DRF-based Object Detection Using the Object Adaptive Patch in the Satellite Imagery

Hyoungmin Choi, Kyoung-Mu Lee, Sang-Uk Lee

School of Electrical Engineering
ASRI
Seoul National University

E-mail: choihyoungmin@diehard.snu.ac.kr, kyoungmu@snu.ac.kr, sanguk@ipl.snu.ac.kr

ABSTRACT

In this paper, we propose a DRF-based object detection method using the object adaptive patch in the satellite imagery. It is a Discriminative Random Fields (DRF) based work, so the detection is done by labeling to the possible patches in the image. For the feature information of each patch, we use the multi-scale and object adaptive patch and its texton histogram, instead of using the single scale and fixed grid patch. So, we can include contextual layout of texture information around the object. To make object adaptive patch, we use "superpixel lattice" scheme. As a result, each group of labeled patches represents the object or object's presence region. In the experiment, we compare the detection result with a fixed grid scheme and shows our result is more close to the object shape.

Keywords: object detection, multi-scale DRF, object adaptive patch, texton histogram

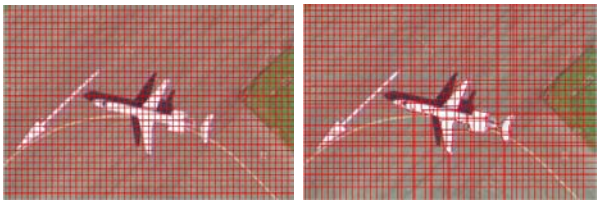
1. Introduction

The practical use of the satellite imagery is increasing in these days, not only for the military surveillance purpose but also science research. And object detection and segmentation is a major tool for those kinds of works. Generally the satellite imagery contains both target objects and various surroundings, and has a large size so it makes the target object being a relatively small. Recently, the "TextonBoost" algorithm [3] has been showed impressive object detection and classification results. However, most existing approaches, including [3], are not aimed at the satellite imagery. So, the target object in image is distinctive so that we can easily extract the object's intrinsic information. It would be of interest to study the randomly distributed and relatively small size object detection method in satellite images. In this work, we want to localize objects in the satellite imagery with patch labeling method. Most object detection algorithms use intrinsic object feature information, such as color, texture, and geometric characteristics, learnt from training images. However, we assume that there is not enough object feature information due to its size in satellite images. So, we use surrounding information of the object and its boundary, instead of implicit object feature information. Our approach has two contributions. First, it does not only use an implicit object feature but also use an object surrounding feature information for training. We divide the image into superpixel patches [9] and extract texton histogram from each patch by scales. We adopt multi-scale feature extract scheme from [1][4] so that we can reduce the ambiguity which is caused by the similar histogram pattern. This scheme is based on the fact that the specific object has constant surroundings, such as car and road, airplane and runway, so combining the object feature and surrounding feature can be useful to infer the existence of objects when we don't have enough intrinsic information. Second, we use an object adaptive patch so that we can extract the feature information more efficiently than fixed size grid patch (Figure 2) and make the labeling result more close to the object (Figure 1).

Figure 1. The object detection result from our method which use the texton histogram and object adaptive patches. The red line shows a boundary of a detected (labeled) region .

2. Related Works

We treat the object detection task as a labeling work using DRF framework, which is proposed by Kumar et al. [1] for the first time. And they had applied it to detect man-made objects from the natural scene. They had proved that the performance of DRF is better than the MRF based detection work. However, they had used the intensity gradient of each patch as a feature which is not appropriate to detect the specific object, like an airplane or car. Xuming He et al. [4] proposed the multi-scale conditional random field framework for labeling images into a predefined set of class labels by considering the relationship between objects and its surroundings. The regional label features and global label features provide different information from different aspects of the image so



(a) (b)

Figure 2. Comparison between the 8x8 fixed grid patch (a) and the object adaptive grid patch (b). Although (a) and (b) image have same number of patches but patches in (b) show that their shape is fit to the target object.

that they could label more accurately. There are some object detection researches using the texton information. Lin Yang et al. [6] proposed a unified object classification method, which use a color, texton histogram, keypoint and object boundary information. The texton histogram of each patch was one of the important information to discriminate objects with others. And they constructed a texton histogram classifier using the SVM. However, they focused on extracting the intrinsic information from objects in this work. J. Shotton et al. [3] proposed the "TextonBoost" algorithm which use a texton, color and object locations information. They used a shape-texture filter response as a one of the member of unary potentials. It uses a texton map to produce the shape-texture filter, and each texton represents an object class. This filter is a kind of classifier to detect objects and provide an initial probability of object's presence. However, the last two methods much depend on the implicit information of objects for detection. So, if we want to detect relatively small objects, which have not enough features or texton information, these methods cannot provide the proper detection result. Our goal is to detect these relatively small objects from the satellite image by using only the texton histograms.

3. Proposed Algorithm

We use a Discriminative Random Field (DRF) framework [1] to learn and detect the object from given images. The use of a DRF allows us to incorporate the information not only from the object but also from object's surroundings. For a discriminative work, we divide the image ($\mathbf{I}$) to 8-by-8 grid and call it 'patch'. We define the conditional probability of the object label $\mathbf{x}$ given an image $\mathbf{I}$ as

$$\log P(\mathbf{x} \mid \mathbf{I},\mathbf{s},\boldsymbol{\theta}) = \sum_i \psi_i(x_i, f_i; \boldsymbol{\theta}_{\boldsymbol{\psi}}) + \sum_{(i,j)\in\varepsilon} \phi(x_i, x_j, g_{ij}(f_i, f_j); \boldsymbol{\theta}_{\boldsymbol{\varphi}}) - \log Z(\boldsymbol{\theta},\mathbf{I}) \quad (1)$$

where ε is the set of edges in the graph, $Z(\boldsymbol{\theta},\mathbf{I})$ is the partition function, $\boldsymbol{\theta} = \{\boldsymbol{\theta}_{\boldsymbol{\psi}}, \boldsymbol{\theta}_{\boldsymbol{\varphi}}\}$ are the model parameters, and i (j) is index node in the graph (corresponding to each patch in the image).

3.1 Multi-scale texture (ψ) and pairwise potential (ϕ)

The multi-scale texture potentials ψ use texton histograms of each patch at different scale as a feature vector f_i and the pairwise potential ϕ use feature vectors to compare their similarity.

To compute the texton histogram, we produce a texton which is a set of clusters made from MR8 filterbank [2] convolution with training images and by K-means. Using this texton, we can make a texton map to extract the histograms. Then we divide the image into a set of superpixel patches which is proposed by [9] and use the multi-scale feature extraction scheme, the feature vector f_i at patch i has N (a number of scales) different texton histograms h_N^i and each histogram is a K (a number of textons) dimensional vector.

$$h_N^i = \sum_{k\in K} count(Texton = k) \quad (2)$$

$$f_i = \sum_N h_N^i \quad (3)$$

Therefore, the multi-scale texture potential is defined as,

$$\psi(x_i, f_i; \theta_\psi) = \log P(x_i \mid f_i)\,, \quad (4)$$

and by the linear logistic regression,

$$\log P(x_i = 1 \mid f_i) = \frac{1}{1+\exp^{-(\boldsymbol{\theta}_{\boldsymbol{\psi}}^{\mathbf{T}} f_i)}} \quad (5)$$

The pairwise potential is expressed as,

$$\phi(x_i, x_j, g(f_i, f_j); \theta_\phi) = \log P(x_i, x_j \mid g(f_i, f_j)) \quad (6)$$

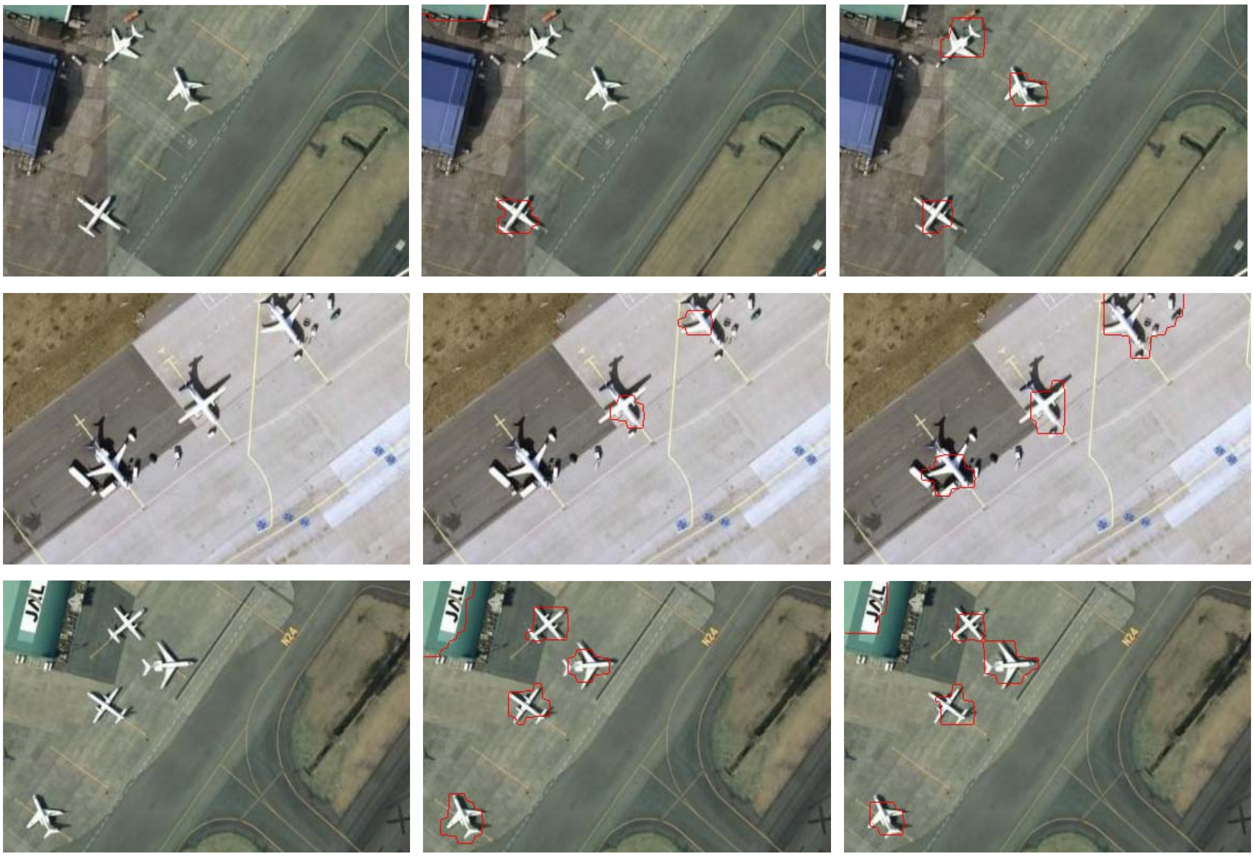
where $g(f_i, f_j)$ is a similarity measure between neighboring histograms. And we slightly modified the X^2 distance measure using $\boldsymbol{\theta}_{\boldsymbol{\varphi}}$ as,

$$g(f_i, f_j) = \frac{1}{2}\sum_n \theta_\phi^n \frac{\left(f_i - f_j\right)^2}{f_i + f_j} \quad (7)$$

where n is a size of the parameter which is same as a size of the feature vector. The advantage of this multi-scale feature extraction scheme is that we can reduce the ambiguity which can be occurred if two histograms have a similar pattern.

3.2 Parameter Estimation

From the (1), we have a set of parameters of our DRF framework which is $\boldsymbol{\theta} = \{\boldsymbol{\theta}_{\boldsymbol{\psi}}, \boldsymbol{\theta}_{\boldsymbol{\varphi}}\}$. The DRF framework is similar to the MRF, so it is possible to use the technique which is used for learning the MRF parameter to get our model parameters. So it is a Maximum a posteriori (MAP) estimation problem with respect to the parameters and the maximum-likelihood approach is the most popular approach for the parameter estimation.



(a) (b) (c)

Figure 3. (a) is an input image, (b) is a result from using a fixed grid patch scheme, and (c) is a result from the proposed method. Our method shows better performance than the fixed grid patch scheme. And the red line represents the boundary of the maximum probability of the object existence region.

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \prod_{m} P(\mathbf{x}^m \mid \mathbf{I}^m, \boldsymbol{\theta}) \tag{8}$$

where m is a number of training data.

We apply the stochastic gradient optimization method with a gain vector adaptation which had been proposed in [7] and it is fast when the training data is large.

4. Experimental Result

The satellite images for our experiment are captured from the "Google Earth" and size of 256x384. We first extract the feature information from the texton map which is produced by the textons. It is a set of groups that clustered by MR8 filterbank response of all the training images with K-means. And we fix the number of K to 30 in this experiment. The MR8 filterbank [2] is consist of 38 filters which are 2 anisotropic filters, which have 3 different scales and 6 different orientations of an edge and a bar filter, and 2 rotational symmetric filters (a Gaussian and a LoG). As we set the scale to 3, so each patch going to be a 90 dimensional feature vector. And we have to train the object boundary using BEL algorithm [5] to make object adaptive patch.

By the linear logistic regression, we can get the initial parameters for the model and put them to the equation (8) to estimate the optimal parameters.

We compare our method with the results from the 8x8 fixed grid patch experiment. The figure 3 shows that the object adaptive patch improves the detection result and more close to the object.

The difference in the result mainly comes from the patch scheme that used in extracting the object and contextual feature information. In case of the fixed grid scheme, due to the variable object's size, one patch might include the whole object or include the just part of it. So the feature information from the patches are unstable (not constant). However, the adaptive patch scheme makes it possible to extract features more stable.

5. Conclusion

We have presented a method that using the multi-scale DRF framework for the object detection using a texton histogram and object adaptive patch. The experiment results shows that our method can be applied to objects detection in the satellite imagery. However, there're exist false positive result. So we need to apply more information, such as object shape and color.

6. Acknowledgement

This research was supported by the Defense Acquisition Program Administration and Agency for Defense Development, Korea, through the Image Information Research Center at Korea Advanced Institute of Science & Technology under the contract UD070007AD.

7. References

[1] Sanjiv Kumar and Martial Herbert , "Discriminative Random Fields: A Discriminative Framework for Contextual Interaction in Classification" In Proceedings of the IEEE International Conference on Computer Vision, 2003.

[2] Varma, M. and Zisserman, A., “A statistical approach to texture classification from single images”, International Journal of Computer Vision 2005: Special Issue on Texture Analysis and Synthesis.

[3] J. Shotton, J. Winn, C. Rother, A. Criminisi., “TextonBoost: Joint Appearance, Shape and Context Modeling for Multi-Class Object Recognition and Segmentation.”, In Proc. ECCV 2006

[4] Xuming He, Richard S. Zemel, Miguel A.Carreira-Perpinan, "Multiscale Conditional Random Fields for Imaged Labeling In IEEE Conference on Computer Vision and Pattern Recognition, 2004

[5] Pitor Dollar, Zhuowen Tu, Serge Belongie, "Supervised Learning of Edges and Object Boundaries," In IEEE Conference on Computer Vision and Pattern Recognition, 2006.

[6] Lin Yang, Peter Meer, David J. Foran, "Multiple Class Segmentation Using A Unified Framework over Mean-Shift Patches," In IEEE Conference on Computer Vision and Pattern Recognition, 2006.

[7] S. Vishwanathan, N. Schraudolph, M. Schmidt, K. Murphy, “Accelerated Training of Conditional Random Fields with Stochastic Gradient Methods,” International Conference on Machine Learning, 2006.

[8] Kumar, N.P., Torr, P.H.S. and Zisserman, A., "OBJ CUT," In IEEE Conference on Computer Vision and Pattern Recognition, 2005.

[9] Moore, A., Warrell, J., Mohammed, U., Prince, S. J. D. “Superpixel Lattices”, IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[10] Jan-Mark Geusebroek, Arnold W. M. Smeulders, “Fast Anisotropic Gauss Filtering”, IEEE Trans. on Image Processing, 2003.