Using locality-constrained linear coding in automatic target detection of HRS images

M Rezaee¹, Z Mirikharaji², Y Zhang¹

¹ University of New Brunswick, Department of Geodesy and Geomatics Engineering, Fredericton, New Brunswick, Canada
² University of New Brunswick, Faculty of Computer Science, Fredericton, New Brunswick, Canada

E-mail: mrezaee@unb.ca

Abstract. Automatic target detection with complicated shapes in high spatial resolution images is an ongoing challenge in remote sensing image processing. This is because most methods use spectral or texture information, which are not sufficient for detecting complex shapes. In this paper, a new detection framework, based on Spatial Pyramid Matching (SPM) and Locality-constraint Linear Coding (LLC), is proposed to solve this problem, and exemplified using airplane shapes. The process starts with partitioning the image into sub-regions and generating a unique histogram for local features of each sub-region. Then, linear Support Vector Machines (SVMs) are used to detect objects based on a pyramid-matching kernel, which analyses the descriptors inside patches in different resolution. In order to generate the histogram, first a point feature detector (e.g. SIFT) is applied on the patches, and then a quantization process is used to select local features. In this step, the k-mean method is used in conjunction with the locality-constrained linear coding method. The LLC forces the coefficient matrix in the quantization process to be local and sparse as well. As a result, the speed of the method improves around 24 times in comparison to using sparse coding for quantization. Quantitative analysis also shows improvement in comparison to just using k-mean, but the accuracy in comparison to using sparse coding is similar. Rotation and shift of the desired object has no effect on the obtained results. The speed and accuracy of this algorithm for high spatial resolution images make it capable for use in real-world applications.

Introduction

Target detection is one of the most important applications in the remote sensing field. Technological advances have brought high spatial resolution imagery for Earth monitoring, with new challenges in processing. Different attempts have been made to reach an acceptable accuracy in target detection; however, targets with complex shape are difficult to detect using spectral based methods [1]. One of the objects with complex shape is the airplane.

Based on [2], airplane detection can be categorized into two branches: one that tries to detect them based on segmentation result [3], and the other one that works based on edge detection [4]. The segmentation category first segments the image and then tries to detect airplanes between obtained segments. The edge-based methods rely on the unique shape of the airplane and try to detect them by processing the extracted edges. Some hybrid methods that use both categories to detect airplanes. [5], [6] used edges of runways and implemented a region growing segmentation method to detect the airports.
Bag of Features (BoF) is one of the methods of classification, initially proposed for text processing. The BoF tries to find a unique histogram for the specified text and use it as a kernel for matching the content. It is used in various classification applications in the remote sensing field with high spatial resolution imageries [2-3]. Since BoF discards the spatial order of local descriptors, its usage in target detection is limited. Spatial Pyramid Matching (SPM), an extended version of BoF, was proposed by [4]. The method divided the image into $2^l \times 2^l$ segments, with different values for $l$. Then, within each segment, the BoF histogram is generated. These histograms then shape the kernel of the classifier for classification. In the case where $l = 0$, the SPM turns to BoF.

The main disadvantage of the SPM is that it cannot use a linear classifier for the classification. Using a nonlinear kernel in a classifier such as Support Vector Machines (SVMs) improves the complexity of the system in the training phase, but the solution is complex and time consuming. This produces severe limitation in the classification process. In order to solve this problem, [5] proposed using Sparse Coding in SPM (ScSPM) for quantization, instead of k-mean. The method decreased the complexity of the SPM for using a linear classifier and obtained state-of-the-art accuracy using typical imagery data sets. In another study, [6] empirically observed that non-zero coefficients in Sparse Coding are local. This means that these coefficients are usually assigned to the bases around the encoded data. They proposed a modification to Sparse Coding to force the method to use local coefficients. The new method is called Local Coordinate Coding (LCC). Later, [7] proposed Locality-constrained Linear Coding (LLC) as a fast analytical solution. In this paper, the LLC method in an SPM workflow is used for detecting airplanes in a single panchromatic image from a Worldview-2 image. The rest of the paper is as follows: section two explains the proposed method, and section three presents the results and discusses the advantages and concerns of the algorithm.

Methodology
Since the method works in the patch-based system, the training and testing image should be divided into patches. For patching, a slider with the size of 100x100 is moved over the images with steps of 4 pixels between them, and the covered areas are clipped. After that, the Scale-Invariant Feature Transform (SIFT) is run over all the patches and the detected points and their descriptors are all saved. Let $X$ be the extracted descriptors, i.e $X = \{x_1, ..., x_N\} \in \mathbb{R}^{N \times D}$, where $D$ is the dimension of the vector, and $N$ shows the number of descriptors. In the next step, a vector quantization (VQ) method is used for all descriptors. In BoF and SPM, the k-mean clustering algorithm is used for VQ, which can be denoted as:

$$\min_{u, v} \sum_{n=1}^{N} \min \|x_n - u_n v\|^2$$

(1)

where $V$ is the vector of cluster centers, $k$ is the number of clusters, and $u$ is the cluster membership indicator. The cluster centers are called codebook. There are two constraints in the k-mean formula. First is the cardinality constraint, which forces $u$ to pick one non-zero element for each $n$. The second constraint forces the elements in $u$ to be zero or positive.

The cardinality constraint hinders the reconstruction of the $X$, and usually results in a coarse reconstruction. Therefore, in Sparse Coding, this constraint is relaxed using an L1-norm regularization over the $u$. In this case, $u$ can have a small number of non-zero elements. Thus, the VQ formula can be rewritten as follows:

$$\min_{u, v} \sum_{n=1}^{N} \min \|x_n - u_n v\|^2 + \lambda|u_n|$$

(2)
The regularization term (Sparsity regularization term) makes it possible to generate an overcompleted codebook, and it decreases the error related to the quantization. Since Sparsity Coding tries to solve the L1-norm of $u$, the process is time consuming.

Yu et al. [6] reported they observed empirically that Sparse Coding results are local. Therefore, instead of the sparsity term, they suggested the locality term. This new term, which is called Local Coordinate Coding (LCC), assures sparsity as well; however, it still needs to solve the L1-norm regularization. [7] proposed an analytical solution for the LCC. It is called Locality-constrained Linear Coding (LLC). This solution is a fast implementation of LCC. In this solution, the VQ formula can be rewritten as follows:

$$\min_{u,v} \sum_{n=1}^{N} \min \|X_n - u_n V\|^2 + \lambda \|d_n \odot u_n\|$$

where $\odot$ denotes the element-wise multiplication and $d$ is defined as follows:

$$d_i = \exp\left(\frac{\text{dist}(X_n, V)}{\sigma}\right)$$

where $\sigma$ is used for tuning the weight decay speed, and $\text{dist}(X_n, V) = [\text{dist}(X_n, V_1), ..., \text{dist}(X_n, V_n)]^T$. The LLC has a better reconstruction in comparison to k-mean and has an analytical solution, which is fast. In order to use LLC for airplane detection in this paper, the following steps are implemented:

1. Using a k-mean algorithm, the codebook is generated.
2. In each iteration, a small portion of descriptors is selected and Equation 3 is solved over them.
3. The coefficients whose numbers are more than a pre-defined threshold are kept and the rest are set to zero.
4. A gradient descent method is used to update the coefficients.

After extracting the coefficients and codebook, the histogram is generated for each image in different levels. Based on SPM, first a histogram is generated for the entire image (Level 0). Then the image is divided into 4 parts, and for each part a new histogram is generated (Level 1). This can be continued into Level 3 as well.

When the histograms are generated for all the trained images, the whole process should be run on the test image. The histograms of each patch of the test image are compared with the histogram of the trained image. For the comparison, the histograms are pooled to each other in each level. Results of each level are then used in Equation 5 to generate the kernel of classification. In the end, each patch is assigned to one of the classes.

$$\text{weight} = l_3 + \frac{1}{2}(l_2 - l_3) + \frac{1}{4}(l_1 - l_2) + \frac{1}{8}(l_0 - l_1)$$

where $l$ denotes the intersection of the histograms of the trained image and test image, and its subscript shows the level’s number.

**Results and Discussion**

In this section, results of the method are shown and discussed. The data used in this paper are panchromatic images from the Worldview-2 satellite. The images are from different cities’ airports. Since the method works on a patch-based system, the training data and testing data should be patched. For training, two sets of data are created: one for the airplanes and the other for background, which cover most of the area that doesn’t have any airplane inside. Figure 1 shows two sets of training data.
The test scene is also patched using a sliding window, moving over the image. Since the sizes of the airplanes are different, the largest size of the airplanes in the image determines the sliding window’s size. In this image, 64 (pixels) is selected for the window size. Figure 2 shows the test data.

Figure 1: Two sets of training data: a. Airplane, b. Background

Figure 2: The test data. It contains an airport with different airplanes in various scale and rotations.
method is trained using 30 patches of each set. Then the method is run over the test and validation data. Figure 3 shows the results on the test data.

For evaluating the results, two sets of data are generated, one for the airplane and the other for the background. The method is run over these two datasets for training. Then, the target image is divided into patches and it turns into two airplane and background categories as well. The method is run over the target patches, and the obtained results are compared to the real labels. The evaluation shows 84% accuracy in detecting the airplane. This accuracy is almost the same with ScSPM (86%). However, the speed of the algorithm is 20x faster in comparison to the ScSPM, due to the analytical solution of LLC.

**Figure 3**: Results on the test image: each yellow circle represents one patch that contains an airplane or part of one, detected by the method.

**Figure 4**: The airplane is detected in four different patches.
Conclusion
Automatic target detection of complicated shapes in high spatial resolution images is a challenge in the remote sensing field. In this paper, a detection framework based on Spatial Pyramid Matching (SPM) and Locality-constraint Linear Coding (LLC) is proposed to solve this problem. The accuracy and the speed of the LLC method in this paper for detecting objects in a single panchromatic image in calculated and compared to ScSPM. The comparison shows a similar detection accuracy in both methods; however, the speed of processing in LLC is 20 times faster than ScSPM, because of the analytical solution in LLC. The results showed that the LLC method is suitable for use in object detection in high spatial satellite images, using just one panchromatic band. However, there are some issues that need more study. The first one is the dependency of the algorithm on the number of training data. Based on the type of the object to be detected, or the environment that the object exists inside, the number of training data will need to differ. The other point is the number of iterations of the method for solving at a given locality. This parameter can affect both accuracy and speed of the algorithm.

References
[1] S. Bo and Y. Jing, Region-based airplane detection in remotely sensed imagery, Image and Signal Processing (CISP), 2010 3rd International Congress on, vol. 4, pp. 1923–1926, 2010.
[2] S. Xu, T. Fang, D. Li, and S. Wang, Object classification of aerial images With Bag-of-Visual words, Geoscience and Remote Sensing Letters, IEEE, vol. 7, no. 2, pp. 366–370, 2010.
[3] M. Lienou, H. Maitre, and M. Datcu, Semantic annotation of satellite images using Latent Dirichlet Allocation, Geoscience and Remote Sensing Letters, IEEE, vol. 7, no. 1. pp. 28–32, 2010.
[4] S. Lazebnik, C. Schmid, and J. Ponce, Beyond bags of features: spatial pyramid matching for recognizing natural scene categories,” 2006 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Vol. 2, vol. 2, pp. 2169–2178, 2006.
[5] J. Yang, K. Yu, Y. Gong, and T. Huang, Linear spatial pyramid matching using sparse coding for image classification, 2009 IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1794–1801, Jun. 2009.
[6] K. Yu, T. Zhang, and Y. Gong, Nonlinear learning using local coordinate coding, in Advances in Neural Information Processing Systems 22, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, Eds. Curran Associates, Inc., 2009, pp. 2223–2231.
[7] J. Wang, J. Yang, K. Yu, and F. Lv, Locality-constrained linear coding for image classification, Comput. Vis ..., pp. 3360–3367, Jun. 2010.