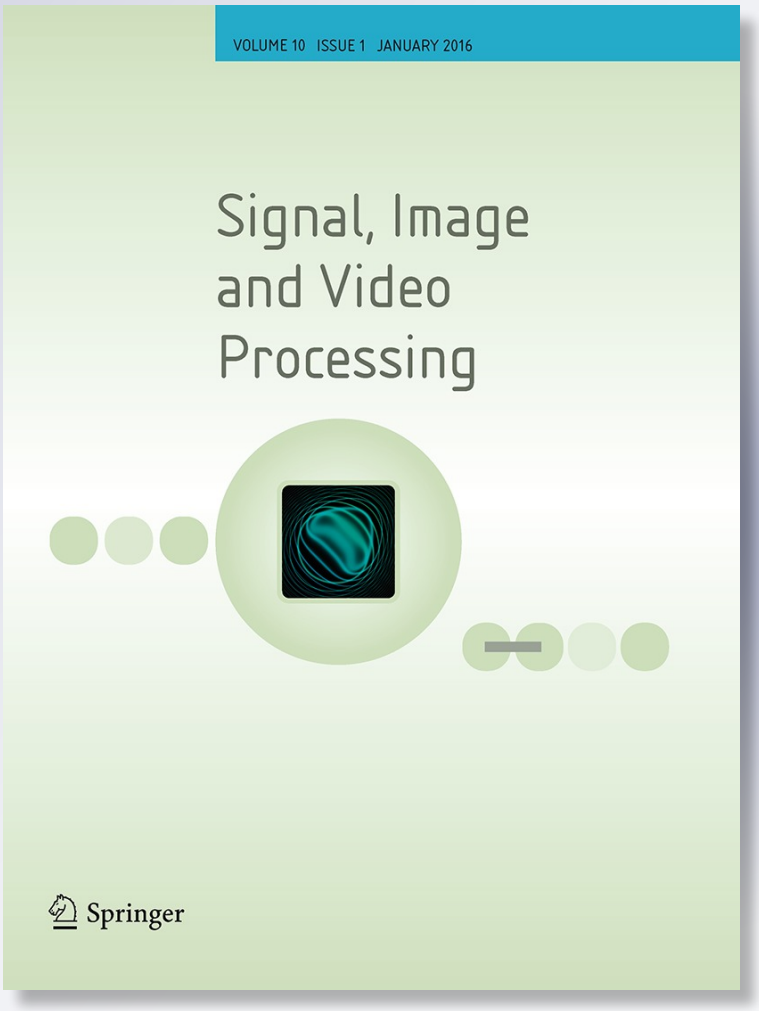
Block-based semantic classification of high-resolution multispectral aerial images

Aleksej Avramović & Vladimir Risojević

Signal, Image and Video Processing

ISSN 1863-1703
Volume 10
Number 1

SIViP (2016) 10:75-84
DOI 10.1007/s11760-014-0704-x



Springer

SIViP (2016) 10:75–84
DOI 10.1007/s11760-014-0704-x

ORIGINAL PAPER

Block-based semantic classification of high-resolution multispectral aerial images

Aleksej Avramović · Vladimir Risojević

Received: 24 January 2014 / Revised: 16 September 2014 / Accepted: 21 September 2014 / Published online: 2 October 2014

Abstract In this paper, we compare different approaches for classification of aerial images based on descriptors computed using visible spectral bands as well as additional information obtained from the near infrared band. We also propose different methods for incorporating dimensionality reduction into descriptor extraction process for both global and local texture descriptors aiming at obtaining low-dimensional descriptors from multispectral images. Furthermore, we examine classification accuracy in cases when small training sets are used. For evaluation purposes, we use an in-house high-resolution aerial image dataset, with images containing visual and near-infrared spectral bands, as well as UC Merced land-use dataset. We achieve the classification rates of over 90 % on in-house dataset. For UC Merced, we obtain classification accuracy of 91 % which is an improvement of about 3 % compared to the state-of-the-art color SIFT descriptors.

Keywords Gist descriptor · SIFT descriptor · Multispectral remote sensing image classification · Land use/land cover

1 Introduction

In recent years, new approaches based on image descriptors computed using both spectral and spatial information available in remote sensing images have achieved state-of-the-art results in classification of high-resolution imagery. Although pixel-based classifiers traditionally use available multispectral data [1], most remote sensing image descriptors are computed using only panchromatic images or, at best, images with visible spectral bands [2–9]. Although this improves classification accuracy, it is noticeable that even in cases where only three spectral bands are used, total dimensionality of descriptor can be very high, thus with additional bands available in multispectral images descriptor dimensionality can increase rapidly.

In this paper, we investigate different schemes to extract texture descriptors from available spectral bands in order to construct low- dimensional image descriptor, while preserving good classification accuracy on the task of block-based aerial image classification. Thus, the proposed schemes use both texture descriptor extraction and dimensionality reduction based on principal component analysis (PCA). We perform experiments using two state-of-the-art image descriptors which combine spectral and spatial information, namely Gist descriptor [10], which is a baseline in scene classification, and scale-invariant feature transform (SIFT) descriptor [11] used in bag-of-words (BoW) framework [12], which has been successful in object recognition. Besides computing, the descriptors using the raw pixel data we also make use of the fact that pixel values in different spectral bands are correlated and decorrelate them using PCA prior to descriptor computation. We show that, in the case of SIFT descriptors, classification accuracies benefit from this decorrelation. Section 2 gives a short review of used descriptors and decorrelation approach. Section 3 presents used data and experiment methodology. Proposed methods are evaluated on two different aerial image datasets. The first one is an in-house dataset obtained by dividing a pair of high-resolution RGB and CIR airborne images into blocks

A. Avramović (✉)
School of Electrical Engineering, University of Belgrade,
Bulevar Kralja Aleksandra 73, 11000 Belgrade, Serbia
e-mail: aleksej@etfbl.net

V. Risojević
Faculty of Electrical Engineering, University of Banja Luka,
Patre 5, 78000 Banja Luka, Bosnia and Herzegovina
e-mail: vlado@etfbl.net

assigned to five different classes. The second one is land-use/land- cover (LULC) dataset from UC Merced which contains images from 21 different classes [2,8]. Since it is often the case that there are only a small number of training block available for some LULC classes, we examined the performances of classifiers trained using small and equal training sets.

The main contributions of this paper are as follows: (1) systematic evaluation of Gist as well as SIFT-based BoW image representations obtained from RGB and NIR spectral bands at the task of aerial image classification, (2) investigation of different possibilities of incorporating dimensionality reduction approaches into descriptor extraction process to circumvent the problem of high descriptor dimensionality obtained for multispectral images, (3) investigation of effects of small training set size on classification accuracy, and (4) introduction of a new LULC dataset with images with RGB and NIR spectral bands.

The significance of texture for aerial image classification, object recognition and scene classification [4,6,7,13–16] raised the issue of efficient way to use texture descriptors in applications where color and multispectral images are available. In [4], different texture representations are compared as representations of multispectral remote sensed imagery in a content-based similarity retrieval scenario. In the remote sensing image analysis, local descriptors have recently been evaluated at the tasks of remote sensing image retrieval, detection of complex geospatial objects [14], and classifying remote sensing images into LULC classes [5,8]. Although information from the NIR spectral band have traditionally been used in remote sensing image analysis, the first efforts to include it into local descriptors came from the general-purpose scene category recognition community. Multi-spectral SIFT (MSIFT) was proposed in [17] and further investigated in [18] where systematic tests of various spectral bands' combinations were performed. Authors in [19] made systematic evaluation from the point of invariance properties of different color descriptors for object and scene recognition tasks. They also reported good performance of local SIFT-based descriptors on the task of object recognition.

2 Image descriptors computed using RGB and NIR spectral bands

2.1 Gist descriptor

Gist descriptor was first proposed in [10] for scene classification. By capturing the *spatial envelope* of the scene, it provides a global image representation which has been shown to be effective in natural scene classification. Its computation starts by filtering the image using a Gabor filter bank which consists of scaled and rotated Gabor filters [3]. The impulse response of a Gabor filter is given by:

$$g(x, y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right) e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+j\Omega x}, \quad (1)$$

where σ_x and σ_y define the bandwidth and Ω is the central frequency of the filter. Function (1) can be considered as Gabor mother wavelet, so its scaled and rotated versions can be seen as Gabor wavelets. The responses of the Gabor filters are then averaged in blocks of a 4×4 non-overlapping grid. Therefore, Gist descriptor consists of means of Gabor filter responses in these 16 blocks.

$$\mathbf{gist} = \left[\mu_{1,1,1} \ \ldots \ \mu_{1,K,1} \ \ldots \ \mu_{S,K,1} \ldots \ \mu_{S,K,16}\right], \quad (2)$$

where $\mu_{i,j,k}$ is the average of the Gabor transform coefficients at scale $i = 1, \ldots, S$, orientation $j = 1, \ldots, K$ and in the block $k = 1, \ldots, 16$. Thus, for each spectral band, we obtain a descriptor whose dimensionality is $16SK$.

Gist descriptor is originally proposed for grayscale images. Taking into account, additional spectral bands in the case of multispectral images is possible by concatenating descriptors obtained for different spectral bands. The overall descriptor dimensionality is $16BSK$, where B number of available spectral bands.

2.2 SIFT descriptor

SIFT descriptor [11] is basically a histogram of gradient orientations in the neighborhood of an interest point. In image classification, it is mainly used in the bag-of-words (BoW) framework. The main idea underlying this framework is to vector quantize the SIFT descriptors and to represent an image using a histogram of codeword occurrences. In this paper, we consider two approaches for computing BoW representation for multispectral images. The first one is *multispectral SIFT (MSIFT)*, proposed in [17]. To obtain MSIFT, we first compute SIFT descriptors for each of the four spectral bands: red, green, blue and NIR and concatenate them. MSIFT descriptors are then vector quantized using k-means algorithm, and the BoW representation is obtained analogously to the grayscale SIFT case. The second approach also starts with computing SIFT descriptors for all spectral bands. Instead of their concatenation, all the original descriptors are vector quantized and BoW representation is obtained. We named this representation *multispectral bag-of-words (MBoW)*. To the best of our knowledge, this is a new approach to adding multispectral information into SIFT-based BoW image representation.

2.3 Decorrelating spectral bands

Besides using the raw image data, in the long-standing tradition of multispectral image analysis in remote sensing [1], we added a preprocessing step and applied PCA to pixel values. If we consider feature extraction from RGB+NIR images, let $\mathbf{m} = [r, g, b, i]^T$, be a 4-D vector of pixel values in red, green, blue, and NIR spectral bands, respectively. By applying PCA, we obtain a new decorrelated color vector:

$$\mathbf{p} = [p_1, p_2, p_3, p_4]^T = \mathbf{P}_{rgb+nir}\mathbf{m}. \quad (3)$$

where $\mathbf{P}_{rgb+nir}$ is decorrelation matrix and $\mathbf{p}$ is transformed vector of pixel values. Decorrelation matrix is computed using 10 % randomly chosen pixels taken from the images in the dataset (to be described in the next section). We normalize the pixel values in the new color vector to the interval $[0, 1]$. Descriptors are then computed using the values from this new color vector. For comparison purposes, we also perform decorrelation of RGB color bands. Although data dependent, decorrelation matrix has similar role as conversion to opponent color space, which, according to [19], can improve classification results. To make the difference between the descriptors calculated from decorrelated data and descriptors calculated from raw image data, in the first case, descriptors are labeled with suffix *PCA*, (eg. *Gist PCA* or *MSIFT PCA*). Furthermore, suffixes *RGB* or *RGB+NIR* are used to distinguish cases were visual spectra is used and when additional data from near-infrared band is used (eg. *Gist PCA RGB* or *MBoW RGB+NIR*).

2.4 Dimensionality reduction

Importance of texture descriptors in the field of aerial and satellite image analysis and classification was de-monstrated in numerous papers. Good discriminative characteristics of texture descriptors often come with the cost of high dimensionality, which increases significantly by putting together data from each spectral band of multispectral images.

Inspired by the earlier work [20], in this paper, we consider two ways for incorporating dimensionality reduction: Applying PCA on the descriptor obtained by concatenating descriptors extracted from each spectral band separately (labeled as **DescriptorConc**, shown in Fig. 1a), and applying PCA on descriptors extracted from each spectral band separately and then concatenating them into a final descriptor (labeled as **DescriptorBand**, shown in Fig. 2b). In the latter case, we define *effective descriptor length* as the total length of the resulting descriptor after the concatenation.

3 Experiment setup

3.1 Used data

In-house dataset The first dataset used in this research is obtained from a pair of RGB and NIR images of the same scene and within the same spatial resolution of 1 m. They have been divided into 850 non-overlapping blocks of size 80×80 pixels, which have been manually classified into five visually distinguished classes, as shown in Fig. 2a.[1] The blocks that could not be reliably classified into one of these classes have been omitted.

UC Merced dataset The second dataset is a dataset from UC Merced which contains high-resolution aerial images of size 256×256 pixels taken from USGS National Map.[2] They have been manually classified into 21 land-use classes. Each class contains 100 images and examples can be seen on Fig. 2b. UC Merced represents a more challenging classification task since it contains visually similar but different land-use classes.

3.2 Methodology

We compare performances of the proposed descriptors in the cases when descriptors are extracted from visual spectra (RGB) as well as extended spectra (RGB+NIR). Moreover, we experiment with the introduction of a preprocessing step which includes pixel decorrelation on both visual spectra (PCA RGB) and extended spectra (PCA RGB+NIR). While conducting the experiments on the in-house dataset for each image, we compute Gist, as well as the two BoW representations. For computing the Gist descriptors, we use Gabor filter bank at 4 scales and 8 orientations which yields 1536-D descriptors for RGB images and 2048-D descriptors for RGB+NIR images. SIFT descriptor is computed for patches of size 8×8 pixels with a step of 4 pixels. SIFT dimensionality for one spectral band is 128-D and MSIFT dimensionality is $3 \times 128 = 384$-D for RGB images and $4 \times 128 = 512$-D for RGB+NIR images. When RGB images are considered, we also compare the proposed descriptors with the best-performing descriptors from [19].

For UC Merced dataset, we compare our approach with the results obtained using best-performing SIFT-based descriptors from [19] as well as with the results given in [2,8,12]. Gist descriptor is computed in the same way as for in-house dataset, while we use patches of size 32×32 pixels with the step of 2 pixels for SIFT-based descriptors.

[1] Available at: dsp.etfbl.net/aerial/rgbnir.zip.

[2] Available at: http://vision.ucmerced.edu/datasets.

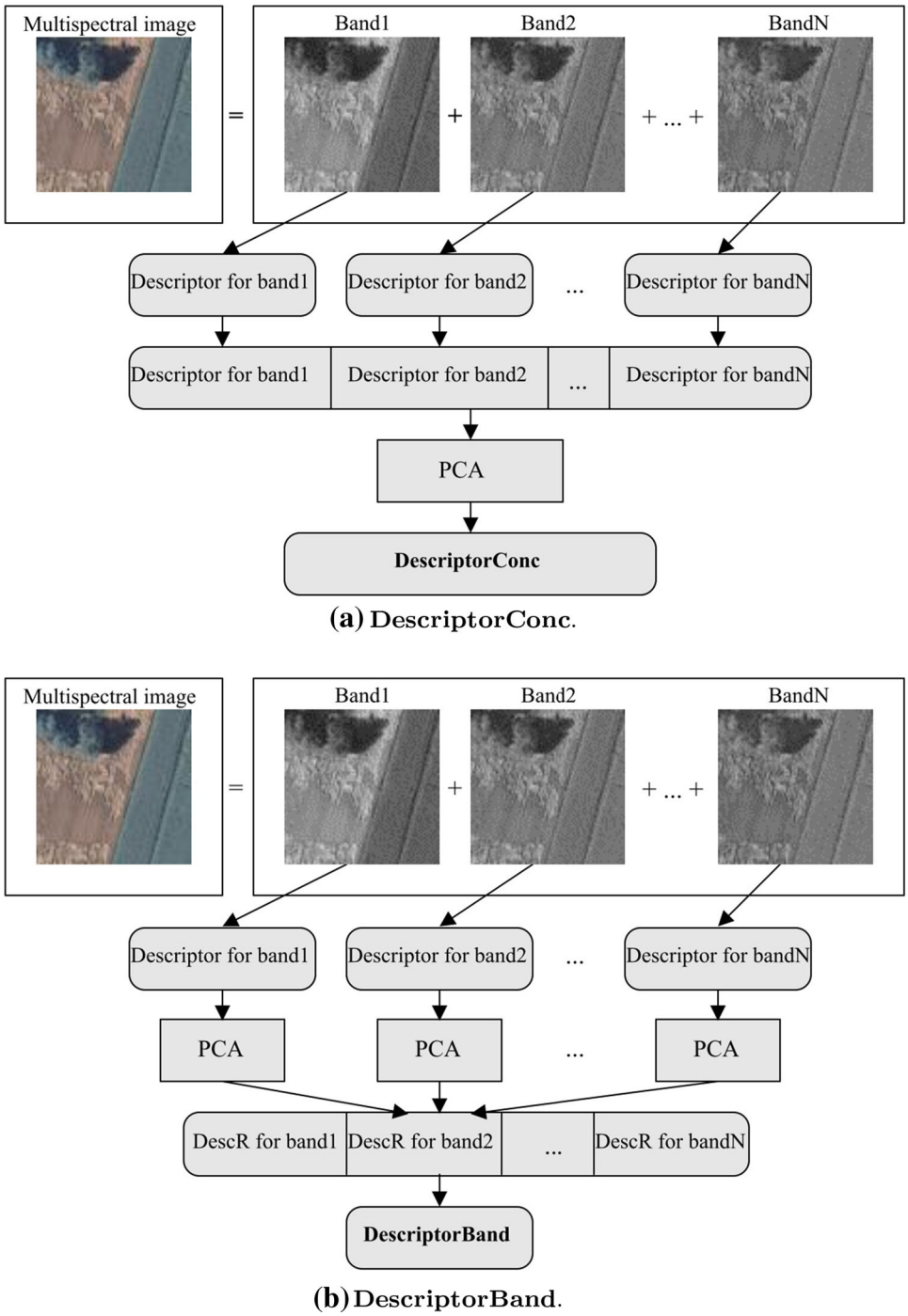
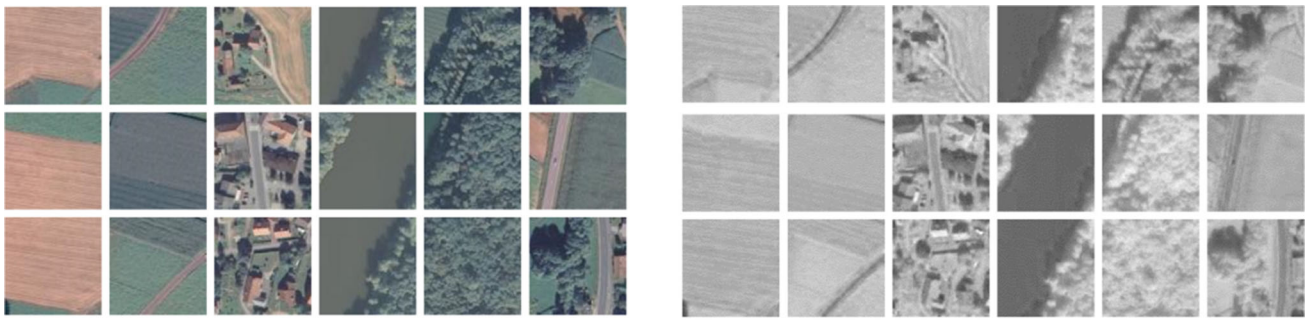


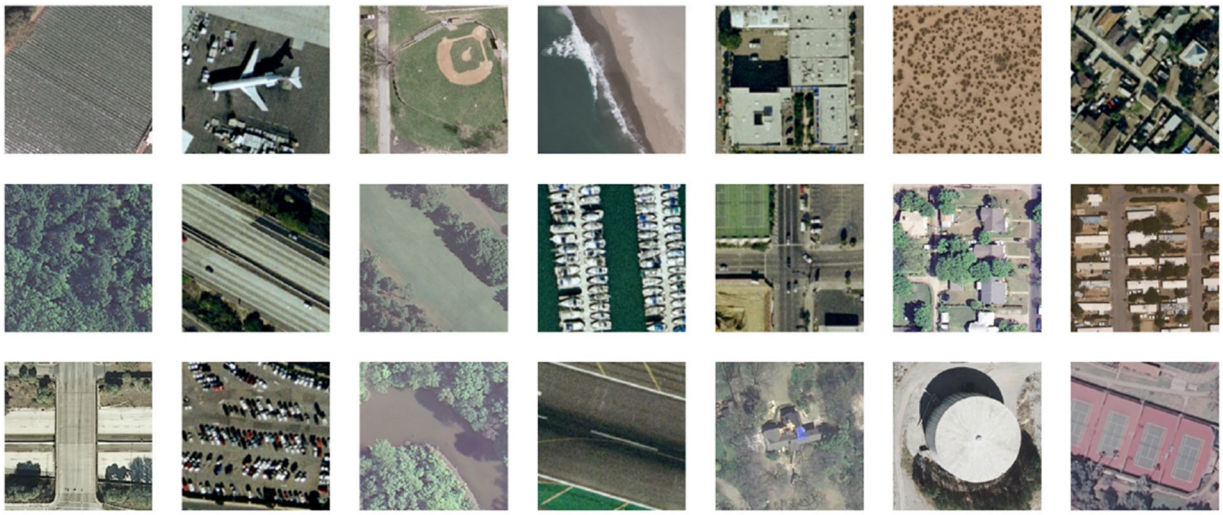
Fig. 1 **a** Scheme for extracting DescriptorConc from multispectral image and **b** scheme for extracting DescriptorBand from multispectral image. *DescR* is used to denote descriptor with reduced dimensionality for a given spectral band

Gist descriptors are computed using the code provided by the authors of [10], SIFT descriptors, and BoW representation using VLFeat library [21], and descriptors described in [19] are computed using the publicly available code.[3] In order to tune up the descriptor extraction for BoW model, we first evaluate the impact of the codebook size on the classifier accuracy. Then we fix the codebook size to the value for which the best results have been obtained and perform further experiments. Multiclass classification is performed in one-versus-all manner. The reported results are obtained by averaging the classification accuracies over 10 different random training/test splits of the dataset. The experiments are performed using MATLAB. For classification, we used support vector machines with radial basis function kernel as implemented in LIBSVM [22].

[3] ColorDescriptor software can be downloaded from http://koen.me/research/colordescriptors/.



(a) Different classes of in-house dataset.



(b) Different classes of UC Merced dataset.

Fig. 2 **a** Examples of blocks by classes (RGB on left and NIR on right). From *left* to *right*: *grayfield*, *greenfield*, houses, river, and woods. **b** Examples of blocks by classes from top to *bottom* and *left to right*: agricultural, airplane, baseballdiamond, beach, buildings, chaparral, denseresidential, forest, freeway, golfcourse, harbor, intersection, mediumresidential, mobilehomepark, overpass, parkinglot, river, runway, sparseresidential, storagetanks, and tenniscourt (color figure online)

Table 1 Mean classification accuracy ± standard deviation (percent) for Gist descriptors on in-house dataset

Gist RGB	Gist PCA RGB	Gist RGB+NIR	Gist PCA RGB+NIR
88.75 ± 2.07	89.53 ± 2.02	90.77 ± 1.55	86.73 ± 1.91

4 Classification results

4.1 In-house dataset

Classification using descriptors with full dimensionality. In this set of experiments, we randomly split each class into the training and test sets of approximately the same size. We compute the image descriptors as described above and train multiclass SVM classifiers. We repeat this process ten times and report the average classification accuracies. The results for Gist descriptor are given in Table 1. Comparing performances when descriptors are extracted from images in the visual spectrum, we can see that pixel decorrelation slightly improves classification accuracy. Furthermore, including the additional information from the near- infrared band improves classification accuracy for about 2 %. However, in this case, pixel decorrelation fails to additionally improve the classification accuracy.

Considering MSIFT descriptor, we first investigate the dependence of the classification accuracy on the size of the codebook. In Fig. 3a, the classification accuracies versus the codebook size for MSIFT descriptors are given. The accuracies are pretty much leveled for the tested range of codebook sizes. The fluctuations are mainly due to the local minima of the k-means algorithm. Adding the data from the NIR spectral band improves the results over the RGB case for 1.5 % for the original, as well as for 2.5 % for PCA transformed spectral bands. PCA transformation of the spectral bands also improves the classification performance over the origi-

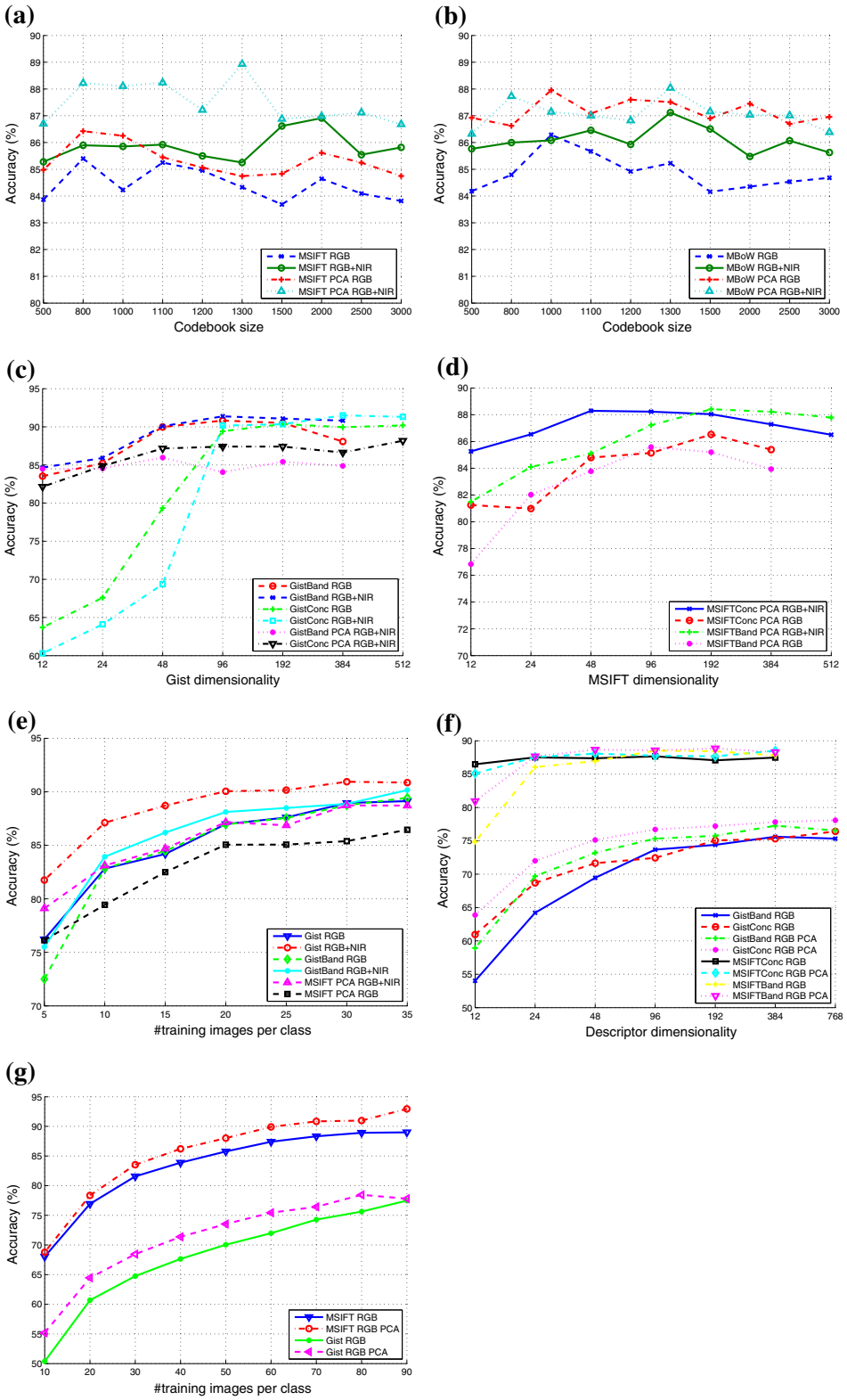


Fig. 3 Plots for various experiments on in-house dataset, from (**a**) to (**e**), and experiments on UC Merced dataset (**f**) and (**g**). **a** Classification performance of MSIFT representation for different codebook sizes. **b** Classification performance of MBoW representation for different codebook sizes. **c** The effect of dimensionality reduction on classification accuracies for Gist descriptor. **d** The effect of dimensionality reduction on classification accuracies for MSIFT. **e** The influence of training set size for best-performing descriptors. **f** The effect of dimensionality reduction on classification accuracy **g** The influence of training set size on classification accuracy

nal spectral bands, both for RGB (1 %) and RGB+NIR (2 %) cases. In Fig. 3b the classification accuracies versus the codebook size in the cases of MBoW representations are given. We can see that, by including the NIR spectral band, the classification accuracy is marginally improved compared to the RGB only case. PCA transformation of the spectral bands again improves the performance over the original bands. Pixel decorrelation in the case when additional NIR band

Table 2 Comparison of classification accuracy (mean ± standard deviation) for in-house dataset

Descriptor	Accuracy
Gist	88.75 ± 2.07
MSIFT	84.33 ± 1.70
MBoW	85.23 ± 2.46
Opponent SIFT	77.93 ± 1.09
cSIFT	85.14 ± 1.75
rgSIFT	82.57 ± 1.89
HSVSIFT	80.88 ± 2.03

Only images in visual spectrum are used

Table 3 Best performance mean classification accuracy ± standard deviation (percent) on in-house dataset

Descriptor	Spectral bands	
	RGB	RGB+NIR
Gist	88.75 ± 2.07	90.77 ± 1.55
GistBand	90.82 ± 1.60 (48)	92.35 ± 1.46 (48)
GistConc	90.36 ± 2.03 (192)	91.31 ± 1.58 (512)
MSIFT	85.40 ± 1.90	86.91 ± 2.43
MSIFT PCA	86.42 ± 1.93	88.93 ± 1.70
MSIFTBand PCA	85.59 ± 2.42 (192)	88.41 ± 2.20 (192)
MSIFTConc PCA	86.53 ± 2.46 (192)	88.30 ± 1.75 (48)
MBoW	86.29 ± 2.11	86.51 ± 1.58
MBoW PCA	87.96 ± 2.40	88.04 ± 1.97

is used almost does not improve the performance, which is different than for the Gist descriptor, Table 1. The reason for this difference can be found in lower dimensionality of SIFT-based descriptors which is less susceptible to noise. Based on the results from the previous experiments, in all further experiments based on BoW, codebook size is set to 1,300.

In order to compare the classification accuracies of Gist and SIFT-based color texture descriptors on the in-house dataset to the other approaches in the literature, we extracted best-performing SIFT-based descriptors described in [19] using patches of size 8 × 8 pixels and step of 4 pixels. This experiment was done using visible spectrum only, since descriptors presented in [19] support only RGB information. Results are summarized in Table 2. We can notice that Gist descriptor outperformed all descriptors based on local texture and color features. Although Gist features are dependent on orientation which can be a drawback when Gist is used for aerial image classification, it seems that classification accuracy on in-house dataset is not significantly affected by this property. One of the reasons may be the fact that blocks from different classes (Fig. 2a) are visually distinct, and there are no significant geometric transformations in the dataset.

Descriptors with reduced dimensionality For this experiment, we use Gist and MSIFT features which showed better classification performances compared to other approaches when extended spectra is used. Dimensionality reduction is incorporated into descriptor extraction process as described in Sect. 2.4. The results for Gist descriptor are given in Fig. 3c. We can see that the two cases when dimensionality reduction is performed on each band separately outperform the other variants. Descriptors extracted from previously decorrelated spectral bands once again failed to achieve noticeable result, especially for low dimensionality. Additional information from the near-infrared band improves GistBand performance for about 1 % compared to the performance based on visible spectra only. It is also interesting to notice good classification accuracy for GistConc RGB+NIR, 91.50 % for dimensionality of 384. GistBand RGB+NIR achieves similar performance with significantly lower dimensionality, namely 91.38 % with 96-D effective length descriptors.

MSIFT descriptor is obtained by concatenating SIFT descriptors for different spectral bands. The obtained results are shown in Fig. 3d for the cases of MSIFT descriptors computed for RGB and RGB+NIR images with PCA transformed spectral bands. For RGB images and low dimensionality (12-D), the results when PCA is applied to the full MSIFT descriptor (MSIFTConc PCA RGB) are better compared to the variant with PCA applied to SIFT descriptors band-wise (MSIFTBand PCA RGB). The reason for this behavior is the extremely low dimensionality of SIFT descriptors computed for individual spectral band in this case (4-D) which is not enough to capture discriminative information. When we add the NIR spectral band to the descriptor, the situation in the low-dimensional case is similar. PCA applied to the full MSIFT descriptor (MSIFTConc PCA RGB+NIR) attains approximately the same value (around 88 %) as without dimensionality reduction (cf. Table 3) for forty-eight-dimensional case. For higher dimensionalities, the accuracies in both cases decrease. The reason for this is the inclusion of dimensions containing small percent of the total descriptor variance which are strongly affected by noise in data.

Size of the training set Selection of the training data is a very important step in remote sensing image classification workflow [1]. Since it is hard to manually annotate a large number of images from each class in order to obtain a representative training set, it is of interest to test the classifier with different sizes of the training set. We varied the number of training images per class from 5 to 35 in steps of 5. If we take only 5 blocks from each of the five classes for training and compare it with total size of original airborne image, it is easy to show that only 2.37 % of the image is used for training, which is significantly smaller set compared to the traditional 50–50 % split. In this experiment, we use only the descriptors with the best performance in the previous experiments Gist,

GistBand and MSIFT PCA. The results are given in Fig. 3e. We can see that, for both descriptors, only 5 training images is needed in the RGB+NIR case to achieve the classification accuracy of almost 80 %. In all cases, the performances are pretty much leveled starting from 20 training images per class.

Summary The best results for all the approaches are summarized in Table 3. First, we report the results for the descriptors with full dimensionality computed on raw spectral bands. In the case of SIFT-based descriptors, we also include the results obtained after decorrelating the pixel values using PCA, as described in Sect. 2.2. Decorrelation of the pixel values in this case improves the classification accuracy which is not the case with Gist descriptors. The classification accuracies for the descriptors after dimensionality reduction are accompanied by the dimensionalities of the resulting descriptors (in parentheses). Reduction of dimensionality improves the performance in the case of GistBand RGB+NIR descriptors for 2 %. In the case of MSIFT-based descriptors, the classification accuracies obtained using descriptors of reduced dimensionality are leveled with the accuracies obtained for the original descriptors. The best accuracy using MBoW representation is around 1 % lower than using MSIFT. For Gist descriptors, the accuracy is over 90 %, which can be preserved even with the dimensionality of 12 per band. The accuracy for BoW approaches is somewhat lower although still close to 90 %. We believe that these descriptors encode complementary information, similarly to the findings of [6] and it is of interest to present the results for both of them.

4.2 UC Merced dataset

To further evaluate descriptor performance, we repeat the experiments done with in-house dataset on UC Merced dataset. In this set of experiments, we train the classifiers using 80 images from each class and test on the rest. For the experiment with descriptors of full dimensionality, we use both Gist: two proposed SIFT approaches and best-performing descriptors from [19]. We also evaluate performances of Gist and MSIFT descriptors in experiments with reduced dimensionality as well as with training sets of different sizes. In order to have fair comparison, we use the same split on training and test images to train classifiers when different descriptors are used.

Classification using descriptors with full dimensionality Classification using descriptors with full dimensionality is done using Gist descriptor and SIFT descriptors with the best performance on the previous dataset. Experiments with SIFT-based descriptors with different codebook sizes are omitted and the codebook size is set to 1300 in order to have fair comparison of different approaches. Classification accuracies are given in Table 4. We can see that local descriptors extracted from RGB images significantly outperform Gist descriptor as well as approaches based on grayscale images reported in the literature. In the case when Gist and MSIFT descriptors are used, we can notice that pixel decorrelation improves classification accuracy by 2 and 3 % for MSIFT and Gist, respectively. On the other hand, MBoW, cSIFT, and rgSIFT do not benefit from pixel decorrelation. As it was described in [19], cSIFT descriptor is extracted from the opponent color space in order to make it scale-invariant with respect to light intensity, thus additional color transformation is not expected to improve classification accuracy. We can notice that pixel decorrelation does not increase performance when MBoW descriptor is considered contrary to MSIFT and Gist cases. MSIFT extracted from decorrelated RGB color space achieves classification accuracy of nearly 91 %, which is the best result reported for UC Merced dataset so far.

Table 4 Comparison of classification accuracy (mean ± standard deviation) for UC Merced dataset with RGB-based descriptors of full dimensionality

Descriptor	Spectral bands	
	RGB	PCA RGB
Gist	74.14 ± 1.93	77.76 ± 2.62
MSIFT	88.92 ± 1.39	90.97 ± 1.81
MBoW	88.60 ± 1.70	88.31 ± 1.38
cSIFT	88.17 ± 1.17	88.76 ± 1.74
rgSIFT	88.24 ± 1.89	87.71 ± 1.33
BoWV [8]	71.86	N/A
SPMK [12]	74.00	N/A
SPCK++ [8]	76.05	N/A
Dense SIFT [2]	81.67 ± 1.23	N/A

The results from the literature were obtained for grayscale images

Descriptors with reduced dimensionality In this experiment, we use two descriptors: two different ways to incorporate dimensionality reduction and two different color spaces (raw RGB and decorrelated RGB). There are eight different descriptor variants in total. Results are given in Fig. 3f. As we can see, MSIFT significantly outperforms Gist descriptor. While classification accuracies obtained with Gist descriptors do not exceed 78 %, classification accuracy in the case when PCA is applied to full MSIFT descriptor (MSIFTConc) can achieve 86 % even for 12-D descriptor. For MSIFTConc pixel, decorrelation slightly improves classification accuracy in most of the cases. The improvement is even larger in the case when PCA is applied to the individual spectral bands (MSIFTBand). In the case when Gist descriptor is used, we can notice that pixel decorrelation improves classification accuracy for about 3–5 %.

Size of the training set Since each class in UC Merced database has 100 images, we start with 10 training images

per class and increase the number of training images in steps of 10 until the number of 90 training images is reached. As in the previous experiment, we use Gist and MSIFT descriptors extracted from raw data as well as from decorrelated pixel values. Results are shown in Fig. 3g. The classification accuracy clearly benefits from more training images, but start to saturate around 60–70 training images. As we can see, MSIFT descriptor gives at least 10 % better classification accuracy compared to Gist descriptor. Also, pixel decorrelation improves classification accuracy for both Gist and MSIFT for about 3–5 %.

Summary Looking at the results, we obtained on UC Merced dataset in the previous experiments, we can give some important concluding remarks. First, we can notice very good classification performance of MSIFT descriptor extracted from previously decorrelated data. In this case, we used MSIFT PCA RGB descriptor of dimensionality 384-D to obtain classification accuracy of nearly 91 %, which represents an improvement for about 3 % compared to the state-of-the-art SIFT-based color descriptors and 10 % compared to the best result obtained using grayscale descriptors. Gist descriptor fails to achieve noticeable classification accuracy which is due to its limited robustness to geometric transformations. We can also notice excellent performance of MSIFTConc descriptor for very low dimensionalities. Considering the influence of the size of training set we can notice regular progression of performance with the increase of training set size for both Gist and MSIFT descriptors. Finally, from the Fig. 3g we can clearly see that pixel decorrelation does improve classification performances for both Gist and MSIFT descriptors.

5 Conclusion

In this paper, we investigated the possibilities for including additional information from near-infrared spectral band into texture descriptors and to examine different ways to incorporate dimensionality reduction techniques into the process of descriptor extraction from multispectral images. We also examined the influence of small training set on classification accuracy as well as pixel decorrelation prior to descriptor extraction. Results of detailed experiments with Gist and SIFT descriptors extracted from images in visual and NIR spectral bands were presented and it was shown that the addition of the multispectral information is beneficial for block-based aerial image classification. Furthermore, we showed that in the case when local descriptor is used (SIFT-based classification), additional improvement is possible by decorrelating pixel values in different spectral bands using PCA. In the case of SIFT-based classification, we concluded that concatenation-based approach, after which the dimensionality reduction is performed, is able to retain good classification performance on both datasets we used. Finally, it is very important to notice that these descriptors are able to achieve good classification accuracies even when small training sets are used.

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

References

1. Campbell, J.B.: Introduction to Remote Sensing. Guilford Press, NY (2006)
2. Cheriyadat, A.M.: Unsupervised feature learning for aerial scene classification. IEEE Trans. Geosci. Remote Sens. **52**(1), 439–451 (2014)
3. Ma, W.Y., Manjunath, B.S.: A texture thesaurus for browsing large aerial photographs. J. Am. Soc. Inf. Sci. **49**(7), 633–648 (1998)
4. Newsam, S.D., Kamath, C.: Retrieval using texture features in high-resolution multispectral satellite imagery. In: Data Mining and Knowledge Discovery: Theory, Tools, and Technology VI, SPIE Proceedings, vol. 5433, pp. 21–32. SPIE (2004)
5. Ozdemir, B., Aksoy, S.: Image classification using subgraph histogram representation. In: Proceedings of 20th ICPR, pp. 1112–1115. Istanbul, Turkey (2010)
6. Risojević, V., Babić, Z.: Fusion of global and local descriptors for remote sensing image classification. IEEE Geosci. Remote Sens. Lett. **10**(4), 836–840 (2013)
7. dos Santos, J.A., Penatti, O.A.B., da Silva Torres, R., Gosselin, P.H., Philipp-Foliguet, S., Falcao, A.X.: Improving texture description in remote sensing image multi-scale classification tasks by using visual words. In: ICPR, pp. 3090–3093. IEEE (2012)
8. Yang, Y., Newsam, S.: Spatial pyramid co-occurrence for image classification. In: Proceedings of ICCV, pp. 1465–1472 (2011)
9. Bayram, U., Can, G., Duzgun, S., Yalabik, N.: Evaluation of textural features for multispectral images. In: Proceedings of SPIE 8180, Image and Signal Processing for Remote Sensing, pp. 81800I–81800I–14 (2011)
10. Oliva, A., Torralba, A.: Modeling the shape of the scene: a holistic representation of the spatial envelope. Int. J. Comput. Vis. **42**(3), 145–175 (2001)
11. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
12. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, vol. 2, pp. 2169–2178 (2006)
13. Filiberto, P., Gema, G., Pedro, G.S., Majid, M., Xianghua, X.: Multi-spectral texture characterisation for remote sensing image segmentation. In: Proceedings of the 4th Iberian Pattern Recognition and Image Analysis Conference, pp. 257–264. Springer, Lecture Notes in Computer Science (2009)
14. Gleason, S., Ferrell, R., Cheriyadat, A., Vatsavai, R., De, S.: Semantic information extraction from multispectral geospatial imagery via a flexible framework. In: Proceedings of IGARSS, pp. 166–169 (2010)
15. Irtaza, A., Jaffar, M.: Categorical image retrieval through genetically optimized support vector machines (gosvm) and hybrid texture features. In: Signal, Image and Video Processing pp. 1–17 (2014)
16. Rajeesh, J., Moni, R., Kumar, S.: Performance analysis of wave atom transform in texture classification. In: Signal, Image and Video Processing, pp. 1–8 (2012)

17. Brown, M., Süsstrunk, S.: Multi-spectral SIFT for scene category recognition. In: Proceedings of CVPR, pp. 177–184 (2011)
18. Salamati, N., Larlus, D., Csurka, G.: Combining visible and near-infrared cues for image categorisation. In: Proceedings of BMVC, pp. 49.1-49.11 (2011)
19. van de Sande, K., Gevers, T., Snoek, C.: Evaluating color descriptors for object and scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. **32**(9), 1582–1596 (2010)
20. Avramović, A., Risojević, V.: Descriptor dimensionality reduction for aerial image classification. In: Proceedings of 18th IWSSIP, pp. 105–108. Sarajevo, Bosnia and Herzegovina (2011)
21. Vedaldi, A., Fulkerson, B.: VLFeat: an open and portable library of computer vision algorithms. http://www.vlfeat.org/ (2008)
22. Chang, C.C., Lin, C.J.: Libsvm: a library for support vector machines. ACM Trans. Intell. Syst. Technol. **2**, 27:1–27:27 (2011). Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm