BAG-OF-VISUAL WORDS AND ERROR-CORRECTING OUTPUT CODES FOR MULTILABEL CLASSIFICATION OF REMOTE SENSING IMAGES

A. Radoi[1], M. Datcu[2]

[1] University Politehnica of Bucharest (UPB), Romania
[2] German Aerospace Center (DLR), Remote Sensing Technology Institute, Oberpfaffenhofen, Germany

ABSTRACT

This paper presents a novel framework for multilabel classification of remote sensing images using Error-Correcting Output Codes (ECOC). Starting with a set of primary class labels, the proposed framework consists in transforming the multiclass problem into binary learning subproblems. The distributed output representations of these binary learners are then transformed into primary class labels. In order to obtain robustness with respect to scale, rotation and image content, a Bag-of-Visual Words (BOVW) model based on Scale Invariant Feature Transform (SIFT) descriptors is used for feature extraction. BOVW assumes an a-priori unsupervised learning of a dictionary of visual words over the training set. Experiments are performed on GeoEye-1 images and the results show the effectiveness of the proposed approach towards multilabel classification, if compared to other methods.

Index Terms— Multilabel classification, BOVW, SIFT, ECOC

1. INTRODUCTION

The continuous advances in satellite technology lead to a tremendous increase in the volume of remote sensing (RS) image archives. In this context, the multilabel classification of remote sensing images has received extensive attention from the RS community. Among the supervised classification methods developed until now, Support Vector Machines (SVMs), alone or combined with other techniques, have shown good performance results [1, 2, 3]. Recent advances have shown that convolutional neural networks (CNNs) achieve high accuracy values for binary classifications [4].

The classification of RS images involves the discrimination between multiple classes. By default, SVMs do not deal with multiple classes directly, but they have to resort to techniques that decompose the multiclass problem into several binary classification problems. The most popular multiclass classification techniques are one-versus-rest (OVR) and one-against-one (OAO) techniques. The OVR strategy consists in conducting one binary classification per class that discriminates between the samples from one class against the samples from the rest of the classes. The OAO strategy consists in pairwise comparisons between classes, yielding a number of $nc(nc-1)/2$ binary classifiers, nc being the number of classes. Numerous experiments have shown that the OAO method is more suitable for practical use than the OVR method, which yields an unbalanced binary classification that becomes even more problematic for an increasing number of classes [5].

A possibility to optimally combine classes into several binary classes is to use Error-Correcting Output Codes (ECOC) [6] that reduce the multilabel classification to binary classification problems. The ECOC approach can make use of algebraic error-correcting codes borrowed from coding theory. The classification scheme presented in [6] associates codewords to class labels and misclassifications are corrected by using a minimum distance approach. In addition, compared to OVR and OAO, the partitions in ECOC may result in a more balanced division of training samples used to learn the binary classifiers.

Assuming that the classification algorithm is established, another critical step in the classification of RS images is feature extraction. One of the widely used methods for feature extraction is the Bag-of-Visual Words (BOVW) that, in many applications, attains remarkable performances [3], but it depends on the extraction of other low level features. In many cases, a common choice for low level feature extraction is the Scale Invariant Feature Transform (SIFT) method that extracts distinctive local features. SIFT features are invariant to scale and rotation, and are robust with respect to distortions, noise or changes in illumination [7]. Sparse representations have also proved to be powerful tools for the classification of remote sensing images, especially in the case of hyperspectral images [8].

In this paper, we propose a new approach towards multilabel classification of RS images based on ECOC defined through algebraic error correcting codes. The features are extracted following a BOVW technique based on SIFT descriptors.

This work has been funded by University Politehnica of Bucharest, through the Excellence Research Grants Program, UPB GEX 2017, Ctr. 32/2017 and by a mobility grant of the Romanian Ministry of Research and Innovation UEFISCDI within PNCDI III (PN-III-P1-1.1-MC-2018-0065).

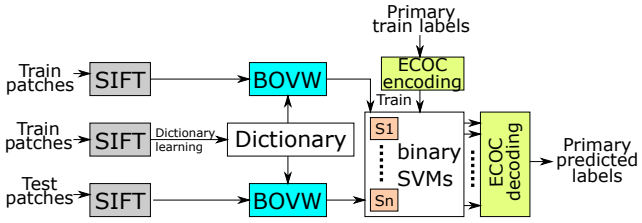


Fig. 1. Scheme of proposed approach for multilabel classification of RS images.

2. PROPOSED APPROACH

Let us consider a RS image divided into N patches and a set of semantic labels of length L. The dataset is divided into two parts: an annotated training set and a testing set containing unlabeled patches. In this work, each patch is characterized by a histogram of word occurrences representing the local distribution of SIFT features with respect to a dictionary. The dictionary is formed by the centroids resulted by clustering SIFT features extracted from available training patches.

As already mentioned, the solution proposed to solve the multilabel classification task is based on encoding and decoding primary class labels into codewords via an algebraic error-correction code. The first step is to define a recombination scheme that provides several different binary partitions of the same set of classes. This step is called *ECOC encoding*. Each combination of primary classes into two classes (0 and 1 master classes) requires training of n separate binary classifiers. Each test sample is passed through the binary classifiers and the inclusion of the test sample into master classes reduces the uncertainty over the class it belongs to. In order to transform the outputs of the binary classifiers into primary class labels, an *ECOC decoding* step is performed. The entire scheme of the proposed approach is in Fig. 1.

3. ERROR-CORRECTING OUTPUT CODES FOR MULTILABEL CLASSIFICATION

3.1. ECOC Encoding

As mentioned above, the classification procedure starts by defining the ECOC encoding scheme that performs multiple partitions of the primary class labels into binary master classes. Considering L classes, each class label can be represented on $k = \lceil \log_2 L \rceil$ information bits. The encoder transforms the sequence of k bits into a *codeword* $\mathbf{c}$ of $n > k$ bits by appending $m = n - k$ control bits. The set of all L codewords $\mathcal{C} = \{\mathbf{c}_0, \mathbf{c}_1, ..., \mathbf{c}_{L-1}\}$ of length n forms the binary block code (i.e., bits are elements in the Galois Field $GF(2)$).

Linear block codes are error-correction codes with special mathematical properties. Each linear code has a given parity check matrix $\mathbf{H}$ and, for each codeword $\mathbf{c}$ in code $\mathcal{C}$, the following property holds true[9]:

$$\mathbf{H}\mathbf{c}^T = 0 \tag{1}$$

Relation (1) provides the encoding rules that express the relations between the control bits and the information bits. All the computations involve linear operations in $GF(2)$ which reduce to multiplications and additions modulo 2.

Linear block codes are widely used in practice for several reasons. Firstly, the encoding and decoding procedures are facilitated by the linearity property of the codes. Secondly, the processing time is smaller than in the case of other codes, e.g., convolutional codes. In this paper, different types of linear codes are used to design ECOC-based multilabel classifiers, i.e., linear cyclic codes and BCH codes. These two codes have special properties. Linear cyclic codes are characterized by the fact that any permutation of a codeword is also a codeword, whereas BCH are powerful error-control codes built to correct a given number of errors. For a more detailed description of these codes, we refer the reader to [9].

3.2. Binary Classifications

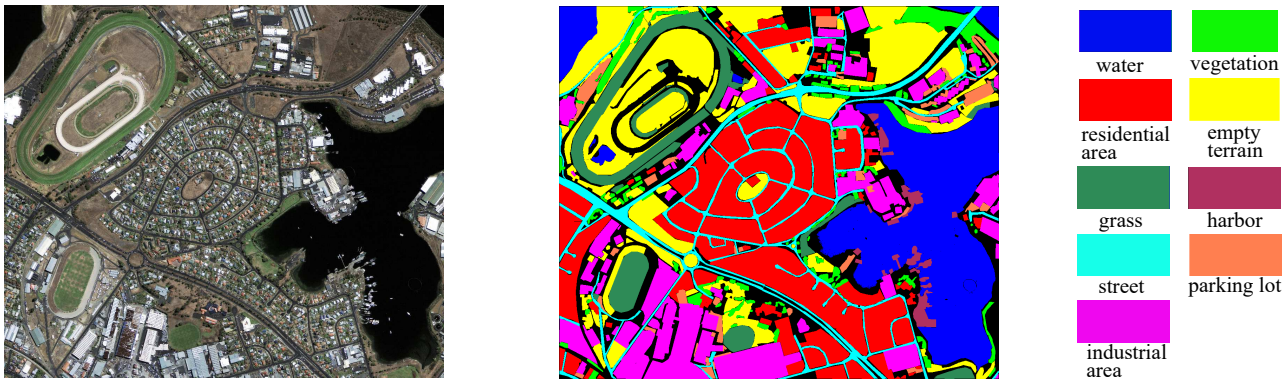
Several well known algorithms can be used as binary classifiers, e.g., SVM, AdaBoost [10]. Since the results reported in [10] for both learners are similar, we choose to use linear SVM as binary classifier [11] to discriminate between master classes 0 and 1. Starting with a training set of t annotated patches $\{(\mathbf{x}_i, y_i)\}_{i=1,...,t}$ ($y_i \in \{0, 1\}$), linear SVM aims at finding a separation hyperplane $\mathbf{w}^T\mathbf{x} + b$ such that the parameters that define the hyperplane ($\mathbf{w}$ and b) minimize the objective function $\frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{t}\epsilon_i$ for some parameter C, subject to linear constraints $y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \epsilon_i$, $i \in \{1, ..., t\}$ and $\epsilon_i \geq 0$.

3.3. ECOC Decoding

In the decoding step, the outputs of the binary classifiers are collected into a vector $\mathbf{r}$ of length n and decoded into the L primary labels. In the case of error-correcting linear codes, syndrome decoding is considered to achieve good performance in terms of computational time and correction properties of the code [9]. In an ECOC-based RS classification framework, errors could appear due to wrong binary classifications. In order to be able to correct some of these errors, once the $\mathbf{r}$ known, a syndrome vector can be computed using:

$$S(\mathbf{r}) = \mathbf{H}\mathbf{r}^T \tag{2}$$

Considering a one-to-one correspondence between possible error configurations and syndromes, the syndrome vector computed for $S(\mathbf{r})$ is used to determine which classifier performed a wrong classification and corrects the corresponding binary output. If the syndrome equals 0, then the decoder decides that no error occurred. The corrected vector is then mapped to one of the primary classes.



(a) GeoEye-1 image (b) Ground truth (c) Legend

Fig. 2. GeoEye-1 and its corresponding ground truth.

Using the syndrome decoding algorithm saves storage and computational time. The look-up table contains only error configurations – syndromes pairs, resulting in a fast correction of errors and decoding. For these reasons, in this work, this decoding strategy is adopted.

4. THE BOVW APPROACH FOR FEATURE EXTRACTION

Feature extraction is performed using a BOVW approach that relies on building a dictionary. One possibility to build a dictionary of D visual words is to use unsupervised learning methods (e.g., K-means clustering) applied on low-level features extracted from training patches. In this work, low-level feature vectors are extracted using SIFT with 4×4 subregions and 8 orientation levels, yielding 128-dimensional feature vectors [7]. In order to provide a better characterization of the patches, SIFT local image descriptors are computed on a dense grid of locations. As the dictionary of D visual words is known, each SIFT descriptor is mapped to the closest visual word in l_2-norm.

Once the dictionary is built, each patch is described by a histogram of occurrences of the dictionary words (i.e., low level feature vectors) in the respective patch. The result is a robust feature extraction method that is invariant with respect to scale, orientation, changes in illumination and objects' positions inside patches.

5. EXPERIMENTS

We tested the proposed algorithms on a GeoEye-1 image (1.65 meters spatial resolution) with multiple complex classes, acquired over Hobart, Tasmania, Australia, on February 5^{th}, 2009. The image contains 3759×3188 pixels and we considered only three spectral bands from the visible domain. Fig. 2 shows the GeoEye-1 image considered in the experiments, along with the reference ground truth containing 9 primary semantic classes, which can be represented on at least $k = \lceil \log_2 9 \rceil = 4$ information bits.

Table 1. Example of codewords (15 bits) associated to each primary label using linear cyclic codes.

Primary label	Codeword
Residential area	0 0 1 1 0 1 0 1 1 1 1 0 0 0 1
Water	0 1 1 0 1 0 1 1 1 1 0 0 0 1 0
Vegetation	0 1 0 1 1 1 1 0 0 0 1 0 0 1 1
Empty terrain	1 1 0 1 0 1 1 1 1 0 0 0 1 0 0
Street	1 1 1 0 0 0 1 0 0 1 1 0 1 0 1
Industrial area	1 0 1 1 1 1 0 0 0 1 0 0 1 1 0
Harbor	1 0 0 0 1 0 0 1 1 0 1 0 1 1 1
Grass	1 0 0 1 1 0 1 0 1 1 1 1 0 0 0
Parking lot	1 0 1 0 1 1 1 1 0 0 0 1 0 0 1

The window size of the input samples is set to 50×50 pixels, leading to a spatial coverage of $82.5 \times 82.5m^2$. These dimensions were selected in order to have an appropriate spatial coverage for the objects of interest. For each patch, feature extraction follows a Bag-of-Visual Words approach with dictionaries of $D \in \{50, 100, 150, ..., 500\}$ visual words. The training set is 25% of the total number of patches, whereas the rest of the patches are used for test.

In the case of cyclic codes, multiple codeword lengths ($n \in \{4, 5, ..., 18\}$) were tested and the mean classification accuracy is shown in Fig. 3. The best result was obtained for a (15,4) systematic cyclic code (i.e., $n = 15$ and $k = 4$). The partitions into binary classes are done according to the 0s and 1s allocations shown in Table 1 which maps each primary label into a codeword. In this case, the first binary classifier considers the first three primary classes as class 0 and the last primary classes as class 1 (on first column). In all experiments, the mapping of primary classes to codewords of code $\mathcal{C}$ is random.

Keeping the same codeword length, we test our framework using BCH codes. Among possible BCH code configurations, we choose BCH(15,5) (i.e., 5 information bits and codewords of 15 bits). Classification results for varying dictionary sizes are shown in Fig. 4 (mean and standard deviation over 10 tests). If compared to SVM OAO approach, the proposed method achieves higher accuracy rates for both cyclic and BCH codes—with highest results for cyclic codes

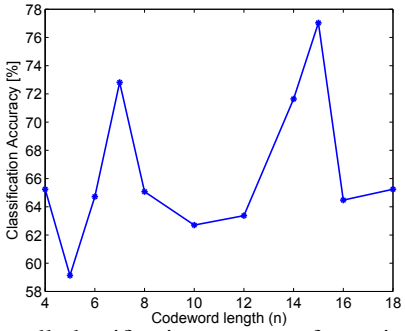


Fig. 3. Overall classification accuracy for various codeword lengths n.

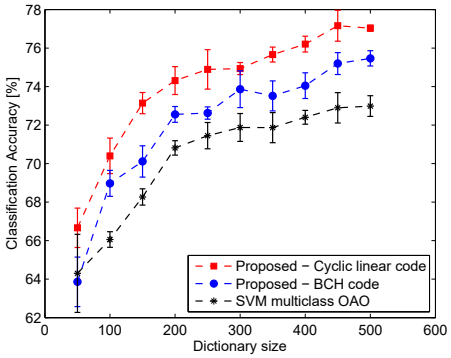


Fig. 4. Comparison of proposed ECOC & BOVW with SVM multiclass OAO & BOVW approach.

Table 2. Comparison of ECOC & BOVW with CNN [%].

Class	ECOC & BOVW	CNN
Residential area	88.64	**95.55**
Water	**97.46**	96.75
Vegetation	**57.58**	50.00
Empty terrain	**83.72**	74.25
Street	**75.38**	47.05
Industrial area	72.15	**75.71**
Harbor	53.37	**80.00**
Grass	71.88	**89.55**
Parking lot	**73.24**	44.44
Overall accuracy	**77.23**	76.21

(77.23%) with a dictionary of 450 words and codewords of 15 bits for ECOC. In Table 2, we compare our results with the ones obtained by applying a CNN-based architecture comprised of four convolutional layers (two with 32 filters and two with 64 filters) and pooling layers, followed by a fully connected layer. In most of the cases, the per-class accuracy of ECOC & BOVW is higher or close to CNN's performance.

6. CONCLUSIONS

In this paper, we have presented Error-Correcting Output Codes in the context of RS image multilabel classification problems when algebraic error correcting codes (i.e., cyclic codes and BCH) are used. The feature extraction module is built using a Bag-of-Visual-Words approach based on SIFT low-level features. The decomposition and recombination of the multilabel problem into binary classification problems is done using the encoding / decoding schemes that characterize the previously mentioned codes. The proposed method outperformed the SVM-OAO approach applied on BOVW features. A final remark is the lower number of binary classifiers needed to be learned in case of ECOC proposed framework ($n = 15$) compared to $9 \times 8/2 = 36$ binary classifiers required by SVM-OAO approach.

7. REFERENCES

[1] G. M. Foody and A. Mathur, "A relative evaluation of multiclass image classification by support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 6, pp. 1335–1343, 2004.

[2] A. C. Grivei, A. Radoi, C. Vaduva, and M. Datcu, "An active-learning approach to the query by example retrieval in remote sensing images," in *Intl. Conference on Communications*, 2016, pp. 377–380.

[3] B. Demir and L. Bruzzone, "A novel active learning method in relevance feedback for content-based remote sensing image retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2323–2334, 2015.

[4] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 645–657, Feb 2017.

[5] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.

[6] T. G. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes," *Journal of Artificial Intelligence Research*, vol. 2, no. 1, pp. 263–286, 1995.

[7] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[8] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 8, pp. 4186–4201, Aug 2015.

[9] E. R. Berlekamp, *Algebraic Coding Theory - Revised Edition*, World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2015.

[10] E. L. Allwein, R. E. Schapire, and Y. Singer, "Reducing multiclass to binary: A unifying approach for margin classifiers," *J. Mach. Learn. Res.*, vol. 1, pp. 113–141, 2001.

[11] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.