb{R}^{4N}$ and a Lipschitz constant $L$, which define a quadratic form $\varphi(s)$ that approximates $f(s)$:

$$f(s) \leq f(w) + \nabla f(w)^T (s - w) + \frac{L}{2} \|s - w\|_2^2 \approx \frac{L}{2} \|s - \left( w - \frac{1}{L} \nabla f(w) \right) \|_2^2 = \varphi(s). \quad (18)$$

Inspired by the results obtained with the FISTA method [31], we develop its version to operate in the AQ. Unlike the original FISTA, which uses $l_1$-norm, we implement Eq. 17 with the $l_{1,2}$-norm. The proposed QFISTA method is applied to approximate the solution of Eq. 9 with the help of Eq. 18. In order to implement the novel QFISTA and calculate the sparse code $\hat{x}$, we develop the novel Algorithm 3 and its sub-procedure Algorithm 4 to solve equation $p_L(w) = \arg\min_{h(s)} \{h(s)\}$ for $s$, where $h(s) = g(s) + \varphi(s)$.

To apply Algorithms 3 and 4, the formulas given below are calculated, where $w, L$, and $\nabla f(w)$ are real-valued constants:

$$\nabla f(s) = -2B^T (u - Bs) = 2 \left( B^T B - B^T u \right). \quad (19)$$
$$\nabla g(s) = \lambda \left[ \frac{\partial g(s)}{\partial s_1} + \frac{\partial g(s)}{\partial s_2} + \cdots + \frac{\partial g(s)}{\partial s_4} \right] \quad (20)$$
$$\nabla \varphi(s) = L \left( s - \left( w - \frac{1}{L} \nabla f(w) \right) \right). \quad (22)$$

3) LABEL ASSIGNMENT STAGE

To identify the label of the test sample, we compute the class dependent residual for each class $k$, $1 \leq k \leq K$ as: $r_k(\hat{y}) = \|\hat{y} - D_k \hat{x}\|_2$, where $D_k$ denotes the characteristic function that selects the coefficients from $\hat{x}$ associated with class $k$. For $\hat{x} \in \mathbb{H}^N$, the non-zeros entries of $D_k(\hat{x}) \in \mathbb{H}^N$ are associated with the class $k$. Finally, the test quaternion vector $\hat{y}$ is assigned to the class providing the minimal residual.

VI. COMPUTATIONAL COMPLEXITY

The computational complexity for the SRCQW algorithm is estimated through calculating the number of arithmetic operations required to determine $s_{k,z+1}$ with Algorithms 3 and 4 for solving Eq. 9 in Algorithm 2. For the purpose of simplicity, we assume that i) the number of training samples (number of dictionary atoms) is same for every class and equals to $n = \max\{N_k\}_{k=1}^K$; ii) Each iterative algorithm requires the same number of iterations $q$ to converge.

We observe, from Eqs. 19 and 20, that the most computationally expensive expressions are: $B^T B - B^T u$ and $\nabla g(s) + \nabla \varphi(s)$, where $B^T u$ can be precomputed. Note $B \in \mathbb{R}^{4M \times 4N}, u \in \mathbb{R}^{4M}$, where $N$ denotes the total number of atoms in the quaternion dictionary. Hence, we determine that $B^T u = O(NM)$. Since $N = Kn$, where $K$ is the number of
of classes, we arrive at $B^Tu = O(nKM)$. Consider that $s \in \mathbb{R}^{4N}$. There are two different ways to calculate the computational complexity of the chain: $B^T(Bs) = (B^T B)s$. For the left one we have $B^T(Bs) = O(MN + MN) = O(MN)$, while for the right chain of matrix multiplication we have $(B^T B)s = O(NMN + N^2) = O(N^2M)$. Therefore, we select $O(NM) = O(nKM)$ for the computational complexity of $B^T B$s.

Consider Eqs. 20-22. It is straightforward to show that $\nabla g(s) = O(nK)$ and $\nabla \psi(s) = O(nK)$. Now, consider that each of the Algorithms 3 and 4 requires $q$ iterations to converge. Therefore, following Eqs. 17, 19, 20, we evaluate the computational complexity of the SRCQW method as:

$$
\begin{align*}
B^T u &+ q [O(nKM) + q [O(nK) + O(nK)]] \\
B^T B s &+ \nabla g(s) + \nabla \psi(s) \\
&= O(nKM) + O(qnKM) + O(q^2nK) \\
&\approx O(qnKM) + O(q^2nK).
\end{align*}
$$

The last big-O expression implies that the SRCQW method has a computational complexity that equal to the max $\{O(qnKM), O(q^2nK)\}$.

VII. EXPERIMENTAL RESULTS

In order to validate the SRCQW capabilities, we carried out experiments on five public datasets and compare its results with contemporary methods in the field including SRC [9], LC-KSVD [11], FDDL [12], LRSDL [32], and SRWC [20] as well as NNs, namely Centralined PLN [39], Distributed PLN [39], PCANet1 [40], and Deep/Wide Net [41].

A. CROSS-VALIDATION

To evaluate the efficiency of the proposed SRCQW method, we applied the Monte Carlo cross-validation [42]. It randomly splits the dataset into training and test sets and repeats this process $k$ times. For each split, a sample appears in either the training set or the test set, but not in both. The results are then averaged over the $k$ splits. The advantage of using the Monte Carlo cross-validation is that it can substantially reduce the variance of the split sample error estimate and the proportion of the training-test random splits does not depend on the number $k$ [43]. In our experiments, $k$ is set to 10.

B. DETAILS OF DATASETS

The datasets we used for validation are: the Extended YaleB face [44], the AR face [45], the AR gender [45], the SVHN [46] and a multi-class object category – the COIL-100 [47]. Figure 5 shows examples from these datasets, whose descriptions are summarized in Table 1.

1) The Extended YaleB dataset [44] contains 2414 face images of 38 people under various poses and lighting conditions. For each person, we randomly select 30 images for training, making a total number of 1140 training and 1274 test images.

2) The AR face dataset [45] contains frontal faces of 126 people and 26 images are available for each face. In this experiment, we use 100 classes (50 males and 50 females). For every person, 20 face images are randomly selected for training and 6 for testing.

3) The AR gender dataset [45] is generated by choosing 14 non-occluded images per individual for 100 people (50 males and 50 females), having total of 1400 images. We randomly select 350 images from each class (male/female) for training; the remaining 700 images are used for testing.

4) The COIL-100 dataset [47] consists of 7200 color images of 100 objects. Hence 72 images are available for every 3D object, under different view points and illumination conditions. For every experiment, we randomly select 10 images of each object for training and the rest (62 images) for testing.

5) The Street View House Number (SVHN) dataset [46] consists of over 600,000 labeled digit images, which are taken from house numbers in Google Street View images. For every experiment, we randomly select 1600 images for training and 200 for testing, from a subset of total 1800 images out from 600,000 ones.

C. EFFECT OF VARYING $\lambda$ ON THE CLASSIFICATION ACCURACY

Figure 6 shows the performance of the proposed SRCQW on three datasets using different values of parameter $\lambda$ (Eq.9, Algorithm 3) from the set $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ recommended by [29]. We observe that the highest accuracy is achieved with $\lambda = 10^{-3}$ for the face datasets and with $\lambda = 10^{-2}$ for COIL-100. Hereafter, we utilize these $\lambda$ values.

---

**TABLE 1. Description of the five datasets. In columns 3, 4, and 5: number of classes, total number of training samples, and total number of test samples, respectively.**

| Database      | Image size | #Class | #Training | #Test  | Feature dim |
|---------------|------------|--------|-----------|--------|-------------|
| YaleB         | 192x188    | 38     | 1140      | 1274   | 2400        |
| AR face       | 165x120    | 100    | 2000      | 600    | 2200        |
| AR gender     | 165x120    | 2      | 700       | 700    | 1700        |
| COIL-100      | 128x128    | 100    | 1000      | 6200   | 400         |
| SVHN          | 32x32      | 10     | 1600      | 200    | 180         |

---

**FIGURE 5. Examples from the five datasets.**

(a) YaleB [44]  
(b) COIL-100 [47]  
(c) AR face [45]  
(d) AR gender [45]  
(e) SVHN [46]
TABLE 2. Mean accuracy (%) of SRCQW and contemporary SR-based methods. Numbers in parentheses show the training set size per class. Best results are in bold.

| Method       | SRC [9] | LC-KSVD1 [11] | LC-KSVD2 [11] | PDDL [12] | LRSiDL [32] | SRWC [20] | SRCQW |
|--------------|---------|---------------|---------------|-----------|-------------|-----------|-------|
| Ext. YaleB (30) | 97.5    | 97.1          | 97.8          | 97.5      | 98.8        | 98.1      | 98.9  |
| Ext. YaleB (55) | 99      | -             | -             | 98.77     | 99.3        | 98.7      | 99.85 |
| AR (20)      | 97.6    | 97.8          | 97.7          | 96.2      | 98.8        | 98.4      | 98.9  |
| ARgender (356) | 92.5    | 88.4          | 90.1          | 93.7      | 95.4        | 98.3      | 98.6  |
| COIL-100 (110) | 81.2    | 81.4          | 81.4          | 77.5      | 84.4        | 81.3      | 84.8  |
| SVHN (160)   | 15.71   | 22.5          | 35.3          | 22.54     | 35.3        | 46        | 55.9  |

D. OVERALL CLASSIFICATION ACCURACY

To evaluate the SRCQW classification performance, we compare it with five contemporary SR-based methods on the five datasets. They represent four different kind of subjects (Fig. 5) in four different sizes (Table 1): human faces, genders, objects, and numbers. The chosen methods comprise a fundamental one [9], its elaborations [11], [12], [32], and a SRC in the wavelet domain [20]. For each method we apply the Monte-Carlo cross validation approach with $k = 10$ in Matlab environment. The average classification rates are reported in Table 2 for training/test images reported in Table 1. It is evident that the proposed SRCQW outperforms the competitors in all databases. In particular, a substantial improvement, up to 9.9% has been made for the SVHN dataset, while in the classification of the Ext. YaleB database, the SRCQW achieved the very high 99.85%. With ARgender dataset, SRCQW has made an improvement of 2.1% compared to SRWC method.

To further reveal the advantages of the proposed SRCQW, we compare it with six contemporary NNs approaches (Table 3). In [39], the authors trained a large deep architecture with a progressive learning network in a distributed setup and used the ADMM optimizing algorithm. PCANet1 [40] is a deep learning network, which is constructed on cascaded PCA, binary hashing, and blockwise histograms. The authors employed PCA to learn multistage filter banks. Then binary hashing and block histograms for indexing and pooling were implemented. In [41], deep and wide convolutional NNs (CNNs) are implemented within the Energy Efficient Deep Neuromorphic Networks framework released by IBM in 2016. Table 3 reveals that the proposed SRCQW outperforms the five NNs on the two face datasets despite using less training images, as it can be seen in the penultimate row from Table 3. For instance, by training the SRCQW with 1596 images we obtained an accuracy of 99.6% (see Table 3), while training with 1140 images we obtained 98.9%. When using nearly the same number of training images, as it is depicted in the bottom row from Table 3, SRCQW achieves very high accuracies of 99.6% on the YaleB and 98.5% on the AR classification. With COIL-100 dataset, SRCQW ranks 2nd in the table.

E. ACCURACY VERSUS FEATURE DIMENSIONS

In order to study the classification accuracy of the novel SRCQW method, we conducted experiments with varying dimensions of the atoms used to create the dictionary. For this
purpose, we used the databases YaleB, AR face, AR gender, and COIL-100 and illustrated the results in Figure 7. The maximum overall accuracies on the Ext. YaleB, AR face, AR gender, and COIL-100 are 98.9%, 98.92%, 98.57%, and 84.74% respectively, obtained for dimensions 2400, 2200, 1700, and 400 respectively. These results tell that if the dimension of the atom increases, the accuracy of face recognition increases as well, while in case of COIL-100, the accuracy of objects recognition decreases when the dimension increase from 400 to 1100.

F. ACCURACY VERSUS SIZE OF TRAINING SET

In this sub-section, we study the SRCQW capabilities and robustness as a function of the size of the training set. Hence, we conducted additional experiments where we vary the number of training samples per class. For the purpose of comparison, we use Ext. YaleB, AR face, AR gender, and COIL-100 datasets. Figure 8 shows that the accuracy curve of SRCQW is above the curves of the other methods. Moreover, the novel SRCQW method achieved an accuracy of 90% on the YaleB and 96% on the AR face databases using only 10 training samples.

G. SPARSITY VISUALIZATION

Finally, we visualize and compare the sparsity of the representation coefficients obtained from the proposed SRCQW, SRWC, and SRC methods using the visualization of the sum of absolute sparse codes for different test samples from one and the same class. One testing class from the AR face dataset is used for illustration. For this purpose, we calculate the sum of sparse codes of six test samples from ‘class 50’ in the AR face dataset. Each sparse code has 2000 entries as illustrated along the x axis in Figure 9. Each colored rectangle from the horizontal bars represents one class (1 to 100 classes) for a subset of dictionary atoms. One can see that the three graphs possess high peaks at the 50th colored rectangle, which means that the test samples are well labeled as ‘class 50’. From Fig. 9 one may observe that the proposed SRCQW provides the highest discrimination between the coefficients associated with ‘class 50’ and those associated with other classes. In the present study, we go further and show that the SRCQW improve sparsity, which leads to the high discriminative ability.

Another way to visualize the sparsity is the sparseness measure proposed in [48]. We adapt this concept for a quaternion sparse vector \( \mathbf{x} \) as follows:

$$\text{sparseness}(\mathbf{x}) = \frac{\sqrt{N} - \|\mathbf{x}\|_1/\|\mathbf{x}\|_2}{\sqrt{N} - 1}, \quad (23)$$

where \( N \) is the dimension of \( \mathbf{x} \). The bigger \( \text{sparseness}(\mathbf{x}) \) is, the sparser \( \mathbf{x} \) is [48]. To illustrate this concept, we apply Eq. 23 and calculate the sparseness values of the sparse codes obtained with the AR face dataset. With 600 test samples (Table 1), we can estimate 600 quaternion sparse codes and then calculate 600 corresponding sparseness values. These values are illustrated by the histogram in Figure 10. One can see that the sparseness values of the SRCQW are averagely bigger than those of the SRWC and SRC (0.64 vs 0.62 and 0.57). More precisely, the largest sparseness value for each of the methods SRCQW, SRWC, and SRC is 0.752, 0.716, and 0.665, respectively (Figure 10). Hence, we conclude that SRCQW provides the highest sparseness among the three methods, which proves its highest discriminative ability.

H. CONVERGENCE RATE

To compare the convergence rate of the proposed SRCQW versus conventional SRC [9] and SRWC [20] methods, we apply them on the AR face dataset. Figure 11 shows the cost functions and runtime of the three methods on the interval of \([0, 300]\) iterations. One may observe that SRCQW has smaller rate of convergence. The higher cost for SRCQW...
comes from the fact that it works in the 4D space of the AQ, and is a trade off for the very high accuracy of 99.85% and 99.6% (Table 2,3).

VIII. DISCUSSION AND CONCLUSION
The main contribution of this paper is the development of the novel and efficient sparsity-promoting SRCQW method for image classification in the AQ, where features are described by the QWT low-frequency sub-bands. Hence, the SRCQW method is the first SR-based one that uses information from the quaternion wavelet domain to solve the QWLasso minimization problem for classification in the 4D space of the AQ. As part of the SRCQW, we developed the novel gradient-based method QFISTA and QWlasso to calculate the sparse vector in the AQ defined over wavelets. Also, for the QFISTA method, we determined an upper bound of f(s) (Eq. 18), which simplifies the calculations.

Conducting the classification in the 4D space of the AQ increases the sparsity and the accuracy due to the additional information that comes from the orientation of the 4D vector constructed by the quaternion coefficients. The above statement holds because the additional information increases the separability of the atoms, resulting in improved classification accuracy (99.85% with YaleB), which comes on the expense of increased computational complexity. To the best of our knowledge, no other method has reported such an accuracy of face identification. Another advantage of the SRCQW compared to other SRC, is that the QWT has near shift-invariance and promotes higher sparsity of features than DWT. Hence, the QW domain makes the SR a very suitable representation learning method, providing robustness and high accuracy of classification.

Moreover, the SRCQW method is robust according to the number of training images as shown in Figure 8. For example, to classify COIL-100 dataset, the SRCQW used only 10 images per object for training and 62 for testing. This observation also shows the advantage of the SRCQW method over the NNs, which are very powerful and flexible tools for classification but need a large number of training samples to achieve high accuracy. In this sense we may state that, the NNs are useful classifiers if hundreds or at least tenth of thousands of training samples are available, while the SRCQW approach proved (see Table 3) to be efficient when medium-sized training dataset is available.

In view of the FISTA algorithm [31] and owing to the convexity of Eq. 17, the proposed QFISTA guarantees global convergence, which is not the case for the methods in [28], [29]. The same statement holds for the NNs which usually need a number of parameters and functions to be correctly selected. On the other hand, the SRCQW method needs the selection of a single parameter (λ, needed to implement Algorithms 3 and 4 with Eq. 21)) for which a few values should be tested (section VII-C).

Recall, we validated the SRCQW method advantages and contributions using five image databases, which contain four kind of subjects: human faces, genders, house numbers on streets, and objects. The database images are of four different sizes: 192 × 168; 165 × 120; 128 × 128 and 32 × 32 and a number of training/testing splits were used for conducting the classification experiments.

A drawback of the novel SRCQW is its relatively high computational complexity which is the trade off for the very high accuracy of 99.85%.

In our future study we will test the SRCQW method on medical datasets of larger and more complicated images. Further, this method promises a strong connection with deep learning for the purpose of non-linear mapping.
Clifford product of quaternions

$$\tilde{ab} = (a_0b_0 - a_1b_1 - a_2b_2 - a_3b_3)$$
$$+ (a_1b_0 + a_0b_1 - a_3b_2 + a_2b_3)i$$
$$+ (a_2b_0 + a_3b_1 + a_0b_2 - a_1b_3)j$$
$$+ (a_3b_0 - a_2b_1 + a_1b_2 + a_0b_3)k$$  \( (29) \)

**ACKNOWLEDGMENT**

The authors thank the Research Commission at the Université Sorbonne Paris Nord for the partial support of this research through the visiting professorship for Dr. N. M. Sirakov. We also thank Vu et al. [32] for providing a toolbox, which includes the source code of SRC, LC-KSVD, FDDL, and LRSDDL methods.

**REFERENCES**

[1] C. C. Aggarwa, Data Classification: Algorithms Application. Boca Raton, FL, USA: Chapman & Hall, 2015.

[2] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification. Hoboken, NJ, USA: Wiley, 2001.

[3] C. M. Bishop, Pattern Recognition and Machine Learning, 1st ed. New York, NY, USA: Springer-Verlag, 2006.

[4] W. H. Zhao, X. Gao, “Multimedia intelligence: When multimedia meets artificial intelligence,” IEEE Trans. Multimedia, vol. 22, no. 7, pp. 1823–1835, Jul. 2020.

[5] T. Georgiou, Y. Liu, W. Chen, and M. Lew, “A survey of traditional and deep learning-based feature descriptors for high dimensional data in computer vision,” Int. J. Multimedia Inf. Retr., vol. 9, no. 3, pp. 155–170, Sep. 2020.

[6] A. Badri, “A survey on data-efficient algorithms in big data era,” J. Big Data, vol. 8, no. 1, pp. 1–54, Dec. 2021.

[7] B. A. Olshausen and D. J. Field, “Sparse coding with an overcomplete basis set: A strategy employed by V1?” Vis. Res., vol. 37, no. 23, pp. 3311–3325, 1997.

[8] M. Elad, Sparse Redundant Representations: From Theory to Application. Boca Raton, FL, USA: CRC Press, 2010.

[9] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 2, pp. 210–227, Feb. 2009.

[10] Q. Zhang and B. Li, “Discriminative K-SVD for dictionary learning in face recognition,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 2691–2698.

[11] Z. Jiang, Z. Lin, and L. S. Davis, “Label consistent K-SVD: Learning a discriminative dictionary for recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 11, pp. 2251–2264, Nov. 2011.

[12] M. Yang, L. Zhang, X. Feng, and D. Zhang, “Sparse representation based Fisher discrimination dictionary learning for image classification,” Int. J. Comput. Vis., vol. 109, no. 3, pp. 209–232, 2014.

[13] B. A. Olshausen, “Principles of image representation in visual cortex,” Vis. Neurosci., vol. 2, pp. 1603–1615, Oct. 2003.

[14] B. A. Olshausen and D. J. Field, “Emergence of simple-cell receptive field properties by learning a sparse code for natural images,” Nature, vol. 381, no. 6583, pp. 607–609, 1996.

[15] V. Liu, R. Kamaraswamy, L. Le, and M. White, “The utility of sparse representations for control in reinforcement learning,” in Proc. AAAI Conf. Artif. Intell., vol. 33, 2019, pp. 4384–4391.

[16] L. Le, R. Kamaraswamy, and M. White, “Learning sparse representations in reinforcement learning with sparse coding,” in Proc. 26th Int. Conf. Artif. Intell., 2017, pp. 2067–2073.

[17] X. Sun, N. M. Nasrabadi, and T. D. Tran, “Supervised deep sparse coding networks for image classification,” IEEE Trans. Image Process., vol. 29, pp. 405–418, 2019.

[18] M. Abavisani and V. M. Patel, “Deep sparse representation-based classification,” IEEE Signal Process. Lett., vol. 26, no. 6, pp. 948–952, Jun. 2019.

[19] D. M. Paiton, C. G. Frye, S. Y. Lundquist, J. D. Bowen, R. Zarcone, and B. A. Olshausen, “Selectivity and robustness of sparse coding networks,” J. Vis., vol. 20, no. 12, p. 10, Nov. 2020.
LONG H. NGO received the B.E. degree in electrical and electronics engineering from the Ho Chi Minh City University of Technology, Vietnam, in 2017, and the M.S. degree in engineering and innovation in image and network and the Ph.D. degree in machine learning and image processing from the Université Sorbonne Paris Nord, France, in 2018 and 2021, respectively. His research interests include machine learning and image processing.

NIKOLAY M. SIRAKOV received the Ph.D. degree from the Institute of Mechanics and Biomechanics, Bulgarian Academy of Sciences. He held positions with the same institution; International Laboratory of Artificial Intelligence, Slovak Academy of Sciences; CVRM-Instituto Superior Tecnico, Portugal; Northern Arizona University, USA. He is currently a Professor of mathematics and computer science with the Department of Mathematics, Texas A&M University-Commerce. Together with his collaborators and students, he developed methods and algorithms in the fields of active contours, vector fields, features extraction, classification, tracking objects in video, and prediction. The fields of application of his methods are skin lesions, neonates’ bradycardia, threat assessment, 3D subsurface objects reconstruction, and faces classification.

NIKOLAY M. SIRAKOV received the Ph.D. degree from the Institute of Mechanics and Biomechanics, Bulgarian Academy of Sciences. He held positions with the same institution; International Laboratory of Artificial Intelligence, Slovak Academy of Sciences; CVRM-Instituto Superior Tecnico, Portugal; Northern Arizona University, USA. He is currently a Professor of mathematics and computer science with the Department of Mathematics, Texas A&M University-Commerce. Together with his collaborators and students, he developed methods and algorithms in the fields of active contours, vector fields, features extraction, classification, tracking objects in video, and prediction. The fields of application of his methods are skin lesions, neonates’ bradycardia, threat assessment, 3D subsurface objects reconstruction, and faces classification.

MARIE LUONG received the Engineer Diploma degree from the Ecole Nationale Supérieure d’Electricité et de Mécanique de Nancy, in 1991, and the Ph.D. degree in automatic, diagnosis, signal processing from the L’Institut National Polytechnique de Lorraine, Lorraine University, France, in 1996. She is currently an Associate Professor with the Université Sorbonne Paris Nord, France. Her research interests include image analysis, machine learning, and medical imaging processing.

EMMANUEL VIENNET received the Ph.D. degree in multi-layered neural networks for image recognition and human faces detection from the University Paris Sud (now Université Paris-Saclay), in 1993. Since 2008, he has been a Full Professor at the L2TI, University Sorbonne Paris Nord, France, where he leads a research group on machine learning for images and complex data structures.

THUONG LE-TIEN (Member, IEEE) was born in Saigon, Ho Chi Minh City, Vietnam. He received the bachelor’s and master’s degrees in electronics engineering from the University of Technology, National University of Ho Chi Minh City (HCMUT), Vietnam, and the Ph.D. degree in telecommunications from the University of Tasmania, Australia. Since May 1981, he has been with the EEE Department, HCMUT. He spent three years in the Federal Republic of Germany as a Visiting Scholar with the Ruhr University from 1989 to 1992. He served as the Deputy Department Head for many years and had been the Telecommunications Department Head from 1998 to 2002. He had also appointed for the second position as the Director of Center for Overseas Studies from 1998 to May 2010. He is currently a Full Professor with HCMUT. He has published more than 190 scientific articles and the teaching materials for university students related to electronic circuits 1 and 2, digital signal processing and wavelets, antenna and wave propagation, and communication systems. His research interests include communication systems, signal processing, and electronic circuits.

* * *