QUERY EXPANSION FOR VHR IMAGE DETECTION

Zheng Huaxin, Zhang Huigang, Bai Xiao, and Zhao Huijie

Beihang University
37 Xueyuan Road, Haidian District, Beijing, China, 100191.
E-mail:zhenghx007@gmail.com

ABSTRACT

In order to detect the objects of interest, many different approaches have been proposed. One kind of popular approaches are based on template matching, which use a template of the object class to match the image at different positions. The matching can be computed using similarity measures such as the correlation coefficient. These approaches, although easy and robust, has the limitation of not containing to much variability of the object class, especially for shape information. Despite the statistical variation in each kind of object, collecting enough training samples is another problem which is time consuming. Inspire by template matching and incremental learning, a new object-oriented object detection methodology for very high resolution remote sensing images is proposed in this paper. We obtain the first initial query results via the bag-of-visual-words method. Then we introduce two query expansion baseline expansion and PAS expansion to obtain a new incremental model for re-query. In the experiment part, we compare and evaluate the performance of our proposed methods.

Index Terms— Object detection, VHR, Query expansion, PAS

1. INTRODUCTION

The very high resolution(VHR) satellite images provide abundant information to researchers. And it became possible to detect different kinds of man-made or natural objects with the advent of the VHR satellite imagery i.e. Ikonos and Quickbird [1]. Extracting man-made objects in VHR satellite images may help researchers in various aspects, such as automatic map making and marking. Therefore, there requires robust detection techniques for man-made objects. Many VHR satellite and aerial images object-based detection methods have been introduced. Y.Li et al. [1] used a fuzzy segmentation method to extract manmade objects from high resolution satellite imagery. Mueller et al. [2] proposed an edge and region-based segmentation technique for extraction of large, man-made objects in high-resolution satellite imagery. Jin et al. [3] combined structural, contextual, and spectral information to detect buildings in urban area. Benediktsson et al. [4] used mathematical morphological operations to extract structural information to detect the urban area in satellite images. Sirmacek and Unsalan [5] associated scale-invariant feature transform (SIFT) with graph theory to detect urban areas and buildings in grayscale Ikonos images. Besides, they used Gabor filters to extract spatial building characteristics (such as edges and corners) in different orientations in [6]. Michel et al. [7] incorporated multi-scale segmentation and spatial reasoning graphs for object detection in remote sensing images. Most exiting detection methods are based on constructing statistical models from VHR image. However, current methods are hard to satisfy the VHR image detection requirements in accuracy and efficiency. Despite the statistical variation in each kind of object, collecting enough training samples is another problem which is time consuming.

To address these problems, we propose a statistical incremental method for VHR image object detection. Rather than simply constructing the model from few training set or query images, we incrementally enrich it with the additional information from the testing image by using ideas from incremental learning. We start by getting a quick initial query by using bag-of-visual-words [8] [9]. This initial detection results may contain inaccuracy results. To get a refined or incremental results, we apply query expansion to refine the model from the initial results. Two query expansion strategies have been introduced which depends on the VHR image information.

Among the two expansions, we construct the shape model by PAS (Pair of Adjacent Segments) [10] which is a robust shape descriptor proposed recently. After the query expansion, we can construct object models for detection. In this paper, we proposed a robust and efficient object detection method for VHR image which only need few training or query samples.

The proposed framework has been tested for high spatial resolution images from commercial satellites, such as QuickBird. The paper is structured as follows. Section 2 presents the general object detection scheme (Section 2.1)and two query expansion schemes (Section 2.2). Section 3 presents the experimental results for detection including evaluation(Section 3.1) and performance(Section 3.2). Finally, conclusions are outlined in Section 4.

2. METHOLOGY

Given one or few query images of an object, our objective is to retrieve the same kind of instances in a VHR satellite image. Rather than using the the query image to model the object, we using an incremental learning to get more information and refine our model from the test image. Hence, we need to get the effective information and abandon the useless ones.

2.1. Object detection

We first apply a bag-of-visual-words retrieval to get initial search results:

Segmentation Image patches are extracted at different scales compared with the query image using sliding-window. We use a sliding-window with the size of query image to segment the test image into small patches. Besides, we obtain patches of different scales using window with 0.5 and 1.5 times of the size of query image.

Feature extraction For each patch, a set of local regions which are stable and affine invariant over different scales are extracted using the MSER(maximally stable extremal regions, [11]) detector. In addition, for each of these affine regions, we compute a 128-dimensional SIFT descriptor [12].

Quantization A visual vocabulary is constructed using clustering algorithm i.e. k-means. Each visual descriptor is assigned, via nearest neighbor search, to a single cluster center, which giving a standard bag-of-visual-words model. These quantized visual features are then used to index the patches for the search. Hence, each patch is described by a sparse vector.

Search The query image and each patch from the test images is represented as a sparse vector of term (visual word) occurrences and search then proceeds by calculating the similarity between the query vector and each patch vector. We use the standard tf-idf weighting scheme [13], which down-weights the contribution that commonly occurring to the relevance score. Sort the patches by the similarity as the first query result.

Although simple, the initial query results from previous step may not accurate due to the noisy and variation within each image patch. In addition, the bag-of-visual-word have ignored the spatial configurations between features. Therefore, we explore query expansion from the initial results.

2.2. Query expansion

In this section we describe several methods for computing new models as query expansion. Each method starts by evaluating the original query Q_0. A new model is then constructed from the verified images returned from Q_0, and new query Q_1 is given from this model.

Query expansion baseline: This method is a straight forward naive application of query expansion which has been used in text-retrieval. Take the top m results from the original query Q_0, average the term-frequency vectors computed from the entire result patches and re-query once. The new query could be represent as:

$$d_{avg} = \frac{1}{m}\sum_{i=1}^{m} d_i \tag{1}$$

where d_i is the normalized tf vector of the i-th result. The new query results of Q_1 are appended to initial results(the top m).

Shape expansion: We first take query image and the top m images from the original query Q_0 as training images. Then we extract edges with the excellent Berkeley edge detector and to chain them. The extracted PAS (Pair of Adjacent Segments) features is given in figure 1. The PAS dissimilarity measure is given in Equation (2) which is used to form a PAS codebook.

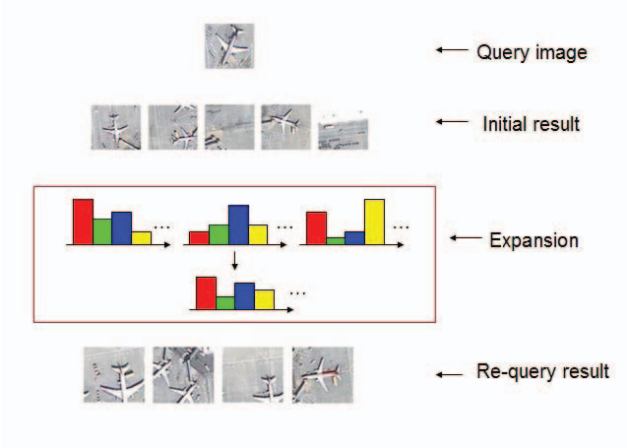
$$D(d^p, d^q) = w_r||r^p - r^q|| + w_\theta \sum_{i=1}^{2} D_\theta(\theta_i^p, \theta_i^q) + \sum_{i=1}^{2} |log(l_i^p, l_i^q)| \tag{2}$$

Here the d is the PAS descriptor $d = (\theta_1, \theta_2, l_1, l_2, r)$ which encodes the shape of the PAS, by the segments orientations θ_1, θ_2, lengths l_1, l_2, and the relative location vector r, going from the center of the first segment to the center of the second. The weights w_r,w_θ are fixed to the same values ($w_r = 4$, $w_\theta = 2$). We then determine model parts as PAS frequently reoccurring with similar locations, scales, and shapes(with the help of codebook). An initial shape is assembled by selecting a particular PAS for each model part from the training examples. To learn a shape model, we we iteratively matching the initial model back onto the training images. This model is used to re-query once.

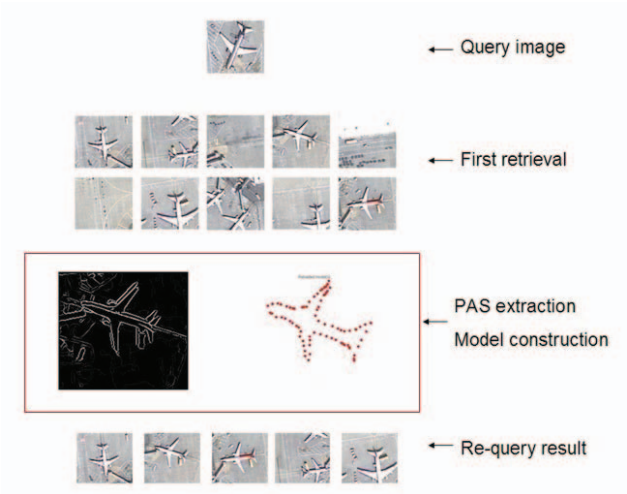
Fig. 1. Example of PAS.

The shape model we chosen has several attractive properties. First of all, we build an explicit shape model formed by continuous connected curves. This allows us to model explicitly different objects by unified representation in the VHR satellite images. The learning method avoids the pairwise image matching used in previous approaches [14] [15], therefore can computationally cheaper and robuster to clutter edges. Moreover, we model intra-class deformations and enforce them at test time, which is adaptive to the different

size and deformation of object in remote sensing images. We show the main processes of our method in Figure 2.



(a) Query expansion baseline.

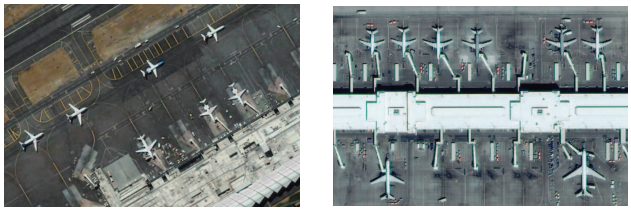


(b) Shape expansion.

Fig. 2. Process of the expansions.

3. EXPERIMENT

The proposed query expansion algorithm has been tested on two types of high-resolution satellite images, including 0.6 m QuickBird and 0.5 m Geoeye-1 image data (see Figure 3). Each time we give two images as query images and the results are the average of twenty trails.

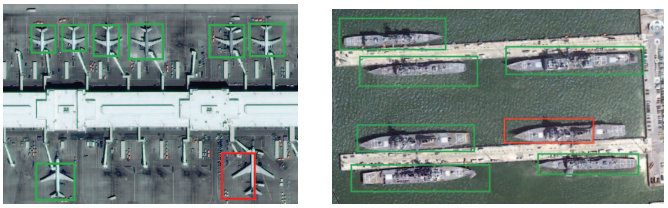


(a) Quickbird (b) Geoeye-1

Fig. 3. The test images.

3.1. Quantitative Evaluation

For quantitative evaluation, we manually label the objects appearing in all testing images as ground truth, the total number of which is NP. The correct detections are added to true positives(TP), and so spurious detections of the same object are counted as false positives(FP). The recall precision curve is also used to plot the tradeoff between recall and precision, where $Recall = TP/NP$, and $Precision = TP/(TP + FP)$. The area under the recall and precision curve takes the whole curve into account and so gives a better measure for comparison purpose. The results is given in Table 1.



(a) plane test results (b) ship test results

(c) stadium test results

Fig. 4. Experimental results of different test objects.

3.2. Performance

The different object detection results have been shown in the Figure 4. Even in some tough cases such as objects rotated in different angels, boundary regions being blurred with background noises, we observe that our method can achieve good performance. The different query expansion quantitative results have been shown in the Table 1.

Recall %	Original	Query Expansion Baseline	Shape Expansion
airplanes	74.5%	85.0%	93.2%
ships	71.7%	82.5%	94.4%
submarines	69.8%	77.5%	85.6%
stadiums	78.6%	87.5%	88.4%

Table 1. Target detection results by our methods.

4. CONCLUSION

To conclude, a new method for VHR target detection is presented. With the additional information from initial result, we form a more robust model for query target. Since this method merely uses a target query image without any other information, it exhibits reliable results for different target detection. In the future, we plan to incorporate more complex incremental learning algorithms in our work. We also plan to test graph based methods for object detection.

5. ACKNOWLEDGEMENT

This work is supported Fundamental Research Funds for the Cental Universities and Open Projects of National Laboratory of Pattern Recognition(201001102).

6. REFERENCES

[1] Y. Li, J. Li, and M. Chapman, "Automatically Extracting Manmade Objects from Pan-Sharpened High-Resolution Satellite Imagery Using a Fuzzy Segmentation Method," *Geo-information for Disaster Management*, pp. 641–653, 2005.

[2] M. Mueller, K. Segl, and H. Kaufmann, "Edge-and region-based segmentation technique for the extraction of large, man-made objects in high-resolution satellite imagery," *Pattern Recognition*, vol. 37, no. 8, pp. 1619–1628, 2004.

[3] X. Jin and C.H. Davis, "Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information," *EURASIP Journal on Applied Signal Processing*, pp. 2196–2206, 2005.

[4] J.A. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 41, no. 9, pp. 1940–1949, 2003.

[5] B. Sirmacek and C. Unsalan, "Urban-area and building detection using SIFT keypoints and graph theory," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 4, pp. 1156–1167, 2009.

[6] B. Sirmacek and C. Unsalan, "Urban area detection using local feature points and spatial voting," *Geoscience and Remote Sensing Letters, IEEE*, vol. 7, no. 1, pp. 146–150, 2010.

[7] J. Michel and J. Inglada, "Multi-scale segmentation and optimized computation of spatial reasoning graphs for object detection in remote sensing images," in *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*. IEEE, vol. 3, pp. III–431.

[8] L. Weizman and J. Goldberger, "Urban-Area Segmentation Using Visual Words," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 3, pp. 388–392, 2009.

[9] S. Xu, T. Fang, D. Li, and S. Wang, "Object Classification of Aerial Images With Bag-of-Visual Words," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 2, pp. 366–370, 2010.

[10] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid, "Groups of adjacent contour segments for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 36–51, 2007.

[11] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.

[12] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[13] R. Baeza-Yates, B. Ribeiro-Neto, et al., *Modern information retrieval*, vol. 463, ACM press New York., 1999.

[14] M. Brown, R. Szeliski, and S. Winder, "Multi-image matching using multi-scale oriented patches," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, vol. 1, pp. 510–517.

[15] B. Xiao, E.R. Hancock, and R.C. Wilson, "A generative model for graph matching and embedding," *Computer Vision and Image Understanding*, vol. 113, no. 7, pp. 777–789, 2009.