Descriptor Dimensionality Reduction for Aerial Image Classification

Aleksej Avramović and Vladimir Risojević
Faculty of Electrical Engineering
University of Banja Luka
Bosnia and Herzegovina
Email: aleksej@etfbl.net

***Abstract*—It is often the case in image classification tasks that image descriptors are of high dimensionality. While adding new, independent, features generally improves performance of a classifier, it increases its cost and complexity. In this paper we investigate how descriptor dimensionality reduction techniques, namely principal component analysis and independent component analysis affect classification accuracy. We test their performance for the task of semantic classification of aerial images. We show that, even with much lower dimensional descriptors, classification accuracy is still near 90%.**

***Index Terms*—Image classification, Image texture analysis, Gabor filters, Principal Component Analysis, Independent Component Analysis.**

I. Introduction

Growing amount of aerial imagery requires efficient methods for acquisition, storage, transmission, browsing, retrieval, and analysis of images. Aerial images are used for land use monitoring, urban planning, crops monitoring, weather forecasting, etc. One of the long standing goals of aerial and satellite image analysis is building of thematic maps, i.e. semantic segmentation of images into a predefined set of classes, e.g. urban, field, forest, etc. However, aerial and satellite images are of high resolution and very often multispectral so this task is extremely computationally intensive. This is the reason why some researchers decided to address the problem of aerial and satellite image classification, where a large image is partitioned into fixed size blocks and each of these blocks is classified into one of the predefined semantic classes. Classification can subsequently help in land use analysis, object recognition, image retrieval and so forth.

Parulekar et al. [1] classify satellite images into four semantic categories in order to enable fast and accurate browsing of the image database. Fauquer et al. [2] classify aerial images based on color, texture and structure features, and use linear discriminant analysis for dimensionality reduction. The authors tested their algorithm on a dataset of 1040 aerial images from 8 categories. In a more recent work [3], Ozdemir and Aksoy use bag-of-words model and frequent subgraph mining to construct higher level features for satellite image classification. The algorithm is tested on a dataset of 585 images classified into 8 semantic categories.

Texture is generally considered as an important cue for image classification. There is a wide variety of different texture descriptors and classification algorithms. In this paper we use Gabor texture descriptor [4] and Gist descriptor proposed for scene classification [5]. Both descriptors showed good performance for aerial image classification [6], with classification accuracy near 90%. However, their dimensionality is very high, and it is of interest to investigate whether it could be reduced without affecting the performance of the classifier. We have been inspired by [7] where dimensionality reduction of Gist descriptors for outdoor scene classification was performed using principal component analysis (PCA) and independent component analysis (ICA). However, to the best of our knowledge, there is no such comparative study for aerial image classification.

The main contribution of this paper is investigation how descriptor dimensionality reduction techniques, namely principal component analysis and independent component analysis affect classification accuracy. We test their performance for the task of semantic classification of aerial images. Images are originally represented using high dimensional Gabor and Gist descriptors computed from multispectral images. We consider dimensionality reduction techniques applied to full descriptors, as well as to descriptors computed for individual spectral components, and compare their performance. We show that, even with much lower dimensional descriptors, classification accuracy is still near 90%.

This paper is organized in the following way. In Section II we discuss dimensionality reduction techniques with focus on principal component analysis and independent component analysis. Experimental results are presented in Section III. In Section IV we summarize our results and identify directions for future research.

II. Descriptor Dimensionality Reduction

It is often the case in pattern classification tasks that feature vectors, i.e. descriptors, in the case of image classification, are of high dimensionality. While adding new, independent, features generally improves the performance of a classifier, it increases its cost and complexity. However, complex models have some drawbacks. They require large storage space, both learning and classification become slow due to computational complexity, and overfitting may occur [8]. In order to overcome this problem various data preprocessing and dimensionality reduction techniques have been proposed. In this paper we evaluate two such methods, namely principal component analysis and independent component analysis.

A. Principal Component Analysis

Principal Component Analysis (PCA) is a linear transform that allows dimensionality reduction obtaining minimal sum of squared errors. Let us consider a zero-mean n-dimensional image descriptor $\mathbf{x}$, and a linear combination $y = \mathbf{w}^T\mathbf{x}$, where $\mathbf{w}$ is an n-dimensional weight vector. If $\mathbf{w}$ is chosen in a such way that y has maximal variance, y is called the first principal component of $\mathbf{x}$ [8]. To reduce descriptor dimensionality PCA matrix is calculated based on descriptors from the training set. Training set descriptors are assembled in the training matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k]$ where k is the number of training images. Correlation matrix is calculated as follows:

$$C = E\left\{\mathbf{X}\mathbf{X}^T\right\} \tag{1}$$

Finding principal components of an image descriptor is equal to calculation of eigenvectors of descriptor correlation matrix $\mathbf{C}$. Each eigenvector $\mathbf{e}_i$ is associated with eigenvalue σ_i. Eigenvalue is equal to the variance of the associated eigenvector, so lower eigenvalue means less significant eigenvector. Thus, eigenvectors with small variances often can be discarded without noticeable loss of information. By retaining m most significant eigenvalues, a transformation matrix $\mathbf{P}$ is defined in a way to reduce the dimensionality of descriptors with minimal sum of squared errors introduced. Total amount of energy preserved is equal to the sum of retained eigenvalues. If we denote a set of training descriptors with $\mathbf{X}$, then

$$\mathbf{Y} = \mathbf{P}^T\mathbf{X} \tag{2}$$

is the low-dimensional set of training descriptors. Transformation matrix $\mathbf{P}$ is used to reduce dimensionality of the test set, as well.

B. Independent Component Analysis

After PCA is applied, the obtained low-dimensional image descriptors can be used for classification. However, it is possible to seek directions in the feature space that are most independent from each other, and may reveal more information that could improve classification. This can be done using another linear transform, Independent Component Analysis (ICA) [8]. After PCA transformation matrix is calculated and dimensionality reduction is performed, the obtained low-dimensional descriptors are used to estimate the parameters of ICA. Now we have a set of m-dimensional descriptors $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k]^T$, which can presented as a product of statistically independent components $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_k]$ and a $k \times k$ mixing matrix $\mathbf{A}$:

$$\mathbf{Y} = \mathbf{A}\mathbf{S} \tag{3}$$

ICA cannot determinate neither the order of independent components nor their energies, so it is essential to assume that the training and test sets have the same statistics. The estimation of $\mathbf{A}$ is done under restriction that marginal density function of $\mathbf{Y}$ is nongaussian, because the sum of gaussian variables is also gaussian. It is well-known that the sum of independent components is more gaussian that the original components. Therefore, if we consider a product $z = \mathbf{c}^T\mathbf{x}$, where $\mathbf{c}$ is an m-dimensional vector, z is an independent component if $\mathbf{c}$ is chosen in a way that z has a maximal nongaussianity [9].

The estimation of the mixing matrix $\mathbf{A}$ and independent components $\mathbf{S}$ is a problem of finding the maximum of some nongaussianity measuring function. A natural measure of gaussianity is kurtosis, but there are many less computationally expensive equivalent measures used in practice, such as negentropy, mutual information, maximum likelihood, entropy maximization, etc. [9].

In this paper, maximum entropy method, Infomax, is used to calculate independent components. Bell and Sejnowski in [10] used a maximization of network entropy for blind separation and blind deconvolution, which is proven to be equivalent to maximum likelihood estimation principle for ICA estimation. After matrix $\mathbf{A}$ has been estimated from the training set, both training and test set descriptors are transformed as follows

$$\mathbf{S} = \mathbf{A}^{-1}\mathbf{X} \tag{4}$$

where $\mathbf{X}$ is the descriptor matrix.

III. Experimental Results

For evaluation of the proposed methods we used a 4500×6000 pixel multispectral (RGB) aerial image of the part of Banja Luka, Bosnia and Herzegovina. In this image there is a variety of structures, both man-made, such as buildings, factories, and warehouses, as well as natural, such as fields, trees and rivers. We partitioned this image into 128×128 pixel tiles. In our experiments we used only tiles that could be classified manually to a single category, therefore 606 images were used. We manually classified all images into 6 categories, namely: houses, cemetery, industry, field, river, and trees. Examples of images from each class are shown in Fig. 1. It should be noted that the distribution of images in these categories is highly uneven. In our experiments we used half of the images for training and the other half for testing.

For experiments in this paper we use two descriptors which have shown good performance for aerial image classification [6], namely Gabor texture descriptor [4], and Gist descriptor [5]. Both descriptors are based on filtering the image with a Gabor filter bank. Gabor texture descriptor assumes homogeneous texture and consists of means and standard deviations of magnitudes of Gabor coefficients. On the other hand, Gist takes into account the spatial distribution of magnitudes of coefficients and uses their means on a 4×4 grid. Since our images are multispectral, in both cases we compute descriptors for all 3 spectral bands in an image, and concatenate the obtained vectors.

For both descriptors we tested PCA and ICA for dimensionality reduction. More specifically, we tested three variants:

Fig. 1. Samples of images from all classes. From left to right, column-wise: houses, cemetery, industry, field, river, trees. (Best viewed in color.)

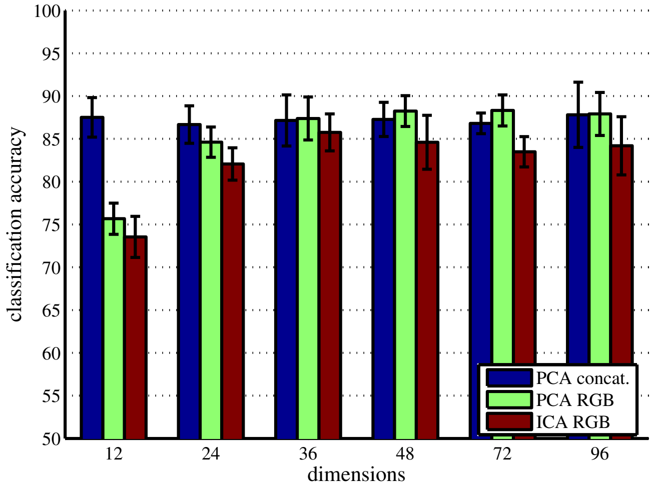

Fig. 2. Classifier performances with Gabor descriptors and various dimensionality reduction techniques (see text).

(i) PCA applied to concatenated vectors, (ii) PCA applied to individual spectral bands, and (iii) ICA applied to individual spectral bands. In the latter two cases the obtained vectors are concatenated after the dimensionality reduction. In order to compare the results, in these cases we used three times smaller number of principal components per spectral band so that the concatenated descriptor has the same dimensionality as in the first case.

We estimated the parameters of both PCA and ICA using the training set, and applied the transforms to both training and test sets. We applied PCA and ICA to both Gabor and Gist descriptors and trained SVM classifiers with radial basis function kernel using thus obtained descriptors. For testing our classifiers we used 10-fold cross validation, each time with different random partition of the dataset, and averaged the results.

A. Gabor Descriptor

We computed Gabor descriptors at 4 scales and 8 orientations for all images from the dataset. Concatenation of descriptors for individual spectral band yields $3 \times 4 \times 8 \times 2 = 192$-dimensional descriptors before dimensionality reduction. In the experiments we tested reductions to 12, 24, 36, 48, 72 and 96 dimensions.

In Fig. 2, results for Gabor descriptors are given. We can observe that PCA applied in a straightforward manner to concatenated descriptors (labeled with PCA concat.) gives the best results, with classification accuracy of 87% obtained with 12 principal components, and it remains nearly constant as descriptor dimensionality increases. Both PCA and ICA applied to individual spectral bands (labeled with PCA RGB and ICA RGB, respectively) have significantly poorer performance for 12 dimensions, which is due to using only 4 principal components per spectral band. However, their performance improves as descriptor dimensionality increases. Nevertheless, in all cases, ICA is outperformed by both PCA-based techniques and its performance does not justify additional computational complexity.

TABLE I
PERFORMANCE OF REDUCED DIMENSIONALITY GABOR DESCRIPTORS COMPARED TO THE PERFORMANCE OF THE ORIGINAL HIGH-DIMENSIONAL DESCRIPTOR.

Descriptor	Dimensionality	Accuracy (%)
Gabor	192	88.4
PCA concat.	96	87.5
PCA RGB	72	88.3
ICA RGB	36	85.8

In Table I the best performances obtained with reduced dimensionality descriptors are compared to the performance obtained with original descriptor (labeled with Gabor) [6]. We can see that dimensionality reduction does not deteriorate the performance of the classifier. Actually, as can be seen from Fig. 2 in some cases it is possible to achieve even more dimensionality reduction at the expense of only slightly reduced classification accuracy.

B. Gist Descriptor

We compute Gist descriptors using a Gabor filter bank with 4 scales and 8 orientations. Since images are multispectral (RGB), we concatenate Gist descriptors computed for individual color components into a $3 \times 4 \times 8 \times 16 = 1536$-dimensional descriptor before dimensionality reduction. In the experiments we tested reductions to 12, 24, 36, 48, 72 and 96 dimensions.

In Fig. 3 results for Gist descriptors are given. In this case, PCA applied to individual spectral bands in most cases outperforms other techniques, achieving 87% classification accuracy for 12-dimensional case. This improvement, compared to the case of Gabor descriptors, is due to the spatial information included in Gist descriptor. The performance of PCA applied to concatenated descriptors improves with the increase of dimensionality. ICA varies in performance compared to PCA applied to concatenated descriptors, but it is consistently poorer

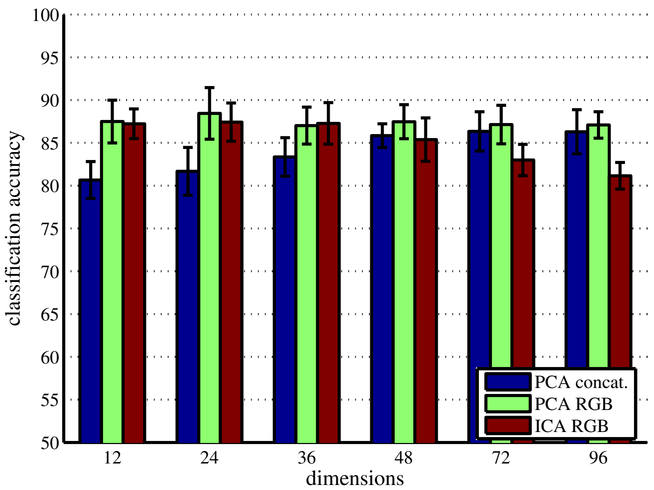

Fig. 3. Classifier performances with Gist descriptors and various dimensionality reduction techniques (see text).

TABLE II
PERFORMANCE OF REDUCED DIMENSIONALITY GIST DESCRIPTORS COMPARED TO THE PERFORMANCE OF THE ORIGINAL HIGH-DIMENSIONAL DESCRIPTOR.

Descriptor	Dimensionality	Accuracy (%)
Gist	1536	89.3
PCA concat.	72	86.3
PCA RGB	24	88.4
ICA RGB	24	87.4

than PCA applied to individual spectral bands. Therefore, its performance again does not justify computational expenses of its application.

In Table II the best performances obtained with reduced dimensionality descriptors are compared to the performance obtained with original descriptor [6]. We can see that dimensionality reduction does not deteriorates the performance of the classifier. Again, from Fig. 3 we can see that, in some cases, it is possible to achieve even more dimensionality reduction at the expense of only slightly reduced classification accuracy.

IV. CONCLUSION

In this paper we evaluated two popular dimensionality reduction techniques, namely principal component analysis and independent component analysis, in the context of aerial image classification. We showed that classifiers trained using descriptors with significantly reduced dimensionality, on this task achieve virtually the same performance levels as classifiers trained using original high dimensional descriptors. The consequence is that it is possible to build effective and efficient aerial image classifiers using state-of-the-art descriptors and standard dimensionality reduction techniques.

In the future we plan to test the described techniques on larger datasets with more semantic categories and to compare them with bag-of-words methods, which are also popular for image classification.

REFERENCES

[1] A. Parulekar, R. Datta, J. Li, and J. Z. Wang, "Large-scale satellite image browsing using automatic semantic categorization and content-based retrieval," in *IEEE International Workshop on Semantic Knowledge in Computer Vision, in conjunction with IEEE International Conference on Computer Vision*, Beijing, China, 2005, pp. 1873–1880.

[2] J. Fauqueur, N. G. Kingsbury, and R. Anderson, "Semantic discriminant mapping for classification and browsing of remote sensing textures and objects," in *Proc. ICIP*, 2005, pp. 846–849.

[3] B. Ozdemir and S. Aksoy, "Image classification using subgraph histogram representation," in *Proc. ICPR*, 2010, pp. 1112 – 1115.

[4] B. S. Manjunath and W.-Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, August 1996.

[5] A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope," *Int. J. Comp. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.

[6] V. Risojević, S. Momić, and Z. Babić, "Gabor descriptors for aerial image classification," in *Proceedings of 10th International Conference on Adaptive and Natural Computing Algorithms, ICANNGA 2011, Part II*, ser. Lecture Notes in Computer Science, A. Dobnikar, U. Lotrič, and B. Šter, Eds. Springer Berlin / Heidelberg, vol. 6594, pp. 51–60.

[7] C. Siagian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 300–312, 2007.

[8] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2001.

[9] A. Hywärinen and E. Oja, "Inedpendent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[10] A. J. Bell and T. J. Sejnowski, "An information maximisation approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.