IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)
Volume 4, Issue 2, Ver. IV (Mar-Apr. 2014), PP 01-04
e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197
www.iosrjournals.org

Unsupervised Learning for Satellite Image Classification

Giriraja C.V, Srinivasa C, T.K. Jaya Ram, Avula Haswanth
Department of Electronics and Communication Engineering , Amrita School Of Engineering, Bangalore, India

Abstract: *A new method of classifying satellite images into different categories such as forest, desert, river etc., with the help of Support Vector Machine (SVM) and unsupervised learning method using MATLAB is presented. Dense low-level feature descriptors are extracted and exploited in a novel way to learn a set of basis functions. The low-level feature descriptors are encoded in terms of the basis functions to generate new sparse representation for the feature descriptors. We pool these sparse features by simple averaging and show that this method will help us to choose SVM classifier which is having less cost. Experimental results show its good performance, when tested with different saliency criteria, such as accuracy, less computational time, and the combination of both.*

I. Introduction

The high-fidelity image data provided by the new and advanced space-borne sensors provide fresh opportunities to characterize aerial scenes based on the spatial and structural patterns encoded in the imagery. Efficient representation and recognition of scenes from image data are challenging problems. Most of the previous approaches for high-resolution satellite image analysis [1] focus on classifying pixels or objects (grouping of local homogeneous pixels) into their thematic classes by extracting spectral, textural, and geometrical attributes as classification features. Often, the objective is to model scenes by aggregating the classes in a bottom-up manner. In contrast, we focus on directly modeling scenes by exploiting the variations in the local spatial arrangements and structural patterns captured by the low-level features. Our approach all-lows us to develop a holistic representation for aerial scenes that does not require intermediate stages of segmentation and representation of individual geospatial objects. The proposed unsupervised feature learning and encoding strategy maps low-level feature descriptors to a new representation that is highly accurate in characterizing different aerial scenes. With high-resolution image data, aerial scenes are often comprised of different and distinct thematic classes. For example, an image patch associated with a scene representing commercial or large-facility class might comprise different classes such as forests, river, desert, deciduous vegetation, land cover etc., Encoding the local structural and spatial scene attributes in an efficient and robust fashion is the key to generating discriminatory models to classify such aerial scenes. Direct modeling of aerial scenes based on low-level feature statistics is a popular idea. Bag-of-visual-words (BOVW) [1] is a feature encoding approach that has been well explored for scene classification. Recent studies [2],[3] have shown that sparse coding of features is highly effective for scene classification compared to the traditional BOVW approaches. Our proposed method involves generating a set of basis functions from unlabeled features. The low-level feature descriptors extracted from the scene are encoded in terms of the basis functions to generate spare feature representations. We show that simple statistics generated from these sparse features characterize the scene well producing significant improvement in scene classification accuracies compared to existing approaches reported in [4], [5]. The proposed sparse feature representation works with linear classification model, yet out-performing classification performance of other methods that use complex nonlinear classification models. In this paper, we also evaluated the classification performance of various low-level feature measurements such as raw pixel intensity values, oriented filter responses, and local scale invariant feature trans-formation (SIFT)-based feature descriptors [6].
The major contributions of this paper are:

i. Unsupervised feature learning approach to generate feature representation for various high-resolution aerial scenes.
ii. Extensive experiments with various low-level feature measurements such as raw pixel intensities, oriented filter responses, and SIFT feature descriptors.
iii. Evaluation of the methodology with different and diverse data sets.
iv. Detection system based on the proposed feature extraction and learning approaches for detecting large-facility in large-scale high-resolution aerial imagery.

The rest of the paper is organized as follows.
In Section 2, Block diagram for unsupervised satellite image classification.
In Section 3, we describe our approach on unsupervised feature learning in detail.
Section 4 provides the Algorithm for overall classification framework.

Sections 5 concludes the paper with discussions on the findings and ideas for extending the work

II. Block Diagram

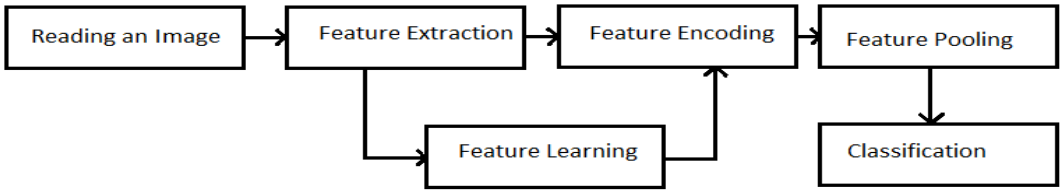

Fig.1. Block diagram for unsupervised feature learning for satellite image classification

III. Unsupervised Feature Learning

Here, the goal is to accurately classify the given image patch into one of the predefined scene categories. Our approach consists of five broad steps

i. Feature extraction
ii. Feature learning
iii. Feature encoding
iv. Feature pooling
v. Classification

We begin by extracting low-level feature descriptors from the image patch. As part of the feature learning process, we compute a set of normalized basis functions from the extracted features in an unsupervised manner. We use a variant of sparse coding called Orthogonal Matching Pursuit (OMP-k) [7] to compute the basis function set. During feature encoding, we project the features onto the learned basis function set and apply soft threshold activation function to generate a set sparse features. We pool the sparse features to generate the final feature representation for the image patch. The final features are then fed to a linear support vector machine (SVM) classifier. Fig. 1 shows the overview of the proposed framework. Next, we describe our dense feature extraction strategies and subsequent steps in detail.

3.1 Feature extraction

For feature extraction we use SIFT (Scale Invariant Feature Transform) [8] based descriptor for each block. Normal SIFT descriptors are operated on only interested points so we use dense SIFT descriptors. Before applying SIFT descriptor each image is divided in to 4X4 sub blocks. For each sub block a magnitude weighted orientation histogram is computed. The orientations are divided into 8 intervals. The magnitudes are further weighted by a Gaussian function with σ equal to one-half the width of the descriptor window. Local histograms are stacked to form the feature vector.

3.2 Feature Learning

The main idea here is that the feature elements representing spatially adjacent pixels might exhibit high correlation. By removing these correlations we can force the model to learn the high-order structure in the data. We will achieve this by randomly sampling low level features in a data set to generate a matrix X. The matrix is normalized by subtracting the mean and dividing by the standard deviation. Next, to whiten the data we apply a Zero Component Analysis (ZCA) [9] transform. By using whitened matrix we learn the basis function D by finding the best solution for minimization of error.

3.3 Feature encoding

The main objective here is to generate a robust representation that effectively and efficiently encodes the patterns in the image. We will achieve this by representing low-level features in terms of basis functions. To represent the features in terms of the basis functions, we project the feature descriptor onto the basis vectors represented in the set D to compute the linear weights [10]. Next we apply a soft threshold activation function to generate sparse features. In this encoding scheme positive and negative weights above and below certain thresholds are retained and remaining elements are forced to zero to generate sparse features. These sparse features are stacked to form the sparse feature.

3.4 Feature Pooling

With the sparse features [11] computed for an image patch, we can estimate the final feature representation based on simple statistics of the sparse features. We are using simple averaging to pool the sparse features, so that the cost for classifying will be less and we can use linear classifiers.

3.5 Classification

3.5.1 Unsupervised Scene Classification

A new approach to unsupervised classification for multispectral imagery. It first implements the pixel purity index (PPI) which is commonly used in hyper spectral imaging for end member extraction to find seed samples without prior knowledge, then uses the PPI-found samples as support vectors for a kernel-based support vector machine (SVM) to generate a set of initial training samples. Since PPI uses a random generator to produce skewers on which data samples are orthogonally projected, the PPI-found samples are not reproducible. In order to resolve this inconsistency issue, an iterative Fisher's linear discriminant analysis (IFLDA) is further developed to implement FLDA [12] iteratively to mitigate the instability caused by the use of skewers and the sensitivity of using PPI-found samples as support vectors by SVM. These resulting IFLDA-classified sample vectors are then used as a final set of training samples which are in fact obtained by a series of processes by first finding seed samples via PPI, then by using SVM to generate initial training samples which are further refined and corrected by IFLDA to produce a final desired set of training samples. So, an algorithm implementing PPI coupled with SVM in conjunction with IFLDA to find unsupervised training samples is called PPI-SVM-IFLDA algorithm. Finally, these PPI-SVM-IFLDA-produced samples are then used as training samples for a follow-up classifier to perform supervised classification on the original data samples in which case this supervised classifier is once again implemented by IFLDA with only difference in that this IFLDA is now performed on the entire data space compared to the IFLDA used in the PPI-SVM-IFLDA which only performs the classification on SVM-generated training samples, not original data samples.

3.5.2 Unsupervised PPI-SVM-IFLDA Classification Algorithm

i. Initial condition: Set the number of classes to the number of spectral bands.
ii. Sphere the data.
iii. Implement PPI-SVM-IFLDA on the sphered data to produce a desired set of training samples.
iv. Using the training samples obtained in Step 3 to perform IFLDA on the original data to produce classification results.

IV. Algorithm : Feature Encoding and Classification

Steps :

i. Extract the low level features vectors from the test image
ii. Find the basis function from the set of feature vectors found in above steps
iii. Normalize the feature vector set obtained in the step1
iv. Apply the Zero Competent Analysis (ZCA) to remove the redundancy
v. Represent the feature vectors in terms of basis function
vi. Apply the threshold on the above (step 5) obtained set of values to get the sparse features
vii. Pool the sparse features
viii. Perform SVM binary decision

V. Discussion

In contrast to previous works on satellite image classification where the focus was on pixel or object-level thematic classification, here we explore a method to directly model scene by exploiting the local spatial and structural patterns in the scene. Our approach model scenes by passing the complicated steps of segmentation and individual segment classification. The proposed classification framework involves dense feature extraction, learning, encoding and pooling. Rather than using the low-level feature measurements directly in the classification framework, we derive sparse feature representations by encoding these features in terms of a learned basis function set. The basis function set is generated in an unsupervised manner. We show that the pooled sparse features employed with a linear SVM kernel out performs existing methods in terms of classification accuracy. In the case of large-facility detection, we obtain a high measure producing excellent detections on large-scale high-resolution satellite imagery.

As future extensions, we plan to extend this approach to encode high-level spatial information and shape based features as part of the feature encoding process. The current feature encoding process is simple, yet provides good classification accuracy for broad neighborhood classes. However, to model complex structures it would be highly beneficial to encode high-level information. Another straightforward extension would be to follow a supervised framework for the basis function set generation. Other interesting applications can result

from combining our proposed scene detection approach with scene parsing to identify the individual geospatial objects within the scene. A similar idea has been explored earlier with a few simple object categories such as buildings, cars, trees, and road.

Acknowledgments

This work was fully supported by Dr. Rakesh S.G, principal of Amrita School Of Engineering, Bangalore, along with Dr.Shikha Tripathi, HOD of Electronics and Communication and many more. We are also very grateful for the comments made by the Principal, HOD of the department, faculties etc., their input helped in improving the quality of the final version of this paper. Thank you very much!

References

[1] Y.Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in Proc. ACM Int. Conf. Adv. Geogr. Inf. Syst.,2010, pp. 270–279.

[2] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and Y. Shuicheng, "Sparse representation for computer vision and pattern recognition,"Proc. IEEE, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[3] Y.-L. Boureau, F. Bach, Y. LeCun, and J. Ponce, "Learning mid-level features for recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2010, pp. 2559–2566.

[4] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," inProc. IEEE Conf. Comput. Vis. Pattern Recognit., 2006, vol. 2, pp. 2169–2178.

[5] Y. Yang and S. Newsam, "Spatial pyramid co-occurrence for image clas-sification," inProc. IEEE ICCV, Nov. 2011, pp. 1465–1472.

[6] D. G. Lowe, "Object recognition from local scale-invariant features," in Proc. IEEE Int. Conf. Comput. Vis., Kerkyra, Greece, 1999, pp. 1150–1157.

[7] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in Proc. Asilomar Conf. Signals, Syst. Comput., 1993,pp. 40–44.

[8] Hao Tang and Feng Tang, "AH SIFT: Augmented histogram based SIFT descriptors", *IEEE Conf.* 2012, pp. 2357-2360

[9] A. Hyvarinen and E. Oja, "Independent component analysis: Algo-rithms and applications," Neural Netw., vol. 13, no. 4/5, pp. 411–430, May/Jun. 2000.

[10] A. Coates and A. Y. Ng, "The importance of encoding versus training with sparse coding and vector quantization," inProc. Int. Conf. Mach. Learn., 2011, vol. 28, pp. 921–928.

[11] R. Rigamonti, M. A. Brown, and V. Lepetit, "Are sparse representations really relevant for image classification?" in *Proc. IEEE* Conf. Comput. Vis.Pattern Recognit., 2011, pp. 1545–1552.

[12] *Hsian-Min Chen, ChinsuLin, Member, Shih-Yu Chen, Member, Clayton Chi-Chang Chen, "*PPI-SVM-Iterative FLDA Approach to Unsupervised Multispectral Image Classification", *Proc. IEEE Conf. Applied Earth Observations and Remote Sensing, VOL. 6, No. 4, August 2013*