MAT JOURNALS

Journal of Image Processing and Artificial Intelligence
Volume 2 Issue 1

Classification of Multi-Spectral Images using Wavelet Transformation and Ensemble Projection

M. Ushanandhini[1], S. Rajesh[1], M. Rajakani[2]

[1]Department of IT, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India

[2]Department of CSE, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India

E-mail: ushanandhini.it@gmail.com, srajesh@mepcoeng.ac.in, rajakanijes@mepcoeng.ac.in

Abstract

Image classification techniques play a significant role in the remote sensing imagery. Many of the researchers found some difficulties while doing the analysis of satellite images. During the classification task, many questions have arisen in the minds of the experts and they might face many challenging issues. SSEP (Semi Supervised Ensemble Projection) is a newly adopted method that yields better accuracy even when the satellite image dataset comprises of limited labeled data and great quantity of unlabeled data. Initially, it is common to extract the preliminary features like color, structure, textures for the given image. In this article, we have not only proposed a new Gaussian normal affinity to describe the nearest neighbor by ensemble process in an accurate way which ensures the reliability and diversity, but also we have applied wavelet transformation to texture feature for enhancing the accuracy of classification. Sparse coding technique was mainly meant to overcome the redundancy. The effectiveness of the proposed method is successfully illustrated by using high resolution satellite dataset.

Keywords*: Ensemble projection, image classification, wavelet, semisupervised learning*

INTRODUCTION

Most of the classification techniques use labeled samples for training the classifiers. The classification accuracy generally depends on the amount and the quality of the training samples. It is a time consuming process because, it needs to train all the samples. In order to improve the data given as input to the classification algorithms, other methods like active learning and semisupervised learning, have been proposed, which together utilize labeled and unlabeled samples for training classifiers to develop classification accuracy [1]. This paper investigates the difficulty in semisupervised feature learning for satellite image classification. Based on support vector machines many semisupervised methods have been developed to progress the classification performance. Bruzzone *et al*. proposed a transductive SVM for semisupervised classification of remote sensing images by using labeled and unlabeled images [2]. Munoz-Mari *et al.* introduced one class SVM for the classification of remote sensing data [3]. Chi *et al*. proposed method for classifying the hyperspectral remote sensing data [4]. Marconcini *et al*. presented progressive semisupervised SVM for the classification of hyperspectral images [5].

Graph based methods also fruitfully introduced to the semisupervised classification of remote sensing data [6]. In this method, each sample spreads it label information to its neighbors until a global state is achieved. Camps-Valls *et al.* presented classification of hyperspectral images by taking both unlabeled images and contextual information [7]. Bandos *et al*. presented a classification of hyperspectral images using spatial-contextual information [8]. Gomez-chova *et al.* introduced a Laplacian SVM for image classification; it is based on graph theory and kernel machines [9]. Semisupervised learning method used a small amount of labeled and large amount of unlabeled data; it does not catch enough attention. This article proposes to learn an image representation in semisupervised way. Semisupervised feature learning is flexible and easy to use.

The most prominent one is Bag of feature represents an order less collection of local features for an image, not using any spatial information [10]. To eliminate this problem Lazebnik *et al*. proposed Spatial Pyramid Matching method [11]. Yang *et al.* introduced construction of high level features by exploiting spatial pyramid of sparse codes of local features [12]. Xia *et al.* developed structure and texture features for indexing and scene classification for high resolution satellite images [13]. These methods produce better results and also used in the classification of remote sensing images. The above methods contain the information of sole image alone, not the whole dataset. Therefore, this paper aims to learn a high-level feature representation for each image by exploiting both labeled and unlabeled images. Tsai and Lai introduced a 3-D gray-level co-occurrence matrix, for extending the conventional GLCM to its 3-D version [14]. Bau *et al.* proposed a spatial–spectral feature based on the 3-D Gabor filters to combine the orientation, scale and wavelength dependent properties for remote sensing imagery [15]. Tewfik *et al*. proposed a method to select a wavelet for signal representation for minimizing an upper bound [16]. Villasenor *et al.*, we are going to establish the most important criteria for choice of decomposition filters in wavelet-based texture description [17]. Dai and cool, developed an unsupervised feature learning approach based on ensemble projection [18]. They sample a collection of weak training set by using both local-consistency and exotic-inconsistency assumption. To compute the similarity images are projected onto the sampled prototypes. For final feature representation all similarities are concatenated. By doing this, the method shows potential results. It also has some drawbacks. First, it cannot fully utilize the information provided by the training samples because it uses an unsupervised feature learning method. Second, to find the nearest neighbor, uses K-nearest neighbor (KNN) algorithm. In high dimensional feature space KNN is not a powerful one.

Our algorithm is developed to address these problems. Images are represented by projecting it onto collection of Weak training sets sampled from Gaussian approximation of several feature spaces, known as Semisupervised ensemble Projection (SSEP). First of all, we need to extract the preliminary features from an image dataset, from that form a low-level image descriptions. Then, we use the given training samples as a base of the weak training set; enrich the prototype set by adding closest neighbors of the base images. Thus, a collection of weak training set is developed with Gaussian Normal affinity (GNA) [19]. By concatenating the projected values images are represented. In each category projected values are known as similarities, which capture high-level information and shrink the semantic gap. The performance of SSEP is related to diversity and accuracy of the WT sets. To ensure the accuracy of the WT sets, choose labeled images as seed images for generating the WT sets. To guarantee the diversity of the WT sets, use different feature to compute neighbors of an image and launch randomness.

SSEP contains redundant information, because different WT set may share same images. In order to overcome the redundancy, use sparse coding to reduce the redundancy. This method involves three steps:

1) Learn features using semisupervised algorithm via EP.
2) Propose GNA method to find the nearest neighbor in an accurate way.
3) Use sparse coding to reduce the redundant information, yields better results.

SEMISUPERVISED FEATURE LEARNING

The training data contained both labeled data $D_{la} = \{x_i, y_i\}_{i=1}^{kl}$ and unlabeled data $D_{un} = \{x_j\}_{j=kl+1}^{kl+u}$, where x_i is the feature descriptor of image i and $y_i \in \{1,....,k\}$ is its label. K is the number of categories. l is the number of labeled data in each category and u is the number of unlabeled data. Our method aims to learn a high-level image representation S by exploiting the few labeled data and great quantities of unlabeled ones, which is then fed into different classifiers to obtain final classification results. The procedure of semisupervised feature learning by SSEP is shown in Fig 1. First, a new sampling algorithm based on GNA is proposed to produce T WT sets $P^t = \{(s_i^t, c_i^t)\}_{i=1}^{kp}, t \in \{1,....,T\}$ with great diversity from all data, where $s_i^t \in \{1,.....,kl+u\}$ is the index of the ith chosen image, $c_i^t \in \{1,.....,k\}$ is its label and p ($p > l$) is number of final training images for each class [19]. The sampling algorithm is performed in different feature spaces for great diversity. Therefore, images are represented by multiple feature descriptor, $x_i = \{x_i^1,...,x_i^r\}$ where r is the number of feature spaces (e.g., k = 3, l = 1, p = 3, and r = 3 in Figure 1). Second, each WT set is used to train a discriminative classifier, resulting in an ensemble of classifiers. For a new image, each classifier projects it into a similarity vector. The final image representation S is obtained by concatenating all the similarity vectors.

Semisupervised Sampling Algorithm

To have good performance WT sets need to be accurate and diverse.The algorithm 1 works based on two steps. First, by using all labeled training samples, WT is formed. Second, skeleton is enriched by sampling their nearest neighbors. Each WT set consists of all labeled data and randomly selected subset of their neighbors. In each category, the labeled images are served as seeds and select neighbors using random sampling process. Seed images must be very accurate one. Classification accuracy decreases if we use poor seed images. By concatenating the final data in all categories, final WT is constructed. By sampling in multiple feature spaces and introducing GNA, Algorithm 1 ensures both the accuracy and diversity of the WT sets [19].

Accuracy

The WT set need to be clean, in order to have accurate base learners. This is guaranteed by adding nearest neighbors of labeled samples. And introduce Gaussian Normal affinity method to find neighbors in a better way.

Diversity

Each WT set consists of the labeled training data and the enriched samples. As different WT sets have the same labeled training data, we should sample diverse neighbors to ensure the diversity of the ensemble. Data may have different distributions in different feature spaces. For the same labeled data, their neighbors in different feature spaces can also be different. We can obtain more diverse WT sets in different feature spaces than only in one feature space. Therefore, sampling in multiple feature spaces is designed in Algorithm 1. Algorithm 1 also improves the diversity of WT sets. The algorithm is repeated many times to construct the ensemble. Instead of KNN, we use GNA method to find nearest neighbors in an accurate way. Semisupervised method is accurate.

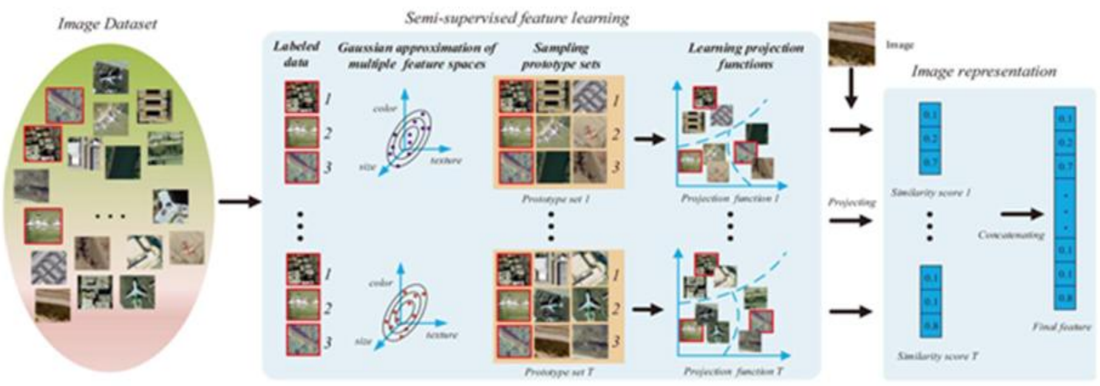

Fig. 1: *Flowchart of SSEP.*

GNA

The key problem in high dimensional feature space is how to find nearest neighbors fast and accurate manner. As a result semisupervised sampling algorithm is proposed. In many applications for finding nearest neighbors Euclidean distance has been used. In high dimensionality conditions, this similarity measures become unreliable.

Algorithm 1: Semi-Supervised Sampling with Gaussian Normal Affinity in t-th Trial

Input : Dataset D = $D_{la} \cup D_{un}$;

Labeled training data $D_{la} = \{x_i, y_i\}_{i=1}^{kl}$

Output: WT set P^t

Find corresponding training image index in D

i.e., $V = \{kl\}$ image index

for $q \leftarrow 1$ to k do

V_q = {indexes of the l labeled images of class q };

For $f \leftarrow 1$ to r do

S = indexes of the ln nearest neighbors of data in V_q in feature space f

$s_{q,f}^t$ = randomly select m indexes from s;

$s_q^t = (s_{q,1}^t, \ldots, s_{q,r}^t) \cup V_q$;

$c_q^t = (q, q, q, \ldots q) \in R^{mr+l}$;

Let $p = mr + l$ then

$s^t = \{s_1^t, \ldots, s_k^t\} \in R^{kp}, c^t = \{c_1^t, \ldots, c_k^t\} \in R^{kp}$;

$P^t = \{(s_i^t, \ldots, c_i^t)\}_{i=1}^{kp}$

A method GNA is introduced to compute image similarity in high dimensional feature space. It is used to find nearest neighbors in an accurate manner. Data distributed in a high-dimensional feature space tend to lie at the periphery, and they are usually separable from the rest of the set. Their manifold can be approximated by a parametric model, a Gaussian, which is fitted by the covariance matrix on all training data. With the covariance matrix, the computation of the normal at a given image is a simple multiplication by the inverse covariance. GNA uses the normal to define the nearest neighbors of the given image. Our image set is D = $D_{la} \cup D_{un}$ and the covariance matrix of D is used to fit the Gaussian

$$\Sigma = \frac{1}{kl+u} \sum_i (x_i - \mu)(x_i - \mu)^P \qquad (1)$$

Where $kl + u$ is the number of images and μ is the mean of their features

$\mu = \frac{1}{kl+u}\sum_i x_i$. Given a labeled image $x \in D_l$ the normal to the Gaussian at x is computed as

$$\omega_x = \Sigma^{-1}(x-\mu) \quad (2)$$

Then, all image features are projected onto the Gaussian normal w_x, and the projected locations are used as the similarities. Therefore, neighbors of x can be easily found. In Algorithm 1, we select the first n images as neighbors of x, which have larger similarity values than the rest. Note that the computation of the simple matrix-vector multiplication is very cheap and Σ and μ can be computed only once.

EP

Information carried by a WT set is quite limited. EP adopts the idea of ensemble learning, which also aims to learn an ensemble of classifiers with accuracy and diversity. After the sampling algorithm, we obtain T WT sets. Over each WT set, a base learner is trained. Thus, an ensemble of base learners is constructed. The feature used in the base learner is the concatenation of multiple typical features. We adopt the concatenation of scale-invariant feature transform (SIFT), combined scattering (CS) and bag of colors (BOC) [20–22]. Our sampling algorithm constructs an ensemble of WT sets with high accuracy and diversity. Therefore, the base learners are also both accurate and diverse. Furthermore, EP introduces a discriminative learning method that is discriminative classifiers as the base learner to further ensure the accuracy. For an input image x, each base learner projects x, resulting in a similarity vector

$$S_i = (S_{i,1}, \ldots S_{i,k}) \quad (3)$$

Where $i \in \{1,\ldots,T\}$. Element $S_{i,c}(c \in \{1,\ldots,k\})$ measures the probability of x belonging to category c using base learner i. The diversity of the base learners is used to handle intraclass variance as different base learners capture different properties of the data. A new image representation $S = [S_1, S_2, \ldots, S_T]$ is obtained by concatenating all the similarity scores for the final classification.

Therefore, the dimension of S is $k \times T$. Because each WT set represents a different aspect of a visual category, these projection values are similarities to each category and by doing so; images are represented with their affinities to a rich set of discovered image attributes for classification, which capture high-level information, thus, shrinking the semantic gap.

Sparse Coding with SSEP

Each WT set consists of both labeled and unlabeled dataset contains redundant information, because WT set share some images. Sparse coding technique is used to reduce redundant information.LLC utilizes the locality constraints to project each descriptor into its local-coordinate system, which suggests that locality is more essential than sparsity. Specifically, the LLC code uses the following criteria

$$\min_{c} = \sum_{i=1}^{N} \| x_i - Bc_i \|^2 + \lambda \| d_i \Theta c_i \|^2, s.t. 1^T c_i = 1, \forall i \quad (4)$$

Where $X = [x_1, x_2, ..., x_N]$ is a set of local descriptors extracted from an image. $C = [c_1, c_2, ...c_N]$ is the set of codes for X, B is the codebook with M entries. ⊙ denotes the element wise multiplication.

Enhancement using Wavelet Approach

A wavelet is a wave-like oscillation with amplitude that begins at zero, increases, and then decreases back to zero. Wavelet is a tool, used to extract information from many different kinds of data. In this paper the multispectral image is decomposed into spectral-spatial sub bands by applying low pass and high pass filters along rows, columns and slices. In this paper, we use continuous wavelet transformation to increase the classification accuracy.

$$C(\tau, s) = \frac{1}{\sqrt{s}} \int_t f(t) \psi^* \left(\frac{t - \tau}{s} \right) .dt \quad (5)$$

Here τ is the translation parameter used to measure the time. s is the scale parameter used to measure the frequency. $\sqrt{s}$ is the normalization constant.

EXPERIMENTAL RESULTS

We tested our algorithm on three different data sets. In the experiments for the 19-class and 21-class data sets, the extraction process of SSEP is based on the whole image, which is actually an object categorization problem.

Data Sets and Experimental Settings

To illustrate the effectiveness of the feature learned by SSEP two experiments were carried out. 1) 19-class satellite scene data set (19-Class):1 19 categories and each of them has 50 images, with a size of 600 × 600 pixels; and 2) 21-class satellite scene data set (21-Class) : 21 categories and it has 100 images for each, with a size of 256 × 256 pixels.

19-Class is composed of 19 classes of scenes, including airport, beach, bridge, commercial area, desert, farmland, football field, forest, industrial area, meadow, mountain, park, parking, pond, port, railway station, residential area, river and viaduct. 21-Class is composed of 21 classes of scenes, including 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. They are both challenging satellite data sets extracted from very large satellite images, as the illumination, appearances of objects and their locations in different images vary significantly. In our experiments we use three feature descriptors are SIFT, CS, and BOC. The Scalable Invariant Feature Transform descriptor (SIFT) is a weighted and interpolated histogram of the gradient orientations and locations in a patch. We compute the bag-of-words representation based on dense SIFT descriptors for each image. Combined Scattering (CS) has been proven highly discriminative for analyzing the textures. It builds invariant, stable, and informative representations through a nonlinear unitary transform, which delocalizes signal information into scattering decomposition paths.

The Bag of Colors (BOC) is a simple yet efficient method that introduces the concept of a laboratory color palette to extract a color signature. These features represent images from three different perspectives, which are structure, texture and color.

Experiment 1

The three feature descriptors SIFT, CS and BOC were first extracted from 19-class and 21-class. For each data set, we randomly selected {1, 3, 5, 8, 10, 15} images per category as the labeled training data. The rest of the images were taken as the unlabeled training data as well as test data. The use of test data as unlabeled training data is common in transductive semi- supervised learning, which is different from the standard setting of supervised learning. Figure 2 shows that the classification performance is improved with the increase of labeled training images. Feature learned by SSEP have the best performance even with fewer labeled images. Finally, feature learned by applying wavelet to the image for extracting texture features improves the accuracy.Table 1 lists the classification results with five labeled training images per class.

The feature learned by SSEP using wavelet performs better than SSEP.The experimental results show that our feature learned by SSEP works well with SVM obtains promising classification results on high-resolution satellite data sets. The

Accuracy values of 19-Class and 21-Class are 74.61% and 70.73%, respectively. Some images of residential area are misclassified into commercial area because images of commercial area consisted of dense houses, vertical and horizontal line. When we apply SSEP using wavelet transformation accuracy values of 19-class and 21-class are 82.56 % and 79.34 % respectively.

Discussions

Classification Results with Semisupervised Classifiers

To illustrate the effectiveness of our feature, two semisupervised classifiers, the Laplacian SVM (LSVM) and meanS3VM were used. We conducted this experiment on 19-Class and 21-Class and used the five random selected images as labeled training data for each class. The rest of the images were used as unlabeled training data as well as test data. Fivefold experiment was performed, and the mean and standard deviation for accuracy were recorded. Table 3 shows that among different feature extraction methods, OF has the least accuracy value, and the feature learned by SSEP using our proposed wavelet approach has higher accuracy, there by outperforming all other existing methods.

Effectiveness of GNA

To illustrate the effectiveness of GNA, we do experiments on 19-class and 21-class data sets of SSEP with KNN and compare

Table 1: *Accuracy of Classification on 19-Class and 21-Class, with Five Labeled Training Data per Class.*

Dataset	Classifier	SSEP	SSEP using wavelet
19 Class	SVM	74.61	82.56
21 Class		70.73	79.34

Table 2: *Accuracy of classification of SSEP with GNA and KNN.*

Dataset	Classifier	With GNA	With KNN
19 Class	SVM	73.46	70.12
21 Class		69.71	66.34

Table 3: *Accuracy of Classification on 19-Class and 21-Class, with Five Labeled Training Data per Class.*

Dataset	Classifier	OF	SSEP	SSEP using Wavelet
19 Class	LSVM	68.21	76.29	84.11
	mean S3VM	70.56	74.43	83.91
21 Class	LSVM	59.35	68.17	76.50
	mean S3VM	61.27	69.53	78.42

it with the performance of SSEP with GNA. We selected five images per class as labeled training data. The result is shown in Table 2, which demonstrates that the GNA is more efficient than KNN.

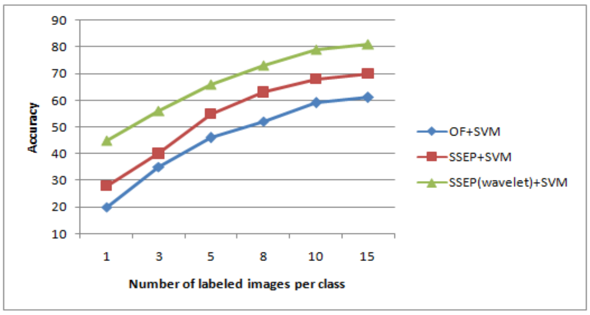

Fig. 2: *Nineteen-Class Satellite Scene Data Set.*

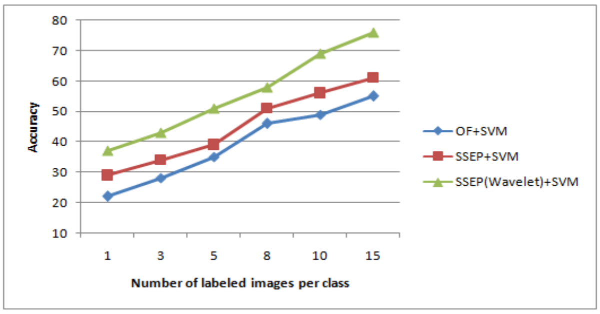

Fig. 3: *Twenty One-Class Satellite Scene Data Set.*

CONCLUSION

A high level image representation is learned for multispectral images using small amount of labeled samples and huge amount of unlabeled samples. To create an ensemble of WT sets, a sampling algorithm is designed and performed in different feature spaces. We used Gaussian Normal Affinity (GNA) to find nearest neighbors in an accurate manner. Results on two challenging datasets illustrates the superiority of our feature learned by SSEP using wavelet transformation and ensemble projection give more accurate results. Our future work will focus on extending this paper towards hyperspectral images.

REFERENCES

1. L. Bruzzone, C. Persello. Recent trends in classification of remote sensing data: Active and semisupervised machine learning paradigms. *In Proc. IEEE IGARSS.* 2010; 3720–3723p.
2. L. Bruzzone, M. Chi, M. Marconcini. A novel transductive SVM for semisupervised classification of remote-sensing images. *IEEE Trans.Geosci. Remote Sens.* 2006; 44(11): 3363–3373p.
3. J. Munoz-Marí, F. Bovolo, L. Gómez-Chova et al. Semisupervised one-class support vector machines for classification of remote sensing data. *IEEE Trans. Geosci. Remote Sens.* 2010; 48(8): 3188–3197p.
4. M. Chi, L. Bruzzone. Semisupervised classification of hyperspectral images by SVMs optimized in the primal. *IEEE Trans. Geosci. Remote Sens.* 2007; 45(6): 1870–1880p.

5. M. Marconcini, G. Camps-Valls, L. Bruzzone.A composite semisupervised SVM for classification of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* 2009; 6(2): 234–238p.
6. F. R. Chung. *Spectral Graph Theory*. Providence, RI, USA: Amer. Math. Soc. 1997; 92.
7. G. Camps-Valls, T. Bandos Marsheva, D. Zhou. Semi-supervised graph-based hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 2007; 45(10): 3044–3054p.
8. T. Bandos, D. Zhou, G. Camps-Valls. Semi-supervised hyperspectral image classification with graphs. *In Proc. IEEE IGARSS*. 2006; 3883–3886p.
9. L. Gómez-Chova, G. Camps-Valls, J.Munoz-Mari et al. Semisupervised image classification with Laplacian support vector machines. *IEEE Geosci. Remote Sens. Lett.* 2008; 5(3): 336–340p.
10. K. Grauman, T. Darrell. The pyramid match kernel: Discriminative classification with sets of image features. *In Proc. IEEE ICCV*. 2005; 2: 1458–1465p.
11. S. Lazebnik, C. Schmid, J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *In Proc. IEEE Conf. CVPR*. 2006; 2: 2169–2178p.
12. J. Yang, K. Yu, Y. Gong et al. Linear spatial pyramid matching using sparse coding for image classification. *In Proc. IEEE Conf. CVPR*. 2009; 1794–1801p.
13. G.-S. Xia, J. Delon, Y. Gousseau. Shape-based invariant texture indexing. *Int. J. Comput. Vis.* 2010; 88(3): 382–403p.
14. F. Tsai, J.-S. Lai. Feature extraction of hyperspectral image cubes using three-dimensional gray-level cooccurrence. *IEEE Trans. Geosci.Remote Sens.* 2013; 51(6): 3504–3513p.
15. T. C. Bau, S. Sarkar, G. Healey. Hyperspectral region classification using a three-dimensional Gabor filter bank. *IEEE Trans. Geosci. Remote Sens.* 2010; 48(9): 3457–3464p.
16. A. H. Tewfik, D. Sinha, P. Jorgensen. On the optimal choice of a wavelet for signal representation. *IEEE Trans. Inform, Theory*. 1992; 38: 747–765p.
17. J. D. Villasenor, B. Belzer, J. Liao. Wavelet filters evaluation for image compression. *IEEE Trans. Image Processing*. 1995; 4: 1053–1060p.
18. D. Dai, L. V. Gool. Ensemble projection for semi-supervised image classification. *In Proc. IEEE ICCV*. 2013; 2072–2079p.

19. M. Gharbi, T. Malisiewicz, S. Paris et al. A Gaussian approximation of feature space for fast image similarity. Massachusetts Instit. Technol., Cambridge, MA, USA, Tech. Rep; 2012.
20. D. G. Lowe. Distinctive image features from scale invariant keypoints. *Int. J. Comput. Vis.* 2004; 60(2): 91–110p.
21. L. Sifre, S. Mallat. Combined scattering for rotation invariant texture analysis. *In Proc. ESANN*. 2012; 127–132p.
22. C. Wengert, M. Douze, H. Jegou. Bag-of-colours for improved image search. *In Proc. ACM Multimedia*. 2011; 1437–1440p.