TELSIKS 2013 Serbia, Niš, October 16 - 19, 2013

Unsupervised Learning of Quaternion Features for Image Classification

Vladimir Risojević[1], Zdenka Babić[1]

Abstract – **Unsupervised feature learning is a very popular trend in image classification. Most of the methods for unsupervised feature learning produces filters which operate either on intensity or color information. In this paper we propose a quaternion-based approach for unsupervised feature learning which makes possible joint encoding of the intensity and color information. The image representation is computed using quaternion principal component analysis and k-means clustering. We experimentally show that our approach outperforms the existing approach for unsupervised feature learning from color images, achieving classification accuracy of 91% on a dataset of remote sensing images.**

Keywords – **Remote sensing image classification, unsupervised feature learning, quaternion principal component analysis.**

I. INTRODUCTION

Unsupervised feature learning has been a very popular area of research in the last decade. It is based on the premise that it is possible to obtain a discriminative image representation starting from the raw pixel values. The main inspiration for this line of research has been [1] were it was shown that it is possible to learn Gabor-like receptive fields by applying sparse coding to patches extracted from natural images.

When applied to color images, unsupervised feature learning typically produces filters that essentially operate either on intensity or color information, exclusively. Furthermore, filters that operate on color information are tuned to specific color antagonisms and do not take into account intensity information. In this paper we propose joint encoding of intensity and color information using quaternion framework for working with color images.

The main contribution of this paper is the local quaternionic representation for color images obtained by unsupervised feature learning. The two main steps in our unsupervised learning scheme are the computation of a quaternion principal component analysis (QPCA) basis from a set of color image patches and learning of feature filters using k-means clustering applied to patches projected onto the QPCA basis. By jointly encoding the intensity and color information this representation enables better discrimination of image categories. We experimentally show that our representation achieves the classification accuracy of 91% on a dataset of remote sensing images. Since using QPCA for feature learning opens the possibility for dimensionality reduction we also investigate the impact of the dimensionality of patches on classification accuracy.

Papers using PCA applied to image patches and clustering thus obtained local features using k-means have recently appeared in the literature on object recognition [2], medical [3] and remote sensing image classification [4]. The main difference between these approaches is that in [3] and [4] projection onto the PCA basis also includes feature reduction, while in [2] the authors do not reduce the dimensionality of the patches and perform their whitening instead. Therefore, another contribution of this paper is an evaluation of the impact of dimensionality reduction on classification accuracy.

K-means clustering of local descriptors is common in computer vision as a part of the bag-of-words approach for image classification [5]. However, it is mostly used with SIFT descriptors [6] and not raw image patches. Quaternion principal component analysis (QPCA) has been introduced in [7] and [8]. In [9] QPCA of local image patches and k-means clustering were used for color texture segmentation. Although computationally similar to our approach it is fundamentally different since in the mentioned paper the patches from a single image are clustered as a part of segmentation algorithm, while we use clustering to learn feature filters using image patches from the unlabeled training images. To the best of our knowledge local quaternionic features have not been used for neither unsupervised feature learning nor in the bag-of-words framework.

The rest of this paper is organized as follows. The unsupervised feature learning scheme based on principal component analysis and k-means clustering is described in Section II. Its quaternion extension is discussed in Section III. The experimental results are presented in Section IV.

II. UNSUPERVISED FEATURE LEARNING WITH K-MEANS

In this paper we adopt the computational model for unsupervised feature learning described in [2]. First, we randomly extract patches of size $w \times w$ pixels from unlabeled training images. Images can be multispectral with d channels so the patches are reordered as vectors $\mathbf{x} \in \mathbb{R}^N$, where $N = w \times w \times d$. Each patch is normalized to have zero mean and unit variance which corresponds to local brightness and contrast normalization.

In the next step, PCA basis is computed using the training set of image patches. The covariance matrix of the training patches is

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

[1]Vladimir Risojević and Zdenka Babić are with the Faculty of Electrical Engineering, University of Banja Luka, Patre 5, 78000 Banja Luka, Bosnia and Herzegovina, E-mail: vlado@etfbl.net; zdenka@etfbl.net

Let $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N]$ be the matrix of its eigenvectors, i.e. PCA basis vectors, and $\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ corresponding eigenvalues. Thus obtained PCA basis can be used for whitening of the patches

$$\mathbf{x}_{white} = \mathbf{U}\,\mathrm{diag}\left(\frac{1}{\sqrt{\lambda_i + \varepsilon}}\right)\mathbf{U}^T\mathbf{x}, \tag{2}$$

where diag $(\cdot)$ denotes diagonal matrix and ε is a small regularization constant, or for dimensionality reduction

$$\mathbf{x}_d = \mathbf{U}_d^T\mathbf{x}, \tag{3}$$

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

After whitening or dimensionality reduction the obtained vectors are vector quantized (VQ) using k-means clustering. The codebook for VQ is given by cluster centroids $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_k \in \mathbb{R}^N$, where k is the number of clusters. The feature vector of a patch $\mathbf{x}$ is then computed using

$$\alpha \in \{0,1\}^k, \alpha_i = 1 \text{ iff } i = \arg\min_{1 \le i \le k} \|\mathbf{x} - \mathbf{c}_i\|_2. \tag{4}$$

After the feature mapping has been learned we proceed with feature extraction in convolutional manner. A sliding window of size $w \times w$ pixels with step size s is used to extract patches which are preprocessed and its feature vectors are computed using (4).

The local feature vectors are finally pooled by summing them up in some image regions. The summation region can be the whole image as well as some smaller parts of the image, e.g. image quadrants. After the pooling the image is described by a single descriptor.

III. Quaternion Feature Learning

In the quaternion framework for color image processing a color image is regarded as a single entity and relationships between color channels are implicitly included in the very definitions of image processing operators. An RGB image pixel can be represented [10] as a pure quaternion $q = ri + gj + bk$, where r, g and b are red, green and blue color channels, respectively, and i, j, k are unit quaternions, $i^2 = j^2 = k^2 = -1$. The image patches are now represented using vectors of pure quaternions and we are computing the basis vectors using QPCA algorithm [7], [8].

The covariance matrix (1) now contains products of pure quaternions, which are of the form [9]

$$z = wq = (w_1 i + w_2 j + w_3 k)(q_1 i + q_2 j + q_3 k). \tag{5}$$

The real and imaginary parts of the result are

$$R[z] = -\mathrm{dot}(I[w], I[q]), \tag{6}$$

$$I[z] = \mathrm{cross}(I[w], I[q]), \tag{7}$$

where the operator $R[\cdot]$ returns the scalar part of the quaternion argument and the operator $I[\cdot]$ returns a 3-D vector containing the three imaginary parts of the quaternion argument. The operators $\mathrm{dot}(\cdot,\cdot)$ and $\mathrm{cross}(\cdot,\cdot)$ denote dot and cross products of their arguments, respectively. Therefore, QPCA basis will contain full quaternions.

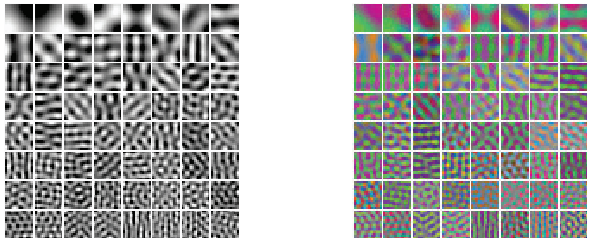

Fig. 1. Examples of QPCA basis filters. (a) Scalar part of the QPCA basis. (b) Vector part of the QPCA basis. Best viewed in color

In Fig. 1a and 1b the examples of QPCA basis filters are shown. We can see that the scalar parts of the filters in Fig. 1a look like a familiar PCA basis obtained by sampling patches from grayscale images [1]. In addition, the imaginary parts of the filters in Fig. 1b show both the color and spatial antagonisms. Thus, the advantage of using QPCA filters for obtaining image representation is that both the intensity and color information will be jointly encoded in the responses of the filters. This leads to better discriminatory ability of the learned features. After image patches have been projected onto the QPCA basis each image patch is represented by a vector of full quaternions. In this step it is possible to reduce the dimensionality of the patches by using only the basis vectors which correspond to the largest eigenvalues, as proposed in [3] and [4]. This approach is taken to the extreme in [9] where the authors claim that, for texture segmentation, it is enough to reduce the dimensionality of the feature vector to one. However, in [2] the dimensionality reduction is not performed and the feature vectors are whitened instead. One of our goals in this paper is to investigate the impact of the dimensionality of the feature vectors on the classification accuracy.

The patches from the input image are sampled using a sliding window 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 essentially the convolution of a full quaternion filter with a pure quaternion image. It was shown in [9] that the convolution of the full quaternion filter with a pure quaternion image is a full quaternion signal whose real part is related to intensity, and imaginary parts are related to red, green and blue color channels.

The codebook for vector quantization is obtained by clustering the feature vectors for patches using k-means algorithm. The extension of k-means to quaternion data is straightforward. Moreover, we note that k-means clustering of quaternion vectors is equivalent to clustering of real vectors whose components are real and imaginary parts of the quaternion elements. In this way we can leverage the existing, efficient implementations of k-means. Prior to clustering we perform z-score normalization of the feature vectors. Finally, the mapped features are pooled as described in Section II.

IV. Experimental Results

We implemented the proposed QPCA-based unsupervised feature learning technique in MATLAB and experimentally compared it with the usual approach for unsupervised feature learning from color images which operates on concatenated vectors of the patch pixel values from red, green and blue color channels. We refer to this approach as RGB PCA.

We test the proposed technique on two image datasets: (i) UC Merced land use dataset[1] [11], [12], and (ii) STL-10 dataset[2] [2].

In the experiments we use the patch size of 5×5 pixels, compute the PCA bases on 10,000 and codebooks on 1 million randomly sampled patches from the training set for the UC Merced land use dataset, and from the unlabeled set for the STL-10 dataset. In the feature extraction phase we use the step size of 1 pixel. Local descriptors are vector quantized using k-means clustering with k=1600.

For classification we use χ^2 kernel mapping [13] and linear support vector machines (SVM). Multiclass classification is performed by training one-versus-all SVMs for all classes and assigning the test image to the class corresponding to the maximum SVM output.

A. Results on the UC Merced Dataset

The UC Merced dataset contains aerial images from 21 land use classes. There are 100 images in each class. All images are color, 256×256 pixels with spatial resolution of 30 cm (1 foot). After computing the local descriptors we pool the local features on the whole image obtaining 1600-D image descriptors.

Following the protocol proposed in [11] we train the classifier on 80 images from each class and test on the rest. We repeat this cycle for five random training/test set splits and report the means and standard deviations of the obtained classification accuracies.

In Table I the classification accuracies for QPCA and RGB PCA-based unsupervised feature learning based on whitening the image patches are given. We can see that the QPCA based whitening outperforms the RGB PCA-based whitening for 1.6%. The reason for this behavior is the ability of the quaternion representation to jointly encode the intensity and color information in the patches thus enabling better discrimination between the image classes. However, this increase in the classification accuracy comes at cost of increased effective dimensionality of the local descriptors. The effective dimensionality is the dimensionality of the equivalent real-valued local descriptor. Since a quaternion is determined with four real numbers the effective dimensionality of quaternion features is four times the number of dimensions of the quaternion descriptor. Therefore, the effective dimensionality of the QPCA-based descriptors is 100-D, compared to 75-D for RGB PCA-based descriptors.

TABLE I
CLASSIFICATION ACCURACIES ON THE UC MERCED DATASET. MEAN CLASSIFICATION ACCURACIES AND STANDARD DEVIATIONS ARE REPORTED.

Method	Accuracy (%)
QPCA	**91.52 ± 0.90**
RGB PCA	89.90 ± 0.80

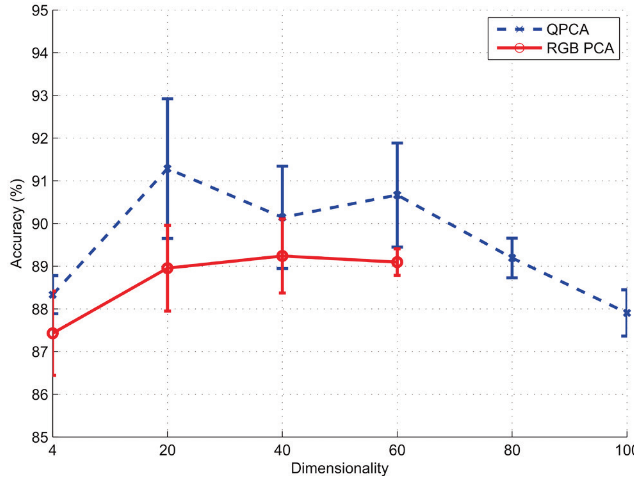

Fig. 2. Classification performance versus the local descriptor dimensionality for the UC Merced datasest. The bars on the plot represent one standard deviation. Since the maximal dimensionality of the RGB PCA-based descriptor is 75 we tested it up to 60-D

Instead of whitening, we can use PCA to reduce the dimensionality of the descriptors. The impact of the number of principal components on the classification accuracy can be observed in Fig. 2. In order to compare the results obtained using QPCA and RGB PCA, the effective dimensionalities of the local features prior to VQ are given on the abscissa of the plot. QPCA-based descriptors always outperform the RGB PCA-based ones, even for very small dimensionalities. It is remarkable that by retaining only one quaternion principal component the classification accuracy of 88% is obtained, and with five quaternion principal components the classification accuracy is 91%, approximately the same as when whitening is used. These results emphasise the importance of jointly coding the local intensity and color information in remote sensing image classification. For larger dimensionalities the accuracy slightly drops due to the noise contained in the dimensions corresponding to the smallest eigenvalues.

B. Results on the STL-10 Dataset

The images in the STL-10 dataset are manually classified in 10 object classes. There are 500 training images and 800 test images per class, as well as 100,000 unlabeled images for unsupervised learning. These images have been sampled from a similar but broader distribution than the labeled images. All images are color, 96×96 pixels. In this case we divide an

[1] http://vision.ucmerced.edu/datasets/landuse.html
2 http://www.stanford.edu/_acoates/stl10/

image into quadrants and pool the local features in each quadrant. The resulting descriptor is a $4 \times 1600 = 6400 - \mathrm{D}$ vector.

TABLE II
CLASSIFICATION ACCURACIES ON THE STL-10 DATASET. MEAN CLASSIFICATION ACCURACIES AND STANDARD DEVIATIONS ARE REPORTED.

Method	Accuracy (%)
QPCA	**60.77 ± 0.55**
RGB PCA	60.75 ± 0.52

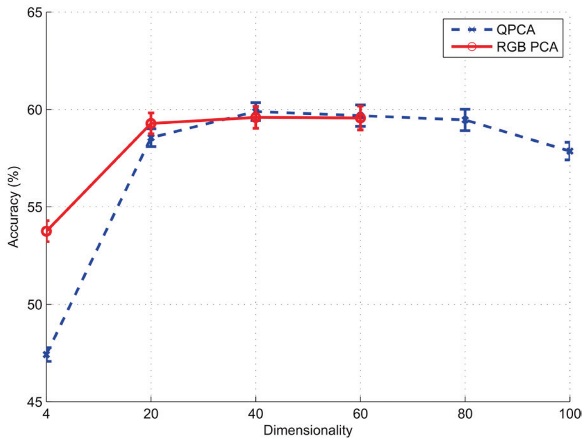

Fig. 3. Classification performance versus the local descriptor dimensionality for the STL-10 dataset. The bars on the plot represent one standard deviation. Since the maximal dimensionality of the RGB PCA-based descriptor is 75, we tested it up to 60-D

The training images in the STL-10 dataset are organized in 10 predefined folds with 100 images per class. In all the experiments with this dataset we trained the classifier using the images from one of these 10 folds and tested on the whole test set. We average the obtained classification accuracies and report means and standard deviations.

In Table II the classification accuracies for QPCA and RGB PCA-based unsupervised feature learning schemes using whitening are shown. We can see that the obtained performances are basically the same. The STL-10 dataset is oriented towards object recognition and local shape information is crucial for discriminating the classes of objects. Color information contained in QPCA-based descriptors does not help in improving the classification accuracy.

In Fig. 3 the dependence of the classification accuracy on the effective dimensionality of the local descriptors is shown. The observations we made above for feature learning using whitening also hold here. The local shape information has more impact on classification accuracy than color information. It is most strikingly seen when only one quaternion principal component is used to represent the local image patch. In that case the effective dimensionality of the descriptor is four. RGB PCA based descriptor of the same dimensionality is obtained using four filters operating on the intensity component of the image. These four filters encode more local shape information than one quaternion-valued which results in better classification accuracy.

V. CONCLUSION

In this paper we proposed a method for unsupervised learning of quaternion image features. Quaternion features enable joint encoding of local intensity and color information thus improving the classification accuracy in cases when color information is an important cue for image classification. We also showed that it is possible to use PCA to reduce the dimensionality of the local image descriptors without loss of the classification accuracy.

ACKNOWLEDGMENT

This work was supported in part by the Ministry of the Science and Technology of the Republic of Srpska under contract 06/0-020/961-220/11.

REFERENCES

[1] 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.
[2] 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.
[3] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, “Xray 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.
[4] L. Weizman and J. Goldberger, “Urban-area segmentation using visual words,” *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 3, pp. 388–392, 2009.
[5] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual categorization with bags of keypoints,” in *ECCV Workshop on statistical learning in computer vision,* 2004.
[6] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, November 2004.
[7] 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.
[8] 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.
[9] L. Shi and B. Funt, “Quaternion color texture segmentation,” *Comput. Vis. Image Underst.*, vol. 107, no. 1-2, pp. 88–96, July 2007.
[10] S. Sangwine, “Fourier transforms of colour images using quaternion or hypercomplex, numbers,” *Electron. Lett.*, vol. 32, no. 21, pp. 1979–1980, 1996.
[11] Y. Yang and S. Newsam, “Bag-of-visual-words and spatial extensions for land-use classification,” in *Proc. ACM SIGSPATIAL GIS*, 2010, pp. 270–279.
[12] ——, “Geographic image retrieval using local invariant features,” *IEEE Trans. Geosci. Remote Sens.,* vol. 51, no. 2, pp. 818–832, 2013.
[13] 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.