

ORIENTATION DIFFERENCE DESCRIPTOR FOR AERIAL IMAGE CLASSIFICATION

Vladimir Risojević and Zdenka Babić

Faculty of Electrical Engineering, University of Banja Luka
Patre 5, 78000 Banja Luka
Bosnia and Herzegovina
{vlado, zdenka}@etfbl.net

ABSTRACT

Texture plays a fundamental role in remote sensing image analysis and texture descriptors based on Gabor filter banks provide a reasonable baseline for aerial image classification. It is known that Gabor wavelet coefficients in different sub-bands are correlated and cross-correlations between Gabor wavelet coefficients at different scales have already been used for image classification.

In this paper we propose using cross-correlations between Gabor wavelet coefficients at different orientations as a descriptor for aerial image classification. We extend our descriptor to color images using new quaternion representation based on symplectic decomposition. On currently the largest publicly available dataset of aerial images we show that the proposed descriptors obtain 85% correct classification rate thus improving state-of-the-art considerably.

Index Terms— Aerial image classification, Texture analysis, Quaternions, Gabor filter banks

1. INTRODUCTION

In general purpose image classification, approaches based on local features such as Scale Invariant Feature Transform (SIFT) are prevalent today. This is understandable since local features have some desirable properties that allow, to a certain extent, for compensation of adversary effects, such as illumination and viewpoint variations, occlusion, and intra-class variability. Local features also give good results in texture classification.

However, in specialized domains some other descriptors may show better results. For example, orthophoto aerial images contain no viewpoint variations, and in land cover/land use analysis occlusions are not very significant. Nevertheless, rotations and translations are present in these images.

Texture plays a fundamental role in remote sensing image analysis. In a recent study [1] it was shown that Gabor texture descriptors yield reasonable performance on large datasets of texture images in the absence of affine and non-affine transformations. Motivated by this result we decided to further

investigate the use of Gabor filters for aerial image classification.

The fact that magnitudes of wavelet image coefficients are correlated was exploited in texture synthesis [2] and retrieval [3]. Gabor wavelets are not orthogonal and there are cross-correlations even among raw coefficients. This was used in [4] where it was shown that center-surround operation in human vision is equivalent to computing correlation coefficients between Gabor wavelet coefficients at different scales and at the same orientation.

However, to the best of our knowledge, cross-correlations between Gabor wavelet coefficients at different orientations and at the same scale have still not been used as image descriptors. The main contributions of this paper are the orientation difference descriptor obtained from the responses of Gabor filters at different orientations and at the same scale, given in Section 2, and its extension to color images obtained using quaternion framework for color image representation, Section 3. We test the proposed descriptors on a challenging 21-class dataset of aerial images and demonstrate that our results considerably improve state-of-the-art compared to both local and global image descriptors, Section 4.

2. ORIENTATION DIFFERENCE DESCRIPTOR

The starting point for the construction of the descriptors in this paper is a Gabor filter bank at S scales and K orientations. The impulse responses of the filters are scaled and rotated versions of the Gabor function.

Suppose we have a multispectral image $I(x, y)$, $(x, y) \in \Phi$, where Φ is the set of image pixels, and let $I_i(x, y)$ is its i -th spectral band. The output of the Gabor filter with impulse response $g_{mn}(x, y)$ at scale $m = 1, \dots, S$ and orientation $n = 1, \dots, K$, is given by the convolution

$$W_{imn}(x, y) = I_i(x, y) * g_{mn}(x, y). \quad (1)$$

In [5] Gabor texture descriptor which consists of means (energies) and standard deviations of the filter responses was proposed. This descriptor proved as a reasonable baseline in aerial image classification [6, 7, 8].

One of the reasons for the popularity of Gabor-based descriptors is their biological plausibility. It is well known that receptive fields of the cells in the retina of the eye have center-surround organization. A cell excited by the incident light in the small central area of its receptive field will be inhibited by the light in the larger region surrounding the center. Because the photoreceptors in the central and surrounding areas can be of the same or different classes, the cells can show spatial and/or chromatic antagonism. Center-surround organization of the receptive fields has been modelled by differences of responses of multiscale filters, and thus obtained color opponent features have been used for texture description [4], visual attention modelling [9], target detection in satellite images [10], and so on.

Let $W_{imn}(x, y)$ and $W_{jm'n}(x, y)$ be the responses of Gabor filters (1) for spectral bands i and j , scales m and m' , and orientation n , respectively. The opponent features $\psi_{ijmm'n}$ are defined as the energies of the differences of normalized filter responses, where normalization is performed using the energies of the filter responses. It was shown in [4] that the opponent features for real-valued Gabor filters can be expressed as $\psi_{ijmm'n} = 2 - 2r_{ijmm'n}$, where $r_{ijmm'n}$ is the correlation coefficient of the filter responses $W_{imn}(x, y)$ and $W_{jm'n}(x, y)$. When complex-valued Gabor filters are used, this relation does not hold any more, but the opponent features can still be regarded as a measure of cross-correlation between the filter responses at different scales and/or for different spectral bands.

Besides cross-correlations between the responses of Gabor filters at different scales, the responses of the filters at the same scale but different orientations are also correlated. Analogously to opponent features, we propose using the energies of the differences of the normalized filter responses at the same scale m and different orientations, n and n' , as an image descriptor. We first compute the differences of the normalized filter responses

$$dW_{ijmnn'}(x, y) = \left| \frac{W_{imn}(x, y)}{\mu_{imn}} - \frac{W_{jm'n}(x, y)}{\mu_{jm'n}} \right|, \quad (2)$$

where μ_{imn} are the total energies of the filter responses for spectral band i , scale m , and orientation n

$$\mu_{imn} = \iint_{\Phi} |W_{imn}(x, y)| dx dy. \quad (3)$$

The orientation difference descriptor consists of the means (energies) and standard deviations of differences (2)

$$\rho_{ijmnn'} = \iint_{\Phi} dW_{ijmnn'}(x, y) dx dy, \quad (4)$$

$$\nu_{ijmnn'} = \sqrt{\iint_{\Psi} |dW_{ijmnn'}(x, y) - \rho_{ijmnn'}|^2 dx dy}. \quad (5)$$

Thus obtained descriptor is $B^2SK(K-1)$ -dimensional, where B is the number of spectral bands.

3. QUATERNION ORIENTATION DIFFERENCE DESCRIPTOR

In this paper we design a color image descriptor using quaternion framework for representation of color images. An RGB image can be represented as a pure quaternion [11]

$$I(x, y) = r(x, y)i + g(x, y)j + b(x, y)k, \quad (6)$$

where i, j, k are unit quaternions, $i^2 = j^2 = k^2 = -1$. In our representation we change the basis of the vector part of (6) from (i, j, k) , to (μ_1, μ_2, μ_3) , where $\mu_1^2 = \mu_2^2 = \mu_3^2 = -1$. The image is then

$$I(x, y) = q_1(x, y)\mu_1 + q_2(x, y)\mu_2 + q_3(x, y)\mu_3. \quad (7)$$

and its symplectic decomposition [11] is

$$I(x, y) = I^{(1)}(x, y) + I^{(2)}(x, y)\mu_2. \quad (8)$$

In our case

$$\mu_1 = \frac{i + j + k}{\sqrt{3}}, \quad (9)$$

$$\mu_2 = \frac{i - j}{\sqrt{2}}, \quad (10)$$

$$\mu_3 = \frac{i + j - 2k}{\sqrt{6}}, \quad (11)$$

so μ_1 corresponds to the gray line in RGB color-space. Hence, the symplectic part $I^{(1)}(x, y)$ in (8) corresponds to luminance signal, and the perplex part $I^{(2)}(x, y)$ corresponds to chrominance signal. This representation is related to the opponent color-space given by

$$q_2 = o_1 = \frac{r - g}{\sqrt{2}} \quad (12)$$

$$q_3 = o_2 = \frac{r + g - 2b}{\sqrt{6}} \quad (13)$$

$$q_1 = o_3 = \frac{r + g + b}{\sqrt{3}}. \quad (14)$$

The image $I(x, y)$ is filtered using quaternion Gabor filter bank obtained by scalings and rotations of the quaternion Gabor function given as

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left[-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi\mu_1\Omega x \right]. \quad (15)$$

The responses of the Gabor filters $W_{mn}(x, y)$ are full quaternions with symplectic decomposition

$$W_{mn}(x, y) = W_{mn}^{(1)}(x, y) + W_{mn}^{(2)}(x, y)\mu_2. \quad (16)$$

The simplex part of the response corresponds to the response of the Gabor filter to luminance signal, and perplex part corresponds to its response to chrominance signal. Since these signals can be regarded as complex-valued we compute the quantities given by (4) and (5) for both parts of the response, dropping spectral band indices, and obtain quaternion orientation difference descriptor. These quantities are computed for all scales and orientations with $n \neq n'$, resulting in a $2SK(K-1)$ -dimensional descriptor.

For computing the dissimilarity between images we adopt distance metric based on the weighted L_1 -norm

$$d(\mathcal{F}_a, \mathcal{F}_b) = \sum_{j=1}^{2SK(K-1)} \left| \frac{f_{a,j} - f_{b,j}}{\sigma(f_j)} \right|, \quad (17)$$

where image descriptors are $\mathcal{F}_a = [f_{a,1}, \dots, f_{a,2SK(K-1)}]^T$ and $\mathcal{F}_b = [f_{b,1}, \dots, f_{b,2SK(K-1)}]^T$, and $\sigma(f_j)$ are the standard deviations of the respective features over the training set.

4. EXPERIMENTAL RESULTS

For classification, we use support vector machines (SVM) with a variant of radial basis function kernel, which uses distance metric (17) instead of Euclidean metric [8]

$$K(\mathcal{F}_a, \mathcal{F}_b) = \exp[-d(\mathcal{F}_a, \mathcal{F}_b)]. \quad (18)$$

Multi-class classification is obtained by training one-vs-all SVMs for all classes and classifying a test sample to the class which corresponds to the maximal SVM response.

We test the proposed descriptor on a dataset of aerial images [7]. All images are multispectral (RGB), 256×256 pixels, with pixel resolution of one foot. They are manually classified into 21 classes, corresponding to various land cover and land use types: agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis courts. Each class contains 100 images. For feature extraction in all experiments we use Gabor filter banks at 4 scales and 6 orientations.

We first test the grayscale orientation difference descriptor. Consequently, all images are first converted to grayscale, using $I(x, y) = 0.299r(x, y) + 0.587g(x, y) + 0.114b(x, y)$, and descriptors are computed using equations (4) and (5). For comparison purposes, we also included the results for Gabor texture descriptors, as well as means and standard deviations of the differences of the filter responses at different scales and same orientation. Since we are considering only grayscale images, there is only one spectral band, i.e. $i = j$ for computation of opponent features as well as in equations (2) - (5).

We averaged the classification rates over 5 different random dataset splits. The results are shown in Fig. 1, where average correct classification rates for training set sizes from

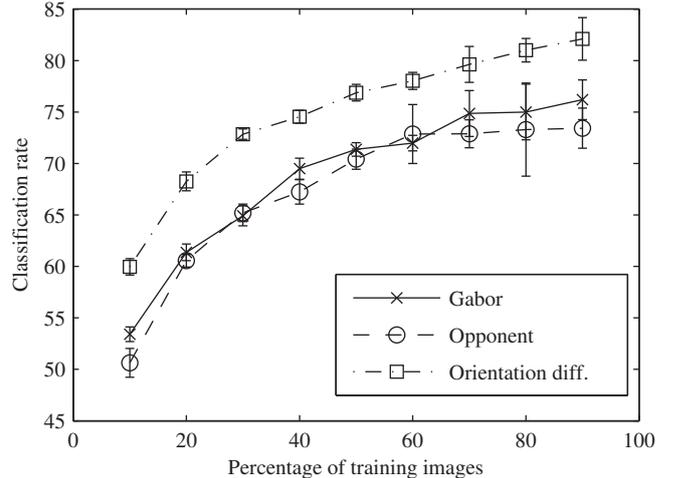


Fig. 1. Classification rates for grayscale descriptors.

10% to 90% of the total dataset size are plotted. Error bars represent the standard deviation of the mean performance over 5 dataset splits. We can see that descriptors based on the differences between filter responses for different orientations outperform all the other descriptors.

For color images we are using quaternion image representation (6) and quaternion Gabor filters (15) at 4 scales and 6 orientations. From the responses of the filters we obtain the descriptor using (4) and (5) for both the simplex and the perplex parts of the response. This descriptor is 240-dimensional.

For comparison we also compute two color texture descriptors. The first is color opponent descriptor, proposed in [4], augmented with standard deviations of differences of filter responses at different scales and the same orientation. This descriptor is 648-dimensional. The second is orientation difference descriptor for color images obtained using (4) and (5). This descriptor is 1080-dimensional.

We also compare our results with the results from [7], where the authors used the classifiers based on bag-of-visual-words, as well as bag-of-visual-words with spatial co-occurrence kernel. We performed 5-fold cross-validation with 80% of the images in the training set, and 20% of the images in the test set. The results are given in Table 1. Confidence intervals for the results from the literature were not available. We can see that grayscale orientation difference descriptors outperform the approaches based on bag-of-visual-words, which are also based on grayscale images. Including color information further improves the performance with quaternion orientation difference and color orientation difference descriptors performing similarly. However, the dimensionality of the quaternion orientation difference descriptor is much lower. Thus, we can see that image descriptors based on cross-correlations between the responses of the Gabor filters at different orientations yield good performance for

Table 1. Classification accuracies.

Descriptor	Accuracy (%)
Grayscale orientation difference	80.67 \pm 0.91
Color opponent features	81.10 \pm 0.55
Color orientation difference	84.86 \pm 1.17
Quaternion orientation difference	85.48 \pm 1.02
Bag-of-words [7]	76.8
BoW + SCK [7]	77.7

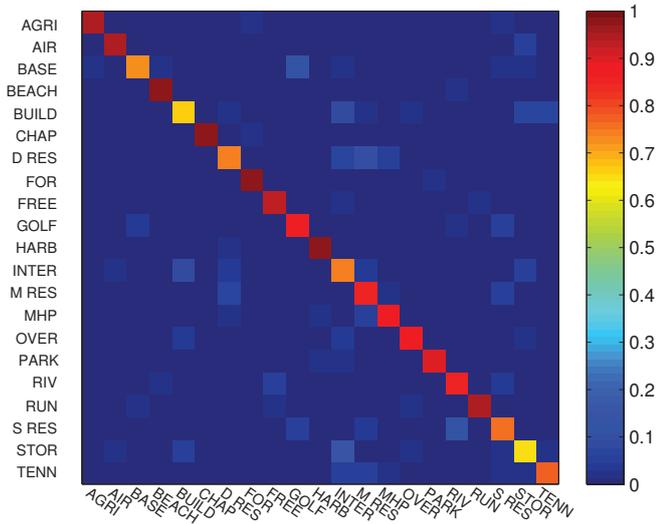


Fig. 2. Confusion matrix for quaternion orientation difference descriptors. (Best viewed in color.)

aerial image classification and including color information considerably improves the state-of-the-art on this dataset.

Confusion matrix for quaternion orientation difference descriptors is shown in Fig. 2. Since our descriptor is focused on texture representation, the obtained results for texture-oriented classes (e.g., beach, chaparral, forest) are very good. However, for object-oriented classes (e.g., buildings, storage tanks, tennis courts) the classification rates are lower.

5. CONCLUSION

In this paper we proposed a descriptor which yields classification rate of 85% on a challenging dataset of aerial images. These results are much better than those obtained using popular bag-of-visual-words descriptor and its spatial extensions. The consequence is that, although bag-of-visual-words descriptor have enjoyed significant popularity in recent years, the “one size fits all” approach may not be the best strategy for specialized domains. Descriptors have to be carefully chosen and the descriptors proposed in this paper may be a good starting point in aerial image classification.

In the future work we will investigate application of di-

mensionality reduction techniques to our descriptor. Since our descriptor is predominantly texture-oriented we plan to investigate combinations of local and global descriptors in order to obtain better results on classes that are more object-oriented.

6. REFERENCES

- [1] U. Kandaswamy, S.A. Schuckers, and D. Adjeroh, “Comparison of texture analysis schemes under non-ideal conditions,” *IEEE Trans. Image Proc.*, vol. 20, no. 8, pp. 2260–2275, August 2011.
- [2] J. Portilla and E. P. Simoncelli, “A parametric texture model based on joint statistics of complex wavelet coefficients,” *Int’l Journal of Computer Vision*, vol. 40, no. 1, pp. 49–71, October 2000.
- [3] J. Zujovic, T.N. Pappas, and D.L. Neuhoff, “Structural similarity metrics for texture analysis and retrieval,” in *Proc. ICIP*, 2009, pp. 2225–2228.
- [4] Amit Jain and Glenn Healey, “A multiscale representation including opponent color features for texture recognition,” *IEEE Trans. on Pat. Anal. and Mach. Intel.*, vol. 7, no. 1, pp. 124–128, January 1998.
- [5] B. S. Manjunath and W.-Y. Ma, “Texture features for browsing and retrieval of image data,” *IEEE Trans. on Pat. Anal. and Mach. Intel.*, vol. 18, no. 8, pp. 837–842, August 1996.
- [6] Y. Yang and S. Newsam, “Comparing SIFT descriptors and Gabor texture features for classification of remote sensed imagery,” in *Proc. ICIP*, October 2008, pp. 1852–1855.
- [7] 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.
- [8] V. Risojević, S. Momić, and Z. Babić, “Gabor descriptors for aerial image classification,” in *Proc. ICANNGA, Part II*, vol. 6594 of *LNCS*, pp. 51–60. Springer Berlin / Heidelberg, 2011.
- [9] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Trans. on Pat. Anal. and Mach. Intel.*, vol. 20, no. 11, pp. 1254–1259, November 1998.
- [10] Z. Li and L. Itti, “Saliency and gist features for target detection in satellite images,” *IEEE Trans. Image Proc.*, vol. 20, no. 7, pp. 2017–2029, July 2011.
- [11] T. A. Ell and S. J. Sangwine, “Hypercomplex fourier transforms of color images,” *IEEE Trans. Image Proc.*, vol. 16, no. 1, pp. 22–35, January 2007.