Unsupervised Quaternion Feature Learning for Remote Sensing Image Classification

Vladimir Risojević, *Member, IEEE,* and Zdenka Babić, *Member, IEEE*

***Abstract*—Bag-of-words image representations based on local descriptors are common in image classification and retrieval tasks. However, in order to achieve state-of-the-art results, complex hand-crafted feature filters and/or support vector classifiers with nonlinear kernels are needed. Compared with hand-crafted features, unsupervised feature learning is a popular alternative, which results in feature filters adapted to the problem domain at hand. Although both color and intensity are important cues for remote sensing image classification and color images are commonly used for unsupervised feature learning, most of the existing algorithms do not take into account interrelationships between intensity and color information. We address this problem by using quaternion representation for color images and propose unsupervised learning of quaternion feature filters, as well as feature encoding using quaternion orthogonal matching pursuit. By using quaternion representation we are able to jointly encode intensity and color information in an image. We obtain local descriptors by soft thresholding and computing absolute values of scalar and three vector parts of the quaternion-valued sparse code. Local descriptors are pooled, power-law transformed and normalized, yielding the resulting image representation. The experimental results on UC Merced Land Use and Brazilian Coffee Scenes datasets are comparable or better than the state-of-the-art, demonstrating the effectiveness of the proposed approach. The proposed method for quaternion feature learning is able to adapt to the characteristics of the available data and, being fully unsupervised, it emerges as a viable alternative to both hand-crafted representations and convolutional neural networks, especially in application scenarios with scarce labeled training data.**

***Index Terms*—Remote sensing image classification, unsupervised feature learning, sparse image representations, quaternion image processing,**

I. Introduction

Hand-crafted local feature representations, such as scale-invariant feature transform (SIFT), and histograms of oriented gradients (HOG), dominate in the field of remote sensing image classification. An image representation is obtained by encoding the local features using a learned dictionary and spatially pooling the feature codes [1]–[3]. Although good classification accuracies have been obtained, hand-crafted descriptors are not adapted to the problem domain at hand and can be expensive to compute.

In the last decade, however, unsupervised feature learning has become an attractive alternative to hand-engineered representations. The main premise of unsupervised feature learning is that it is possible to obtain discriminative image features starting from raw pixel values. This idea stems from [4], where it has been shown that it is possible to obtain Gabor-like filters by applying sparse coding to natural image patches.

Both intensity and color are important visual cues for classification of remote sensing images. Previous work using hand-crafted descriptors [2], [5], has shown that combination of intensity and color information is beneficial for remote sensing image classification. On the other hand, most of the unsupervised feature learning algorithms, e.g. [6]–[8] produce filters that can encode either intensity or color information. Furthermore, filters that respond to color information are tuned to specific color antagonisms and do not take into account intensity information, which has adverse effect on classification accuracy. This observation was the starting point in [9], where quaternion representation for color images along with quaternion principal component analysis (QPCA) and K-means clustering was used for jointly encoding intensity and color information in aerial images. However, although high classification accuracy was obtained, outperforming more traditional approach of concatenating information from color channels, it was necessary to use support vector machine (SVM) classifier with nonlinear χ^2 kernel which resulted in high computational cost. This drawback can be mitigated by using kernel mapping [10], which increases memory requirements. On the other hand, approaches based on unsupervised feature learning using real-valued orthogonal matching pursuit (OMP) [11], with dictionaries learned using K-SVD algorithm [12], in conjunction with linear SVM classifiers, have achieved state-of-the-art classification accuracies on various image classification tasks [6], [7].

In this paper we address the question of learning features which capture interrelationships between intensity and color information in an image, as well as lowering computational complexity of the resulting classifier. In order to achieve this goal, we investigate local features obtained using raw quaternion pixel values, as well as their projections onto QPCA basis, with or without dimensionality reduction. Sparse coding of features is performed using quaternion orthogonal matching pursuit (Q-OMP). We perform the experiments with random dictionary, randomly sampled image patches, as well as dictionaries learned using quaternion K-means (QK-means) and quaternion K-SVD (QK-SVD) algorithms.

Manuscript received September 14, 2015; revised November 10, 2015; accepted December 28, 2015.

This research was supported in part by Norwegian Ministry of Foreign Affairs through HERD ICT NORBOTECH project under contract 2011/1370, and in part by the Ministry of Science and Technology of the Republic of Srpska under contract 19/6-020/961-187/14.

The authors are with the Faculty of Electrical Engineering, University of Banja Luka, 78000 Banja Luka, Republic of Srpska, Bosnia and Herzegovina (e-mail: vlado@etfbl.net, zdenka@etfbl.net)

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

When quaternion representation is used for color images, pixel values in red, green and blue color channels are regarded as a single entity – pure quaternion, and the interrelationships between color channels are embedded into definitions of image processing operators. The response of a quaternion feature filter is a full quaternion [13], whose scalar part is related to intensity, and vector parts correspond to red, green and blue color channels. In this way, the response of a single quaternion feature filter is able to indicate that, for example, there is both an edge in general, and that it is an edge between green and blue regions, at the specific location in an image. On the other hand, a single real-valued feature filter can encode this information only partially, that is, the existence of either an edge in general or a green-blue edge. In this way one part of the information is lost and cannot be used for image classification. When quaternion features are used it is possible to use only one part or complete information, which enables better discrimination between classes. This approach results in state-of-the-art classification accuracies using SVM classifier with a linear kernel. Furthermore, we achieve faster training and classification without penalizing memory usage. To the best of our knowledge, quaternion sparse representations have not been used for image classification thus far.

The main contributions of this paper are unsupervised learning of quaternion feature filters for computing local features and image representation based on sparse encoding of local features using quaternion orthogonal matching pursuit.

The rest of the paper is organized as follows. Related work is reviewed in Section II. In Section III, the quaternion-valued local features are introduced, and in Section IV the proposed image representation is presented. Experimental results are given in Section V. Section VI concludes the paper.

II. Related Work

Bag-of-words image representations based on SIFT or HOG descriptors are common in classification and retrieval tasks for general-purpose [14], as well as for remote sensing images [1], [2], [5], [15], [16]. These representations use vector quantization for image encoding, and non-linear classifiers must be used in order to obtain good classification accuracies. In order to overcome this drawback, in [17] sparse codes of SIFT descriptors were proposed for image representation instead of vector quantizaton. Coupled with spatial pyramid matching and linear SVM this approach yielded high classification accuracies on several common image classification benchmarks. Sparse coding of SIFT descriptors is applied to satellite image classification in [18] and [19].

In the earlier studies, e.g. [20], raw image patches were shown to yield inferior image representations to SIFT descriptors with regard to classification accuracy. Furthermore, in [3] image representations based on three types of features, namely raw pixel values, responses of oriented filters, and SIFT features, along with OMP encoding were evaluated for aerial image classification. The best results were obtained using SIFT features and the worst using raw pixel values. It should be noted that, although the dictionary for feature encoding was learned from examples, the best performance was obtained using hand-crafted features. However, in [6] it was shown that unsupervised feature learning from raw pixels based on pre-whitening of data using zero-phase component analysis (ZCA) and K-means clustering can result in classification accuracies comparable or better than those obtained using more complex deep architectures as well as hand-crafted features. Features for color images are obtained by applying the learning algorithm to concatenated color channels.

Applying principal component analysis (PCA) to image patches and clustering thus obtained low-dimensional local features also yielded improved results in medical [21] and remote sensing image classification [22]. More recently, a simple approach to unsupervised feature learning named PCANet has been proposed [23], which uses PCA, binary hashing and histograming in order to compute an image representation.

In the area of remote sensing image classification, in [24] the authors investigated feature learning using PCA and deep belief networks for classification of high resolution aerial images, and compared the obtained results to several well known hand-crafted features. Sparse coding on a local manifold of raw image patches was proposed in [8]. However, it requires nonlinear histogram intersection kernel in order to obtain state-of-the-art classification accuracy. Saliency information for unsupervised feature learning from local image patches was used in [25]. In [26], semisupervised learning of high-level features is proposed in order to overcome the problem of having few labeled training examples.

Quaternion representation of color images was proposed in [27], and used for face recognition [28], texture [29] and remote sensing image classification [30]. Quaternion principal component analysis (QPCA) has been introduced in [31] and [32]. In [13] QPCA of local image patches and K-means clustering were used for color texture segmentation, and in [9] for aerial image classification.

Recently, sparse representations for quaternion-valued signals were analyzed in detail in [33]. The authors paid attention to non-commutativity of quaternion multiplication and proposed quaternion versions of orthogonal matching pursuit algorithm based on left and right multiplication linear models for quaternion signals. In [34] quaternion-based sparse representation of color images was used for image reconstruction, denoising and inpainting, and in [35] for color image super-resolution. Quaternion-based sparse representation is obtained using quaternion K-SVD [36] for dictionary learning, and quaternion OMP for computing the sparse representation.

III. Unsupervised Quaternion Feature Learning

By learning feature filters from examples instead of relying on manually designed features, we can obtain an image representation that is better suited to the problem at hand. In this case, both the filters and dictionary for feature encoding are learned from unlabeled training examples.

In Fig. 1, an overview of our approach is shown. Images are represented using pure quaternions [27]

$$Q(x,y) = Q_r(x,y)\,i + Q_g(x,y)\,j + Q_b(x,y)\,k, \quad (1)$$

where $Q_r(x,y)$, $Q_g(x,y)$, and $Q_b(x,y)$ are pixel values in red, green, and blue color channels respectively, and (x,y)

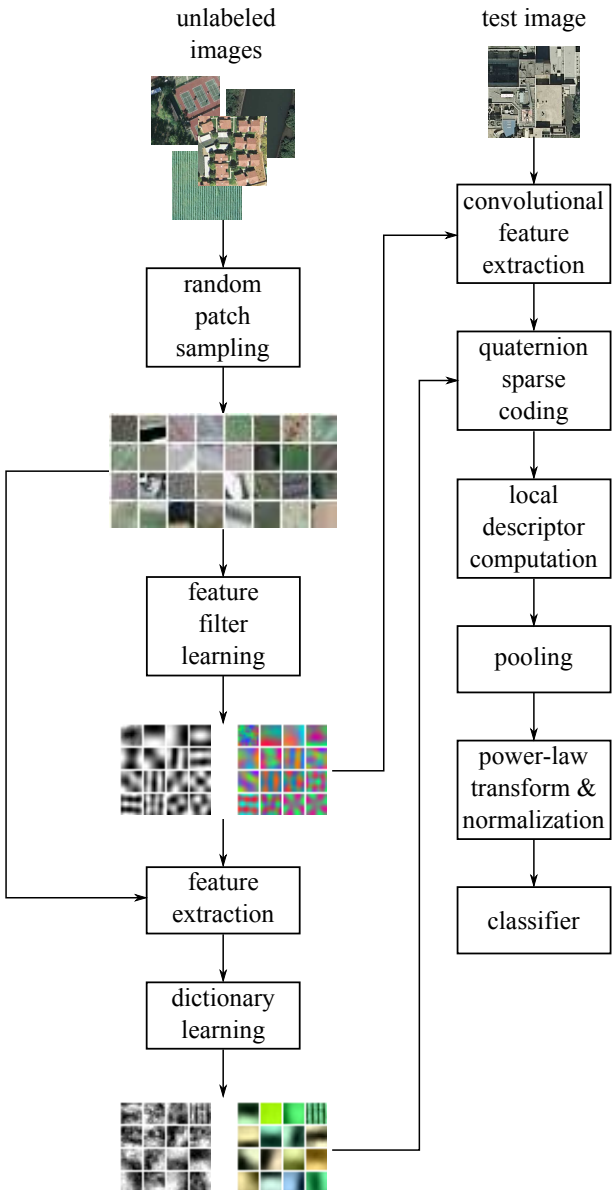


Fig. 1: Overview of the proposed approach for image classification using learned quaternion features.

are the pixel coordinates. Feature learning phase starts with randomly sampling a number of patches from training images and reshaping patch pixel values into a quaternion-valued vector form.

In [6] image patches are normalized to zero mean and standard deviation of one, and in [7] only normalization of mean is performed. We observed that, when quaternion-valued patches are used, the best results are obtained without patch normalization whatsoever.

In this paper we investigate two types of local feature filters: (i) raw pixel values, and (ii) projection of pixel values onto a quaternion PCA (QPCA) basis [31], [32]. Raw pixels are the simplest features, and in that case no feature learning is necessary. On the other hand, when QPCA is used, the feature learning stage consists of learning a QPCA basis.

Let $\mathbf{x}_i \in \mathbb{H}^n, i = 1, \ldots, N$ be the vectors of pixel values in image patches, where N is the total number of training patches. QPCA basis is computed by determining eigenvalues and eigenvectors of the covariance matrix of the training patches

$$\mathbf{S} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^T. \tag{2}$$

Let $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$ be the matrix of its eigenvectors, i.e. QPCA basis vectors, and $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ be the corresponding eigenvalues.

Thus obtained QPCA basis can be used for quaternion zero-phase component analysis (QZCA) whitening of the patches

$$\mathbf{x}_{white} = \mathbf{U} \operatorname{diag}\left(\frac{1}{\sqrt{\lambda_i + \epsilon}}\right) \mathbf{U}^T \mathbf{x}, \tag{3}$$

where $\operatorname{diag}(\cdot)$ denotes diagonal matrix and ϵ is a small positive regularization constant. QPCA basis can also be used for dimensionality reduction of local features

$$\mathbf{x}_d = \mathbf{U}_d^T \mathbf{x}, \tag{4}$$

where $\mathbf{U}_d^T$ is a matrix of basis vectors corresponding to d largest eigenvalues.

Since the covariance matrix (2) contains products of pure quaternions, the obtained QPCA basis will contain full quaternions. Given the QPCA basis, in the feature filtering phase, the patches from the input image are sampled using a sliding window, normalized, and projected onto the QPCA basis. This operation can be regarded as filtering the input image using the set of QPCA filters. The output of a QPCA filter is essentialy the convolution of a full quaternion filter with a pure quaternion image. In [13] has been shown that the result of this convolution is a full quaternion signal whose real part is related to intensity, and vector parts correspond to red, green and blue color channels.

IV. Quaternion Feature Encoding

In this section we review extensions of OMP and dictionary learning algorithms to quaternion-valued signals and propose an image representation based on quaternion-valued sparse codes of image patches.

A. Quaternion Orthogonal Matching Pursuit

Given the overcomplete dictionary $\mathbf{D} \in \mathbb{H}^{n \times M}$ containing M n-dimensional atoms, $\{\mathbf{d}_k\}_{k=1}^M$ as columns, sparse representation, $\mathbf{s} \in \mathbb{H}^M$, of the signal $\mathbf{y} \in \mathbb{H}^n$, is obtained by solving

$$\min_{\mathbf{s}} \|\mathbf{s}\|_0 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{D}\mathbf{s}, \tag{5}$$

where $\|\mathbf{s}\|_0$ is the l_0 pseudo-norm defined as the number of non-zero elements of $\mathbf{s}$. Since the dictionary atoms are L_2 normalized, the elements of the sparse representation reflect the energy of the corresponding atoms in the signal.

The optimization in (5) is NP-hard. Nevertheless, it is possible to obtain an approximate solution by constraining the sparse representation to have at most K nonzero elements. The optimization problem now becomes

$$\min_{\mathbf{s}} \|\mathbf{y} - \mathbf{D}\mathbf{s}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{s}\|_0 \leq K, \tag{6}$$

1: **function** $\mathbf{s} = \text{Q-OMPr}(\mathbf{y}, \mathbf{D})$
2: **Input:** Signal vector $\mathbf{y}$, dictionary $\mathbf{D}$
3: **Output:** Coefficient vector $\mathbf{s}$
4: Initialization: $k \leftarrow 1$, $\epsilon^0 \leftarrow \mathbf{y}$, dictionary $\mathbf{D}^0 \leftarrow \emptyset$
5: **repeat**
6: **for** $m \leftarrow 1, M$ **do**
7: Scalar Products: $C_m^k \leftarrow \langle \epsilon^{k-1}, \mathbf{d}_m \rangle = \mathbf{d}_m^H \epsilon^{k-1}$
8: **end for**
9: Selection: $m^k \leftarrow \arg\max_m |C_m^k|$
10: Active Dictionary: $\mathbf{D}^k \leftarrow \left[\mathbf{D}^{k-1}, \mathbf{d}_{m^k}\right]$
11: Active Coefficients: $\mathbf{s}^k \leftarrow \arg\min_s \|\mathbf{y} - \mathbf{D}^k \mathbf{s}^k\|_2^2$
12: Residue: $\epsilon^k \leftarrow \mathbf{y} - \mathbf{D}^k \mathbf{s}^k$
13: $k \leftarrow k + 1$
14: **until** stopping criterion
15: **end function**

Fig. 2: Right-multiplication Q-OMP algorithm.

where $K \ll M$ is a constant. Orthogonal matching pursuit [11] provides an efficient way to obtain a solution to (6), especially when the number of nonzero coefficients is small. The algorithm greedily selects the dictionary atoms and computes the corresponding representation coefficients in such a way as to minimize the representation error of the residual signal.

Quaternion extension of the OMP algorithm is proposed in [33]. Since quaternion multiplication is non-commutative, two models were considered – left and right-multiplication linear model. In this paper we chose to work with the right-multiplication linear model. We believe that the results would be similar had we chosen the left-multiplication linear model. Right-multiplication quaternion orthogonal matching pursuit (Q-OMPr) finds the sparse representation of the signal by solving the optimization problem (6), where $\|\cdot\|_2$ denotes the L_2 norm in the quaternion space.

The algorithm is summarized in Fig. 2. In each iteration of the algorithm one dictionary atom is selected based on the values of scalar products of the dictionary atoms $\mathbf{d}_m$ and the current signal residual ϵ^{k-1}. The selected dictionary atom corresponds to the scalar product with maximal modulus (Step 9). The active dictionary is subsequently updated with the selected atom and the active coefficients $\mathbf{s}^k$ are computed by orthogonal projection of the signal vector $\mathbf{y}$ onto the active dictionary $\mathbf{D}^k$ (Step 11)

$$\begin{aligned}\mathbf{s}^k &= \arg\min_{\mathbf{s}} \|\mathbf{y} - \mathbf{D}^k \mathbf{s}\|_2^2 \\ &= \left(\left(\mathbf{D}^k\right)^H \mathbf{D}^k\right)^{-1} \left(\mathbf{D}^k\right)^H \mathbf{y}. \end{aligned} \tag{7}$$

The pseudoinversion in (7) can be computed recursively as proposed in [11] and extended to quaternions in [33]. The new signal residual is computed in Step 12 and the described steps are repeated until the stopping criterion is met.

B. Dictionary Learning

Let $\mathbf{Y} \in \mathbb{H}^{n \times N}$ be the training set containing N quaternion-valued n-dimensional signals, $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$. Dictionary learning algorithm finds the dictionary $\mathbf{D} \in \mathbb{H}^{n \times M}$ containing M, n-dimensional quaternion atoms $\{\mathbf{d}_k\}_{k=1}^M$, such as to obtain the best sparse representation of the signals from the training set, i.e. by solving the optimization problem

$$\min_{\mathbf{D},\mathbf{S}} \|\mathbf{Y} - \mathbf{D}\mathbf{S}\|_F^2 \quad \text{s.t.} \quad \forall i,\ \|\mathbf{s}_i\|_0 \leq K, \tag{8}$$

where $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_N]$ is the matrix of sparse representation coefficients for the signals from the training set, and $\|\cdot\|_F$ denotes Frobenius norm.

When a dictionary is learned using QK-SVD algorithm, optimization in (8) is performed iteratively. First, the dictionary $\mathbf{D}$ is held fixed and the sparse representations of the signals from the training set are sought. This step can be performed using any sparse coding algorithm and we use Q-OMPr for efficiency reasons [12]. Next, given the sparse representation, the dictionary is updated one column at a time. All columns but one, $\mathbf{d}_k$, in the dictionary matrix $\mathbf{D}$ are held fixed, and new dictionary atom, $\tilde{\mathbf{d}}_k$, as well as new values of the corresponding representation coefficients are sought such as to obtain the largest reduction of the representation error. These steps are repeated until convergence is reached.

In Fig. 3, the examples of scalar and vector parts of dictionary atoms learned using QK-SVD are shown. We can see that the scalar parts of dictionary atoms capture intensity information, whereas the vector parts are sensitive to color information. Both the scalar and vector parts of a dictionary atom simultaneously contribute to the sparse representation, thus enabling joint encoding of intensity and color information.

When sparsity constraint K in (8) equals one, i.e. when each column of the coefficient matrix $\mathbf{S}$ has exactly one non-zero element, whose value is limited to be one, K-SVD algorithm reduces to K-means. In this case, sparse signal representation is computed by finding the closest dictionary atom using L_2 norm distance, setting the corresponding representation coefficient to one, and all the other coefficients to zeros. Then, in the dictionary update step, each atom is computed as the mean of the signals best represented with that particular atom.

C. Image Representation

Feature extraction proceeds in convolutional manner. Image patches of size $w \times w$ pixels are sampled with step size of r pixels and, after normalization, filtered using learned feature filters. Given a dictionary $\mathbf{D}$, we compute the sparse representation $\mathbf{s} \in \mathbb{H}^M$ for each image patch using Q-OMPr algorithm. It can be written as

$$\mathbf{s} = \mathbf{s}^{(0)} + \mathbf{s}^{(1)} i + \mathbf{s}^{(2)} j + \mathbf{s}^{(3)} k, \tag{9}$$

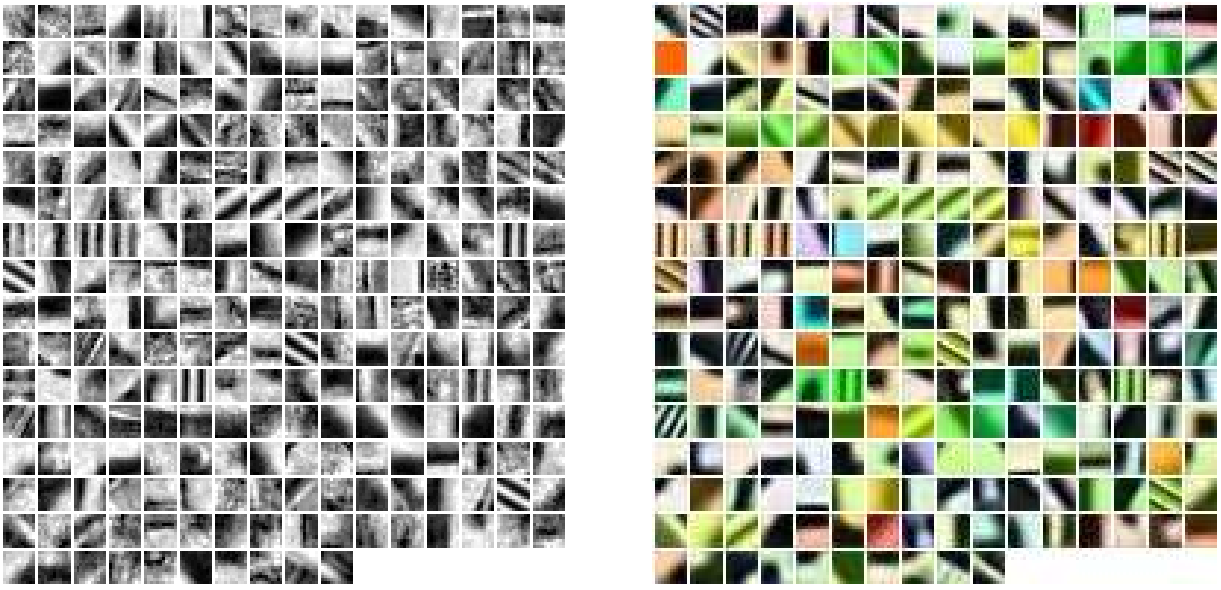
where $\mathbf{s}^{(l)} \in \mathbb{R}^M, l = 0, \ldots, 3$. Analogously to the QPCA feature filters, the scalar part of the sparse representation is related to the intensity information, and the three vector parts are related to red, green and blue color channels.

For each part, $\mathbf{s}^{(l)}$, of the quaternion-valued code, the elements of the local descriptor, $\mathbf{f}^{(l)}$, are computed in the following way

$$f_p^{(l)} = |s_p^{(l)}|, \tag{10}$$

$$f_{p+M}^{(l)} = \max\left(0, s_p^{(l)} - \theta^{(l)}\right), \tag{11}$$

$$f_{p+2M}^{(l)} = \max\left(0, -s_p^{(l)} - \theta^{(l)}\right), \tag{12}$$



(a) (b)

Fig. 3: Examples of (a) scalar, and (b) vector parts of quaternion-valued dictionary atoms. Dictionary $\mathbf{D} \in \mathbb{H}^{121\times250}$ is learned from 11×11 pixels image patches and each column from the dictionary is again reshaped to an 11×11 pixels patch.

where $s_p^{(l)}, p = 1, \ldots, M$ are the elements of the vector $\mathbf{s}^{(l)}$, and $\theta^{(l)}$ is the threshold value. By using thresholding, feature sparsity is further enforced. Therefore, for each part of the quaternion-valued dictionary atom, there are three elements in the local descriptor, resulting in $3M$-dimensional descriptors.

The obtained local descriptors, $\mathbf{f}^{(l)}$, are then spatially pooled by averaging them up over local image regions. The averaging region can be the whole image or some smaller image parts, e.g. image quadrants. In this paper we pool the descriptors from the whole image because it has been shown earlier [1] that, for remote sensing images, pooling schemes based on spatial pyramid do not improve classification accuracy significantly, while increasing descriptor dimensionality. Therefore, the parts of the image descriptor, $\mathbf{F}^{(l)}$, are obtained as

$$\mathbf{F}^{(l)} = \frac{1}{N_p} \sum_{q=1}^{N_p} \mathbf{f}_q^{(l)}, \tag{13}$$

where N_p is the number of patches in the pooling region. We then apply the power-law transform to the elements of $\mathbf{F}^{(l)}$

$$g\left(F_p\right) = F_p^{\alpha},\ p = 1, \ldots, 3M, \tag{14}$$

where $0 < \alpha < 1$. The power-law transform is closely related to Box-Cox transform [37] which has long been used in the statistics community. It is known to make the distribution of data more normal, which has been shown to be beneficial for image classification [38]. Next, we apply L_2 normalization to each part of the image descriptor

$$\overline{\mathbf{F}}^{(l)} = \frac{\mathbf{F}^{(l)}}{\sqrt{\|\mathbf{F}^{(l)}\|_2^2 + \delta}},\ l = 0, \ldots, 3, \tag{15}$$

where δ is a small positive number. Finally, the four parts of the image descriptor are stacked and the descriptor is again L_2 normalized. The resulting image descriptor is a real-valued vector of dimensionality $3 \times 4 \times M = 12M$.

V. Results

In this section we present an experimental evaluation of the proposed approach for unsupervised quaternion feature learning. We compare the classification accuracies obtained using quaternion and traditional, real-valued, local features. The impact of all components and hyper-parameters of the proposed image representation on classification accuracy is also analyzed. Finally, the obtained results are compared to the state-of-the-art in unsupervised feature learning and deep features obtained using pre-trained convolutional neural network (CNN), as well as to the results obtained using SIFT and HOG descriptors.

A. Experimental Setup

For the experiments in this paper we use two publicly available datasets. The first one is UC Merced Land Use (UCM) dataset[1] containing high resolution aerial images manually classified into 21 land use classes: agricultural, airplane, baseballdiamond, beach, buildings, chaparral, denseresidential, forest, freeway, golfcourse, harbor, intersection, mediumresidential, mobilehomepark, overpass, parkinglot, river, runway, sparseresidential, storagetanks, tenniscourt. There are 100 images in each class. All images are color, 256×256 pixels with spatial resolution of 30 cm (1 foot). This dataset was introduced in [1] for image classification, and in [16] it was used for image retrieval. We randomly pick 80 images from the each class for training the classifier, and the remaining 20 images are used for testing. The experiments are repeated

[1] http://vision.ucmerced.edu/datasets/landuse.html

five times and means and standard deviations of classification accuracies are reported.

The second dataset used in the experiments is Brazilian Coffee Scenes (Coffee) dataset[2]. It contains 2876 images taken by the SPOT sensor. The images are manually classified into two classes. The images containing at least 85% coffee pixels are assigned to the coffee class, and the images containing less than 10% of coffee pixels are assigned to the non-coffee class. Each of the classes contains 50% of all the images in this dataset. All images are color infrared, 64×64 pixels. This dataset was introduced in [39]. The authors have also provided five folds with equal contributions of coffee and non-coffee images. This dataset has been chosen because of different spectral and spatial properties compared to the UCM dataset, and because both the textural and spectral information are essential for good classifier performance.

As a classifier we use LIBSVM [40] implementation of linear SVM. Multiclass classification is performed by training one-versus-all SVMs for each class and assigning the test image to the class corresponding to the maximal SVM response.

B. UC Merced Land Use Dataset

1) Feature Learning and Dictionary Size: In this section, we perform experiments with three types of learned filters: (i) raw quaternion-valued patches, (ii) QZCA whitened patches, and (iii) QPCA transformed patches with dimensionality reduction. In all the cases patch size is set to 5×5 pixels and step size is one pixel. For dictionary learning and feature encoding we use QK-SVD and Q-OMPr algorithms, respectively. In both algorithms, we set the sparsity constraint to $K = 1$. We also experimented with other values of sparsity and obtained similar classification accuracies. However, by increasing the value of K, the sparse coding step will be slower, which will result in slower dictionary learning and image encoding.

The local descriptors are computed using (10)-(12). The thresholds, $\theta^{(l)}$, $l = 0, \ldots, 3$, in (11) and (12) can be selected using cross-validation. However, since there are four thresholds, the cross-validation in order to determine each threshold value would be impractical, and we chose to adaptively determine the thresholds for each image as the r-th percentile of the nonzero values of $|s^{(l)}|$, instead. We tested the values for $r \in \{50, 60, 70, 80, 90\}$ and obtained the best results for $r = 60$. In this way the number of hyper-parameters of the encoder is reduced from four to one. After the average-pooling, the power-law transform with $\alpha = 0.5$ is applied and the resulting image descriptors are L_2 normalized.

We compare the classification accuracies obtained using quaternion and real-valued local features. For computing real-valued local features, we first sample image patches and reshape them into vectors by concatenating pixel values from all color channels. Further processing is analogous to the quaternion-valued case except that we found out that the best results are obtained when the patches are normalized by subtracting their mean value. Again, three options are tested for local features: (i) raw patches, (ii) ZCA whitened patches, and (iii) PCA-based dimensionality reduction. Local features are then encoded using OMP with a dictionary learned using K-SVD. The elements of the local descriptor are computed using (10)-(12) for $l = 0$ only. The obtained local descriptors are subsequently pooled, power-law transformed with $\alpha = 0.5$, and L_2 normalized.

In Fig. 4 the impact of dictionary size on classification accuracy is shown. As pointed out in Section IV-C, resulting descriptor dimensionality for quaternion-valued sparse codes is $12M$, where M is the dictionary size. On the other hand, when real-valued sparse coding is applied to concatenated color channels, descriptor dimensionality is $3M$. For fair comparison we present the results obtained for the descriptors of same dimensionalities, which means that the corresponding dictionary sizes are not the same. We present the results obtained using raw and ZCA whitened patches.

We can see that image descriptors obtained using Q-OMPr consistently outperform descriptors obtained using real-valued OMP for 2-4%. It is interesting to note that in the real-valued case the classification accuracy of 88.48% is obtained for descriptor dimensionality of 12000, whereas in the quaternion-valued case classification accuracy of 88.29% is obtained for 1500-dimensional descriptors, i.e. eight times lower dimensionality. The reason for better performance of quaternion feature learning is its ability to jointly encode intensity and color information in images. Moreover, quaternion-valued sparse coding yields more efficient solutions in terms of computational and memory requirements along with better classification accuracy than in the real-valued case.

When the impact of patch whitening using ZCA on classification accuracy is considered, it can be seen that, in the quaternion case, whitening does not improve the classification accuracy, whereas in the real-valued case there is a slight improvement of classification accuracy when patch whitening is used. However, the improvement decreases with increase of descriptor dimensionality. This behavior can be explained by comparing signal representations using ZCA and OMP. ZCA decorrelates patch values by projecting them onto an orthogonal basis. On the other hand, the elements of the dictionary used with OMP do not need to be orthogonal. However, OMP ensures that the signal residual is orthogonal to all the atoms used for signal approximation. Since each atom is chosen in such a way as to be maximally correlated with the current signal residual, it follows that the algorithm chooses atoms which are as orthogonal as possible to the already used atoms. When larger dictionaries are used it is more likely that more orthogonal atoms will be used in signal decomposition thus making the signal decomposition closer to ZCA. Consequently, OMP results in a similar decomposition as ZCA. This is somewhat different from the case of K-means feature encoding [6], where patch whitening has beneficial effect on classification accuracy for all descriptor dimensionalities as discussed in detail in [41].

In order to get better understanding of the reasons for good performance of quaternion-valued local features in Fig. 5, producers' accuracies for real and quaternion-valued local features are shown. The dictionary size in the quaternion-valued case is 250, and in the real-valued case is 1000. Therefore, in both cases the descriptor is 3000-dimensional.

[2]http://www.patreo.dcc.ufmg.br/downloads/brazilian-coffee-dataset/

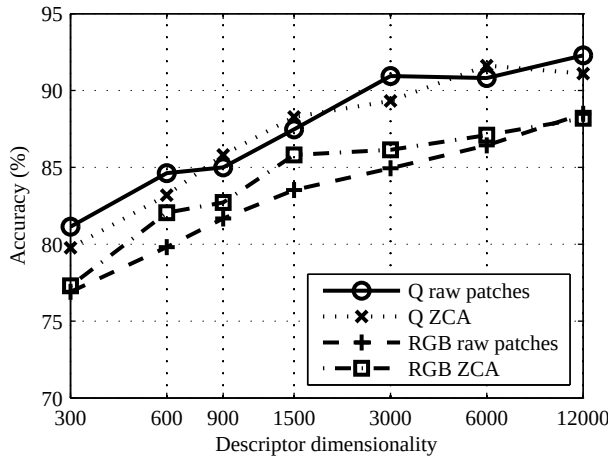


Fig. 4: The impact of image descriptor dimensionality on classification accuracy for UCM dataset.

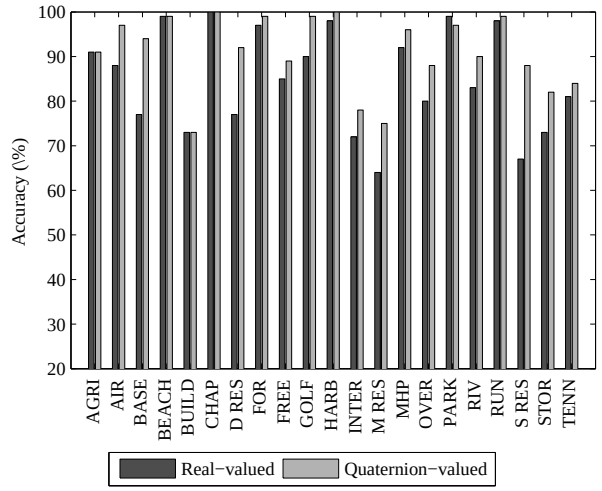


Fig. 5: Producers' accuracies when quaternion and real-valued local features are used for UCM dataset.

We can see that for almost all classes quaternion-valued features are better or leveled with real-valued ones. The only exception is the class parking lot. The largest improvements in classification accuracy of over 10% have been obtained for the three residential classes and baseball field. From the confusion matrices shown in Fig. 6 we can see that the number of confusions between the residential classes is considerably smaller for quaternion-valued local features.

When real-valued ZCA whitening is used for feature filtering, the resulting local features are 75-D real-valued vectors, whereas in the QZCA case the resulting local features are 25-D quaternion-valued vectors. Since for each quaternion value four real numbers have to be stored, a 25-D quaternion-valued feature vector corresponds to a 100-D real-valued vector. Therefore, the effective dimensionality of quaternion-valued local features is four times the dimensionality of the original quaternion-valued vector. Considering that PCA-based feature filtering can also include dimensionality reduction (4), we tested the impact of local feature dimensionality on classification accuracy. The results are shown in Fig. 7(a) for both the QPCA and real-valued PCA cases. Maximum dimensionality of real-valued local features is 75-D, and we report the obtained accuracies for dimensionalities up to 60-D. In both cases the resulting image descriptors are 3000-dimensional. We can see that, again, quaternion-valued features consistently outperform real-valued ones, with 5-7% larger accuracies. It is worth noting that when QPCA is used for dimensionality reduction of local features, 5-D quaternion-valued local features, i.e. local features with effective dimensionality of 20, result in less than 1% decrease of classification accuracy compared to raw patches.

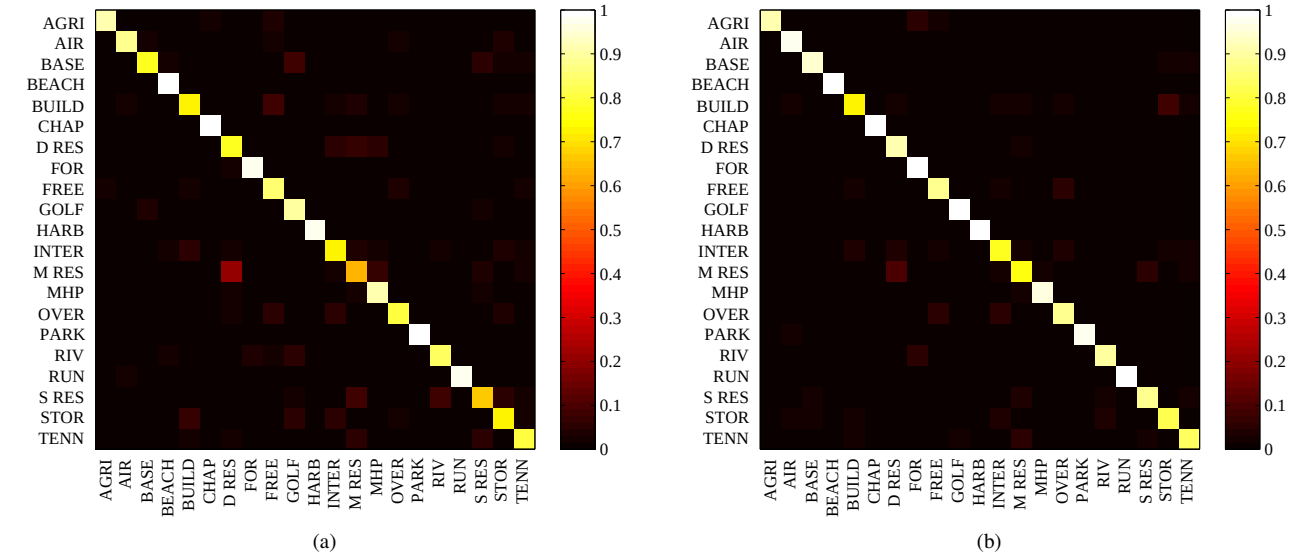
2) Impact of Dictionary Learning and Feature Encoding: In order to obtain better insight into the impact of dictionary learning and feature encoding algorithms on classification accuracy, we analyze the performances obtained using four unsupervised algorithms for learning the dictionary:

1) **Random dictionary (RAND):** The dictionary is populated with vectors of uniformly distributed unit quaternions. Thus obtained atoms are subsequently normalized to unit length.
2) **Random patches (RP):** The dictionary atoms are obtained by randomly sampling image patches from the training set and normalizing them to unit length.
3) **Quaternion K-means (QK-means):** The dictionary is learned using an extension of K-means algorithm to quaternion data.
4) **Quaternion K-SVD (QK-SVD):** The dictionary is learned using an extension of K-SVD algorithm to quaternion data, as described in Section IV-B.

For local feature encoding we tested the absolute values (ABS) of the scalar and vector parts of quaternion-valued sparse codes (10), and thresholded and rectified (TR) features, (11) and (12), as well as their combination. However, different encoding schemes result in image descriptors of different dimensionalities, namely ABS encoding yields $4M$-dimensional descriptors, TR encoding yields $8M$-dimensional descriptors and ABS+TR encoding yields $12M$-dimensional descriptors, where M is the number of dictionary atoms. For fair comparison, we fix the descriptor dimensionality to 3000 and learn dictionaries of appropriate size for each encoding scheme. In these experiments the patch size is fixed to 5×5 pixels with step of 1 pixel, and mean pooling (13) is used.

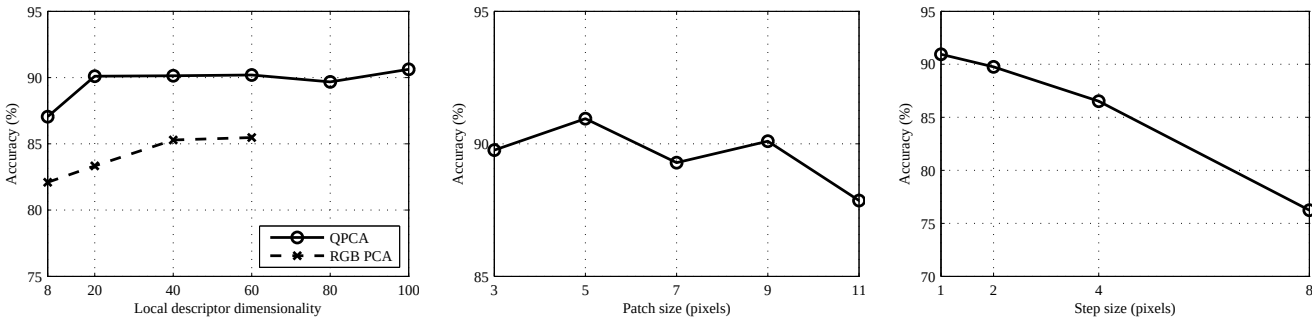
The obtained results are given in Table I. We can see that, except for the random dictionary, the differences in obtained accuracies using raw patches and ZCA whitened patches are small, regardless of the feature encoding method. It is interesting to note, however, that the random dictionary with whitened patches is consistently better than when raw patches are used. Moreover, in combination with thresholded features it yields classification accuracies of over 80%.

In almost all cases thresholded features outperform modulus values of sparse codes and the obtained accuracy is slightly higher when the combination of features is used. Moreover, when the combination of features is used, it is possible to use a smaller dictionary, thus making dictionary learning and sparse coding of patches faster. Dictionaries learned using QK-means and QK-SVD perform similarly, and, surprisingly, the dictionary composed of randomly sampled patches performs



(a) (b)

Fig. 6: Confusion matrices for (a) real-valued local features, and (b) quaternion-valued local features obtained on UCM dataset.



(a) (b) (c)

Fig. 7: Classification performance comparison under different parameter settings on UCM dataset. (a) The impact of local feature dimensionality on classification accuracy. (b) The impact of patch size on classification accuracy for quaternion-valued local features. (c) The impact of patch step size on classification accuracy for quaternion-valued local features.

TABLE I: The classification accuracies on UCM dataset for various combinations of dictionary learning and local feature encoding. The numbers in parentheses are dictionary sizes. All results are percent accuracy.

	Encoding					
	ABS (750)		TR (375)		ABS+TR (250)	
Dict.	raw	ZCA	raw	ZCA	raw	ZCA
RAND	54.62	79.52	65.38	83.48	62.33	82.67
RP	88.24	88.67	89.43	**90.48**	**91.00**	90.81
QK-means	**89.48**	86.76	88.48	89.09	89.81	89.57
QK-SVD	87.05	85.90	89.76	89.00	90.95	89.33

the same or slightly better than the learned dictionaries. These results suggest that the good performance is more due to the feature encoding than to the dictionary learning algorithm.

We also examine the influence of the feature pooling scheme and power law transform on the classification accuracy. Here, a dictionary with 250 atoms is learned using QK-SVD. We vary α in (14) in the range $[0.1, 1]$ with step of 0.1 for mean and max pooling. The results are shown in Fig. 8. Mean pooling consistently outperforms max pooling for 10-15%. This is in contrast to results in object recognition and general scene classification [42] where max pooling yielded better classification results compared to mean pooling. This difference stems from the fact that the pooling region used in aerial image classification comprises the whole image whereas in object recognition and general scene classification spatial pyramids are used. However, it has been already shown [1] that spatial pyramid pooling does not improve classification accuracy for aerial images. Furthermore, the power law transform is essential for good classification performance, with increase in accuracy of 5% for $\alpha = 0.5$ compared to the raw features, i.e. $\alpha = 1$.

3) Impact of Patch Sampling Parameters: We also analyzed the impact of patch size on classification accuracy. Having already shown that image descriptors obtained using quaternion-valued local features outperform the descriptors obtained using real-valued local features, in the following we only present the results obtained for the quaternion-valued

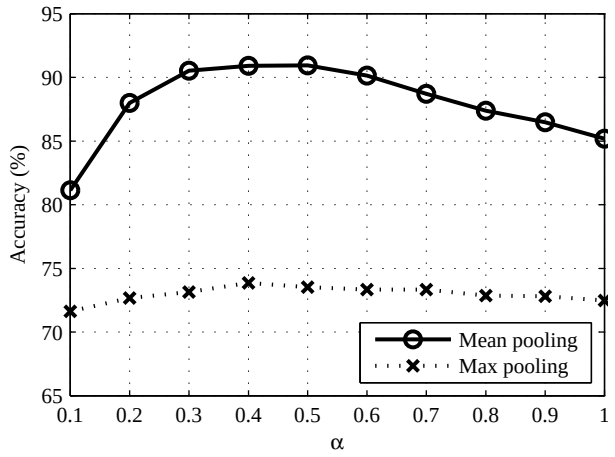


Fig. 8: The impact of power law transform and feature pooling scheme on classification accuracy on UCM dataset.

case. Raw patches are used as local features, patch step is one pixel, and dictionary size is 250, corresponding to 3000-D image descriptors. We varied the patch size from 3×3 pixels to 11×11 pixels in steps of 2 pixels. The obtained results are given in Fig. 7(b). The best results are obtained using patches of size 5×5 pixels. This is important since smaller patch sizes result in lower dimensional local features, which results in lower memory and computational requirements.

We varied the patch step size over 1, 2, 4, and 8 pixels. The patch size is held fixed at 5×5 pixels, raw patches are used as local features, and image descriptors are 3000-D. The results are shown in Fig. 7(c). The classification accuracy is the highest for step size of 1 pixel, and decreases for larger steps. With step of 2 pixels it drops for slightly over 1%, and with step of 4 pixels for 4%. Therefore, in order to obtain high classification accuracy, small step size is needed. This can affect computational complexity, because for encoding each patch Q-OMP algorithm is used for computing sparse representation. Fortunately, as we have already shown, good results can be obtained by choosing $K = 1$ in (6), which reduces Q-OMP algorithm to matrix multiplication and searching for maximum value.

4) Comparison With the State-of-the-Art: We compare our results with the state-of-the-art classification accuracies obtained using three other unsupervised feature learning approaches tested on UC Merced dataset, namely the saliency-guided unsupervised feature learning [25], unsupervised feature coding on local patch manifold (LPM) [8], and quaternion-based feature learning using QPCA and quaternion K-means clustering (QK-means) [9]. We also present the selected results obtained using deep features [39] as well as hand-crafted features, namely multispectral extensions of SIFT-based BoW classifier (MSIFT) [2] and second-order features based on histograms of oriented gradients (HOG) [5].

The results are given in Table II. We report the results obtained using raw quaternion-valued patches and Q-OMP feature encoding using dictionary with 1000 atoms, resulting in 12000-D image descriptors. The results for QK-means are obtained using the code from [9] and with 12000-D descriptors, whereas the results from [8] and [25] are taken from the literature.

TABLE II: Comparison of classification accuracies on the UCM dataset.

Algorithm	**Accuracy (%)**
Learned local features	
Q-OMP	$\mathbf{92.29 \pm 0.71}$
OMP	88.48 ± 1.05
QK-means (χ^2 kernel) [9]	91.38 ± 1.03
QK-means (linear kernel) [9]	83.76 ± 0.46
LPM [8]	90.26 ± 1.51
Saliency [25]	82.72 ± 1.18
Hand-crafted local features	
Dense SIFT [3]	81.67 ± 1.23
MSIFT [2]	90.97 ± 1.81
HOG VLAT [5]	92.3
HOG+RGB VLAT [5]	**94.3**
Deep features	
Caffe [39]	$\mathbf{93.42 \pm 1.00}$
OverFeat [39]	90.91 ± 1.19

We can see that the proposed approach outperforms all the other methods for unsupervised feature learning. Moreover, our classifier is linear SVM, whereas the two closest approaches, QK-means and LPM, use nonlinear χ^2 and histogram intersection kernels, respectively. In comparison with nonlinear SVM, by using linear SVM we reduce computational complexity of classifier training from $O(N^2)$ to $O(N)$, and memory requirements from $O(N^3)$ to $O(N)$, where N is the number of training samples.

When compared to hand-crafted local features, our approach is better than dense SIFT [3], extracted from grayscale images, and MSIFT [2], extracted from color images. The classification accuracy is leveled with HOG-based vectors of locally aggregated tensors (VLAT), and slightly worse than VLAT representation obtained using combined HOG and RGB local features [5]. However, our approach is considerably simpler to compute than elaborate VLAT representation based on second-order statistics of hand-crafted HOG and RGB local features. We believe that our feature learning approach coupled with a more elaborate feature encoding scheme which uses second-order statistics of local quaternion features would further improve the classification performance.

Finally, the classification accuracy obtained with the proposed approach is around 1% worse than that obtained using deep features [39]. However, our feature learning scheme is completely unsupervised and, thus, applicable to scenarios with limited number of training examples.

C. Brazilian Coffee Scenes Dataset

Quaternion image representation (1) has originally been proposed for color, i.e. visible spectrum, images. On the other hand, the Coffee dataset contains color infrared images. We propose to represent color infrared images as pure quaternions

$$Q(x,y) = Q_{nir}(x,y)\, i + Q_r(x,y)\, j + Q_g(x,y)\, k, \quad (16)$$

where $Q_{nir}(x,y)$, $Q_r(x,y)$, and $Q_g(x,y)$ are pixel values in near infrared, red, and green spectral bands, respectively. When this representation is used, a color infrared image is regarded as a single entity and the interrelationships between

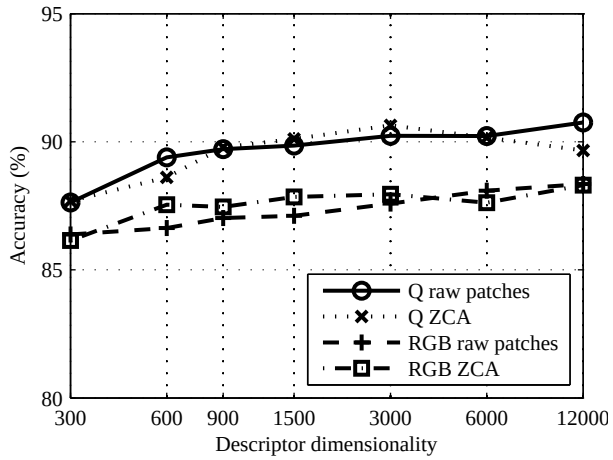


Fig. 9: The impact of image descriptor dimensionality on classification accuracy for Coffee dataset.

the spectral bands are implicitly included in the definitions of image processing operators. This image representation is subsequently used in the same way as for color images.

The values of the encoder hyper-parameters used in the experiments with the Coffee dataset are the same as those used with the UCM dataset, that is patch size is 5×5 pixels with step size of one pixel, QK-SVD is used for dictionary learning and Q-OMP for sparse coding. Subsequently, the local descriptors are computed using (10)-(12), average pooled, power-law transformed with $\alpha = 0.5$, and L_2 normalized.

In Fig. 9 classification accuracies obtained with image descriptors of varying dimensionalities are given. We compare the performances of quaternion and real-valued local features with and without ZCA whitening. Similarly to the results obtained for UCM dataset, quaternion-valued local features outperform real-valued features for 2-3% for all descriptor dimensionalities, thus justifying their use. Furthermore, ZCA whitening does not influence the results in a significant way.

The comparison of the state-of-the-art results with the proposed feature learning algorithms is given in Table III. The Q-OMP feature encoding using 1000 dictionary atoms outperforms both the learning-based and hand-crafted features from [39]. Remarkably, hand-crafted Border-Interior Pixel Classification (BIC) descriptor also outperforms CNN on this task. The reason for this are different spectral characteristics of the images used for training and testing the network, namely visible spectrum and color infrared. Unsupervised feature learning, on the other hand, does not require large labeled training sets and it is able to adapt to the spectral and textural characteristics of the data at hand, yielding more effective visual representation than hand-crafted descriptors. Therefore, quaternion-valued representation can be also used with color infrared images and it is able to leverage the existing interrelationships between near-infrared and visible spectral bands.

VI. Conclusion

In this paper quaternion image representation obtained using raw image patches as features and Q-OMPr for feature encoding is proposed for classification of remote sensing images. In the experiments, it has achieved better classification accuracies than traditional representation computed using real-valued OMP for feature encoding. The reason for success of quaternion image representation is its ability to jointly encode intensity and spectral information by making use of interrelationships between spectral bands in an image.

TABLE III: Comparison of classification accuracies on Coffee dataset.

Algorithm	Accuracy (%)
Q-OMP	$\mathbf{90.75 \pm 0.67}$
OMP	88.35 ± 1.78
Convolutional Neural Network (Caffe) [39]	84.82 ± 0.97
Border-Interior Pixel Classification (BIC) [39]	87.03 ± 1.17

We obtained the highest classification accuracy on UCM dataset, compared to other unsupervised feature learning approaches, and better or competitive results in comparison with hand-crafted features and pre-trained CNN. The obtained results on Coffee dataset are better than those obtained using either hand-crafted features or pre-trained CNN. In this way quaternion feature learning emerged as a viable approach to computing image representation, especially in situations where there are not enough labeled data to train CNN, and pre-trained networks cannot be used because of different characteristics (e.g. spectral) of images.

The proposed image representation uses simple average pooling of encoded features, i.e. only the first-order statistics. However, some of the best results in remote sensing image classification were obtained using second-order features. The extraction and evaluation of second-order features obtained from sparse quaternion image representation is an interesting direction for future research.

The image representation considered in this paper contains a single layer only. On the other hand, hierarchical representations have achieved state-of-the-art classification accuracies in different domains. The proposed algorithm can be easily extended to hierarchical structure for classification. Consequently, in the future work we plan to evaluate quaternion representations at higher levels of hierarchy.

Acknowledgment

The authors would like to thank the anonymous reviewers for their thorough and helpful remarks which greatly improved the quality of this paper.

References

[1] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. ACM Int. Conf. Adv. Geogr. Inf. Syst.*, 2010, pp. 270–279.

[2] A. Avramović and V. Risojević, "Block-based semantic classification of high-resolution multispectral aerial images," *Signal, Image and Video Process.*, pp. 1–10, 2014, to be published.

[3] A. M. Cheriyadat, "Unsupervised feature learning for aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 439–451, 2014.

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

[5] R. Negrel, D. Picard, and P.-H. Gosselin, "Evaluation of second-order visual features for land-use classification," in *Proc. 12th Int. Workshop Content-Based Multimedia Indexing*, 2014, pp. 1–5.
[6] A. Coates, A. Y. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," *J. Mach. Learn. Res. – Proceedings Track*, vol. 15, pp. 215–223, 2011.
[7] L. Bo, X. Ren, and D. Fox, "Hierarchical matching pursuit for image classification: Architecture and fast algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 2115–2123.
[8] F. Hu, G.-S. Xia, Z. Wang, X. Huang, L. Zhang, and H. Sun, "Unsupervised feature learning via spectral clustering of multidimensional patches for remotely sensed scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 8, no. 5, pp. 2015–2030, 2015.
[9] V. Risojević and Z. Babić, "Unsupervised learning of quaternion features for image classification," in *Proc. Int. Conf. Telecommun. in Modern Satellite, Cable and Broadcast. Services*, vol. 1, 2013, pp. 345–348.
[10] A. Vedaldi and A. Zisserman, "Efficient additive kernels via explicit feature maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 480–492, 2012.
[11] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, 1993, pp. 40–44.
[12] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.
[13] L. Shi and B. Funt, "Quaternion color texture segmentation," *Comput. Vis. Image Underst.*, vol. 107, no. 1-2, pp. 88–96, 2007.
[14] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proc. ECCV Workshop Stat. Learn. Comput. Vis.*, 2004.
[15] L. Chen, W. Yang, K. Xu, and T. Xu, "Evaluation of local features for scene classification using VHR satellite images," in *Proc. Joint Urban Remote Sensing Event*, Munich, Germany, 2011, pp. 385–388.
[16] Y. Yang and S. Newsam, "Geographic image retrieval using local invariant features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 818–832, 2013.
[17] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 1794–1801.
[18] G. Sheng, W. Yang, L. Chen, and H. Sun, "Satellite image classification using sparse codes of multiple features," in *Proc. 10th IEEE Int. Conf. Signal Process.*, 2010, pp. 952–955.
[19] D. Dai and W. Yang, "Satellite image classification via two-layer sparse coding with biased image representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, pp. 173–176, 2011.
[20] A. Bosch, A. Zisserman, and X. Muñoz, "Scene classification using a hybrid generative/discriminative approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 712–727, 2008.
[21] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, "X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words," *IEEE Trans. Med. Imag.*, vol. 30, no. 3, pp. 733–746, 2011.
[22] L. Weizman and J. Goldberger, "Urban-area segmentation using visual words," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 3, pp. 388–392, 2009.
[23] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "Pcanet: A simple deep learning baseline for image classification?" *arXiv preprint arXiv:1404.3606*, 2014.
[24] P. Tokarczyk, J. Montoya, and K. Schindler, "An evaluation of feature learning methods for high resolution image classification," in *Proc. ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, 2012.
[25] F. Zhang, B. Du, and L. Zhang, "Saliency-guided unsupervised feature learning for scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2175–2184, 2015.
[26] W. Yang, X. Yin, and G.-S. Xia, "Learning high-level features for satellite image classification with limited labeled samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4472–4482, 2015.
[27] S. Sangwine, "Fourier transforms of colour images using quaternion or hypercomplex, numbers," *Electron. Lett.*, vol. 32, no. 21, pp. 1979–1980, 1996.
[28] C. Jones and A. L. Abbott, "Color face recognition by hypercomplex gabor analysis," in *Proc. 7th Int. Conf. Automatic Face and Gesture Recognition*, 2006, pp. 126–131.
[29] R. Soulard and P. Carré, "Quaternionic wavelets for texture classification," *Pattern Recogn. Lett.*, vol. 32, no. 13, pp. 1669–1678, 2011.
[30] V. Risojević and Z. Babić, "Orientation difference descriptor for aerial image classification," in *Proc. 19th Int. Conf. Systems, Signals and Image Process.*, Vienna, Austria, 2012, pp. 156–159.
[31] N. Le Bihan and S. Sangwine, "Quaternion principal component analysis of color images," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1, 2003, pp. I–809–12 vol.1.
[32] S.-C. Pei, J.-H. Chang, and J.-J. Ding, "Quaternion matrix singular value decomposition and its applications for color image processing," in *Proc. IEEE Int. Conf. Imag. Process.*, vol. 1, 2003, pp. I–805–8 vol.1.
[33] Q. Barthélemy, A. Larue, and J. I. Mars, "Sparse approximations for quaternionic signals," *Adv. Applied Clifford Algebras*, vol. 24, no. 2, pp. 383–402, 2014.
[34] L. Yu, Y. Xu, H. Xu, and H. Zhang, "Quaternion-based sparse representation of color image," in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2013, pp. 1–7.
[35] M. Yu, Y. Xu, and P. Sun, "Single color image super-resolution using quaternion-based sparse representation," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, 2014, pp. 5804–5808.
[36] A. Carmeli, "Quaternion k-svd for color image denoising," Geometric Image Processing Lab, Technion – Israel Institute of Technology, Tech. Rep., 2013.
[37] G. E. Box and D. R. Cox, "An analysis of transformations," *J. Roy. Statist. Soc. Series B (Methodological)*, pp. 211–252, 1964.
[38] F. Perronnin, J. Sánchez, and T. Mensink, "Improving the fisher kernel for large-scale image classification," in *Proc. European Conf. Comput. Vis.*, 2010, pp. 143–156.
[39] O. Penatti, K. Noguiera, and J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2015, pp. 44–51.
[40] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. and Technol.*, vol. 2, pp. 27:1–27:27, 2011.
[41] A. Coates and A. Y. Ng, "Learning feature representations with k-means," in *Neural Networks: Tricks of the Trade*. Springer, 2012, pp. 561–580.
[42] Y.-L. Boureau, F. Bach, Y. LeCun, and J. Ponce, "Learning mid-level features for recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2559–2566.

Vladimir Risojević (S'00–M'14) received the B.Sc. degree in electrical engineering from the Faculty of Electronic Engineering, University of Niš, Serbia, in 1997, and the M.Sc. and Ph.D. degrees in electrical engineering from the Faculty of Electrical Engineering, University of Banja Luka, Bosnia and Herzegovina in 2006 and 2014, respectively.

He is currently an Assistant Professor at the Faculty of Electrical Engineering, University of Banja Luka. His research interests are in the areas of signal processing, machine learning and computer vision.

Zdenka Babić (M'03) received the B.Sc., M.Sc. and Ph.D. degrees in electrical engineering from the Faculty of Electrical Engineering, University of Banja Luka, Bosnia and Herzegovina in 1983, 1990 and 1999, respectively.

She is currently a Full Professor at the Faculty of Electrical Engineering, University of Banja Luka. Her research interests are in the areas of signal processing, image processing, circuits and systems and real time algorithms.