SINGLE-SAMPLE AEROPLANE DETECTION IN HIGH-RESOLUTION OPTIMAL REMOTE SENSING IMAGERY

Bin Pan[1], Liming Wang[2], Xinran Yu[3] and Zhenwei Shi[1]

[1]Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China
[2]State Key Laboratory of Information Security, Institute of Information Engineering Chinese Academy of Sciences, Beijing 100093, China
[3]The 28th Research Institute of China Electronics Technology Group, Nanjing 210000, China
{panbin,shizhenwei}@buaa.edu.cn;

ABSTRACT

In remote sensing images, detecting aeroplanes of special shapes is difficult due to limited number of samples. Without enough training samples, most supervised learning based algorithms will fail. Focusing on the specially-shaped aeroplanes in high-resolution optical remote sensing imagery, this paper presents a single-sample approach. The proposed approach takes one sample as input and directly searches for similar matches from the image. Unlike the supervised learning algorithms which extracts information from positive and negative samples, the hyperspectral algorithm estimates the statistics of background by analyzing the global information of the target image, needless to provide negative samples. Furthermore, this algorithm tries to find a hyperplane projected on which the background is compressed while the target is preserved, making it more data-adaptive than the conventional similarity measurements. Experiments on real data have presented the robustness of the proposed method.

***Index Terms*—** Aeroplane detection, locally adaptive regression kernels, constrained energy minimization

1. INTRODUCTION

Aeroplane detection is of great significance owing to its wide range of applications such as airport management, military reconnaissance and accident surveillance [1]. In particular, with the increasing demand for designing practical systems, detecting aeroplane in high resolution optical imagery has attracted much attention.

Existing aeroplane detectors are usually designed based on the training/learning framework which necessitates a large number of manually labeled training samples (including positive and negative samples), *i.e.*, transforming the detection problem to an equivalent classification problem. However, in some cases, the number of positive samples may be very limited. Single-sample detection, *i.e.*, detecting targets using only one positive sample, can handle the specially-shaped planes with only a small number of samples. Besides, this kind of algorithms can also avoid the practical problems caused by the unbalanced proportion between positive and negative samples. Briefly, single-sample detection is a kind of image-to-patch matching problem so the feature descriptor and the similarity measurement are two keys of the system. Single-sample detection algorithms also have many other applications such as [2–4].

However, single-sample detection in remote sensing imagery lacks deep research. Compared with visual images, remote sensing images are much larger in size and contain more complicated background, which could produce more false detections. On the other hand, the targets in the remote sensing images are relatively small and the detailed information for the target is much less. In consequence, detecting targets in remote sensing images has much more difficulties than visual images.

In this paper, we investigate the problem of single-sample detection in high resolution remote sensing imagery and propose a new method to detect aeroplanes of special shapes. Locally adaptive regression kernels (LARK) [5] of the sample (query image) is extracted to generate the feature vector. Then the LARK features are extracted patch by patch from the image under detection (target image). In this way, each patch in the target image is represented by a feature vector, generating a three-dimensional data cube. This data cube has the same form as the hyperspectral image while the only difference lies in that the spectral vectors are replaced by the feature vectors. Motivated by this idea, the data cube is viewed in a hyperspectral perspective [6] and a hyperspectral algorithm is employed to measure the similarity between the query image and the unlabeled patches. The proposed algorithm aims at finding the optimal hyperplane which is capable of compressing the background while preserving the target. More

Thanks to National Key R&D Program of China under the Grant 2017YFC1405600, the National Natural Science Foundation of China under the Grants 61671037, and the Excellence Foundation of BUAA for PhD Students under Grant 2017057. *Corresponding author: Liming Wang (wangliming@iie.ac.cn) and Zhenwei Shi.*

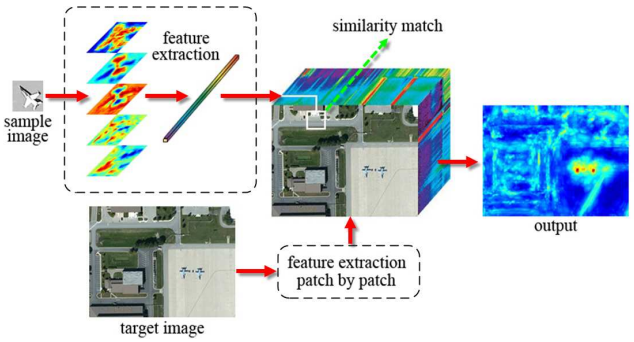


Fig. 1: The outline of the proposed method.

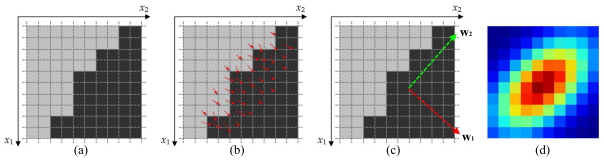


Fig. 2: The detailed explain of LARK. (a) is a simulated 10×10 local window containing an edge pattern. (b) shows the gradient vector of each pixel where the oriented and scaled arrows denote the gradient vectors. (c) shows the principal components of the gradient vectors where $\mathbf{w}_1$ denotes the first principal direction and $\mathbf{w}_2$ denotes the second. (d) is the computed LARK.

importantly, this algorithm takes into account both the local and global information of the data. Accordingly, it is more data-adaptive than the conventional similarity measurements.

The proposed approach has two major contributions. First, we provide theoretically analysis about the superiority of LARK, when compared with some other popular features [7–11]. Second, a hyperspectral algorithm is applied to measure the similarity between feature vectors.

2. LARK BASED FEATURE EXTRACTION

As outlined in the previous section, our approach to detect aeroplanes mainly consists of two stages, *i.e.*, locally adaptive regression kernel (LARK) feature extraction and similarity match. In this section, we will describe the first step in detail.

LARK is extracted based on the idea that the dominant structures or shapes of an image can be robustly described by a group of similar pixels. The similarity between two pixels is not only determined by the pixel value, it is also related to the spatial distance and the gradient information. The LARKs (or similarities) between a pixel $\mathbf{x}$ and its $P \times P$ neighbors $\mathbf{x}_l (l = 1, ..., P^2)$ are computed as follows:

$$Kernel(\mathbf{x}_l - \mathbf{x}; \mathbf{H}_l) = \frac{K(\mathbf{H}_l^{-1}(\mathbf{x}_l - \mathbf{x}))}{\det(\mathbf{H}_l)}, l = 1, ..., P^2, \tag{1}$$

where $K(.)$ is a radially symmetric function (e.g., the Gaussian function), $\mathbf{x}_l = [x, y]^{\mathrm{T}}$ is the spatial coordinates, P^2 is the total number of pixels in the $P \times P$ window, $\mathbf{H}_l$ is a 2×2 matrix named as *steering matrix* that contains the spatial information of the local window around $\mathbf{x}_l$. $\mathbf{H}_l$ is defined by

$$\mathbf{H}_l = h\mathbf{C}_l^{-\frac{1}{2}} \in \mathbb{R}^{2\times 2}, \tag{2}$$

where h is a global smoothing parameter (scalar) and $\mathbf{C}_l$ is the covariance matrix estimated from the spatial gradient vectors within the local window around $\mathbf{x}_l$.

Given a $N \times N$ sample image, it is first cut into a number of patches of the $k \times k$ size and the total number of the patches is $(N/k) \times (N/k)$. LARKs are then extracted patch by patch with each patch corresponds to a $k^2 \times 1$ dimension feature vector. After computing the LARKs for all patches, a collection of $k^2 \times 1$ feature vectors is obtained and principal component analysis (PCA) is applied to the data to remove redundant information. After PCA, the $k^2 \times 1$ feature vectors are reduced to $n \times 1$ feature vectors where n is the first n principal components. Finally, the $n \times 1$ feature vectors are combined together to form the final LARK feature descriptor.

The above description indicates that the sample image is firstly cut into non-overlapping patches. In fact, the non-overlapping patches are simply for ease of description. Instead, we use the overlapping patches in our approach to avoid the loss of information and the overlapping region is half of the patch. In other words, the LARKs are computed within a $k \times k$ sliding window whose step size is $k/2$.

In our approach, the target image also need to extract the LARK features. Given an $H \times W$ target image, the LARK features are extracted in a $N \times N$ sliding window with one-pixel step size. After the window running through the whole image, each pixel will correspond to a feature vector and the whole image will correspond to a three-dimensional data cube, which is named as *feature-spectrum data* in this paper. In fact, the obtained data cube and the hyperspectral image have the same data form: one is made up of the feature vectors while the other is made up of the spectral vectors. Motivated by the similarity in data form, we utilize a hyperspectral algorithm to analyze the feature-spectrum data cube and to search similar matches for the query image.

3. CEM BASED DETECTION

After extracting the LARK features, a three-dimensional data cube is generated. Detecting planes from the target image is equal to searching similar matches for the query feature vector from the data cube. In this paper, constraint energy minimization (CEM) with regularization term [12], a hyperspectral object detection algorithm, is used to measure the similarity be-

tween samples.

Given a query vector $\mathbf{d}$ and a collection of background vectors $\mathbf{S} = \{\mathbf{r}_1, ..., \mathbf{r}_N\}$, where N denotes the number of the feature vectors to be compared, CEM aims at finding a projection direction projected on which the output of background is compressed while target of interest is preserved. Specifically, CEM tries to find a direction $\mathbf{w}$ which keeps the the projection of $\mathbf{d}$ a constraint (e.g. 1):

$$\mathbf{d}^T\mathbf{w} = 1, \tag{3}$$

After adding a regularization term, the Lagrange function is given by:

$$E(\mathbf{w}, \lambda, \beta) = \mathbf{w}^T\mathbf{R}\mathbf{w} + \lambda(\mathbf{w}^T\mathbf{d} - 1) + \beta\mathbf{w}^T\mathbf{w}. \tag{4}$$

and $\mathbf{R}$ is the covariance matrix. Differentiating Eq. (4) with respect to $\mathbf{w}$ gives

$$\frac{\partial E}{\partial \mathbf{w}} = 2\mathbf{R}\mathbf{w} + \lambda\mathbf{d} + 2\beta\mathbf{w}, \tag{5}$$

and setting it to zero, $\mathbf{w}$ can be expressed as:

$$\mathbf{w} = -\frac{\lambda}{2}(\mathbf{R} + \beta\mathbf{I})^{-1}. \tag{6}$$

Using the constraint $\mathbf{w}^T\mathbf{d} = 1$ gives:

$$-\frac{\lambda}{2} = \frac{1}{\mathbf{d}^T(\mathbf{R} + \beta\mathbf{I})^{-1}\mathbf{d}}. \tag{7}$$

Substituting Eq. (7) into Eq. (6), the regularized CEM is given by

$$\mathbf{w}^* = \frac{(\mathbf{R} + \beta\mathbf{I})^{-1}\mathbf{d}}{\mathbf{d}^T(\mathbf{R} + \beta\mathbf{I})^{-1}\mathbf{d}}. \tag{8}$$

After applying the constraint energy minimization (CEM) algorithm, we need to normalize the results to a certain range (e.g., [0,1]) and choose a threshold to segment the detected targets. However, directly applying normalization will be unreasonable, since there could be no targets of interest present in the image. Therefore, before normalization, two kinds of tests need to be judged, i.e., (i) whether there are any targets present in the image, (ii) if there do exist the targets, where are they.

The first test is judged by analyzing the query vector $\mathbf{d}$, the projection hyperplane $\mathbf{w}^*$ and the feature vectors $\mathbf{r}_i, (i = 1, .., n)$. Specifically, assume that the projection of $\mathbf{d}$ on $\mathbf{w}^*$ is p ($p = 1$), the angle between them is α, then we check all the vectors $\mathbf{r}_i, (i = 1, .., n)$. If there exist one vector $\mathbf{r}_i$ whose projection on $\mathbf{w}^*$ belongs to $[0.5p, 1.5p]$ while whose angle to $\mathbf{w}^*$ belongs to $[0.5\alpha, 1.5\alpha]$, then we assert that there exists at least one target, otherwise, there is no target. The second test is judged by a threshold which is set empirically. Suppose that m is the maximum of the detection result and $0.85m$ is set to be the threshold separating the targets out of background. Afterwards, the connected regions in the thresholded image are also analyzed so that the regions with too small or too large sizes will be eliminated.

To detect targets of different orientations, the sample image is rotated for 15 times and each time it is rotated by 22.5 degree. To detect targets of different sizes, the target image is down-sampled for 3 times and the sample rate is 1.0, 0.6 and 0.3. For the outputs of different orientations and scales, their maximums are set to be the final results.

4. EXPERIMENTS

4.1. Dataset

The dataset contains 50 panchromatic remote sensing images cut from the Google-Earth software. The sizes of the images are nearly 5000×5000 and the resolutions range from 0.5m to 1m. The images are all airport scenes of which 10 images contain specially-shaped aeroplanes while the other 40 images contain general aeroplanes. The aeroplanes are of different orientations and different shapes and their sizes range from 30×30 to 100×100.

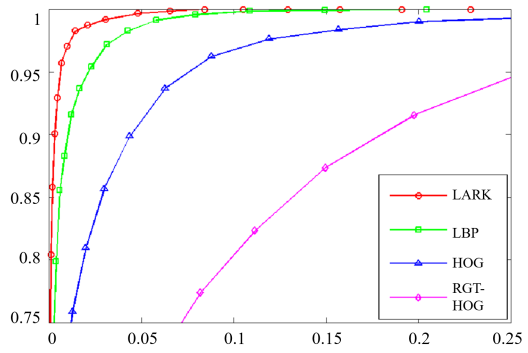
4.2. Quantitative Analysis of the Proposed Approach

In this section, we validate the superiorities of LARK and CEM, respectively. Fig. 3 illustrates the ROC curves of different methods. LARK is compared with the local binary pattern (LBP) [7], histograms of oriented gradient (HOG) [8] [9] [10] and radial gradient transform based HOG (RGT-HOG) [11]. For LARK, LBP and HOG, which are not invariant to the orientation, the sample image is rotated for several times to detect the aeroplanes of different orientation. However, for RGT-HOG, the sample image do not need to rotate because of its rotational invariance. The ROC curves of the detection results are presented in Fig. 3(a) which indicates that LARK outperforms other features in aeroplane detection.

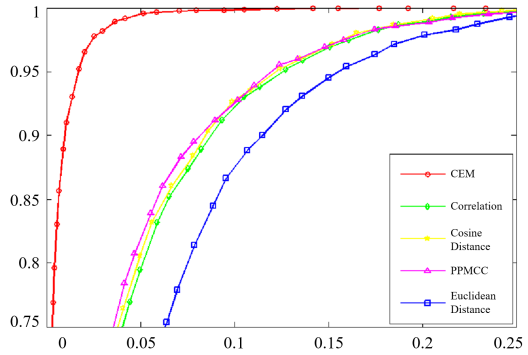
To further study the performance of CEM, the conventional similarity measurements including the Euclidean distance, Correlation, Cosine distance and Pearson product-moment correlation coefficient (PPMCC) are compared, as is shown in Fig. 3(b). It can be seen that CEM presents certain advantage. The Euclidean distance tends to produce missing detections while Correlation, Cosine distance and PPMCC are liable to generate much more false alarms.

5. CONCLUSION

In this paper, we investigate the problem of single-sample aeroplane detection in high resolution optical panchromatic imagery. Firstly, LARKs are extracted from the query and the target image, generating a three-dimensional data cube in the hyperspectral form. To measure the similarity between the



(a)



(b)

Fig. 3: The ROC curves different methods. (a) Comparison of features. (b) Comparison of similarity measurements. The X-axis denotes the false positive rate and the Y-axis denotes the true positive rate.

computed feature vectors, the feature vectors are treated in a hyperspectral perspective and the hyperspectral detection algorithm CEM is employed. Unlike the conventional similarity measurements, CEM tries to find a direction projected on which the output of background is compressed while objects of interest (*i.e.*, the aeroplanes) are preserved. This algorithm analyzes the global information of the target image, including the statistics of background. Therefore, the algorithm takes the background information into consideration even though only one query image is utilized beforehand, which makes it more data-adaptive than the conventional similarity measurements. The experimental results on real data indicate that the LARK feature is better than other features in the single-sample aeroplane detection and the CEM algorithm also outperforms the conventional similarity measurements.

6. REFERENCES

[1] Y. Yang, Y. Zhuang, F. Bi, H. Shi, and Y. Xie, "M-fcn: Effective fully convolutional network-based airplane detection framework," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 8, pp. 1293–1297, Aug 2017.

[2] H. J. Seo and P Milanfar, "Training-free, generic object detection using locally adaptive regression kernels," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1688–1704, 2010.

[3] B. Zhang, S. Luan, C. Chen, J. Han, W. Wang, A. Perina, and L. Shao, "Latent constrained correlation filter," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1038–1048, March 2018.

[4] A. Tejani, R. Kouskouridas, A. Doumanoglou, D. Tang, and T. K. Kim, "Latent-class hough forests for 6 DoF object pose estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp. 119–132, Jan 2018.

[5] H Takeda, S Farsiu, and P Milanfar, "Kernel regression for image processing and reconstruction.," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 16, no. 2, pp. 349–366, 2007.

[6] Bin Pan, Zhenwei Shi, and Xia Xu, "MugNet: Deep learning for hyperspectral image classification using limited samples," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2017.

[7] G. Heusch, Y. Rodriguez, and S. Marcel, "Local binary patterns as an image preprocessing for face authentication," in *International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 6–14.

[8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, June 2005, vol. 1, pp. 886–893 vol. 1.

[9] I. M. Creusen, R. G. J. Wijnhoven, E. Herbschleb, and P. H. N. De With, "Color exploitation in hog-based traffic sign detection," in *IEEE International Conference on Image Processing*, 2010, pp. 2669–2672.

[10] Yanwei Pang, Yuan Yuan, Xuelong Li, and Jing Pan, "Efficient hog human detection," *Signal Processing*, vol. 91, no. 4, pp. 773–781, 2011.

[11] Liu Liu and Zhenwei Shi, "Airplane detection based on rotation invariant and sparse coding in remote sensing images," *Optik - International Journal for Light and Electron Optics*, vol. 125, no. 18, pp. 5327–5333, 2014.

[12] Nasser M. Nasrabadi, "Regularization for spectral matched filter and rx anomaly detector," *Proc Spie*, vol. 6966, 2008.