Aircraft Detection in Aerial Image Based on BOVW of Affine Invariance Features

Feng'an Zhao*, Xiaodong Mu, Peng Zhao and Zhou Yang
Department of Computer Science and Engineering, Xi’an Research Institute of High-Tech, Xi’an, China

*Corresponding author e-mail: zhao_flying123@163.com

Abstract. Aircraft detection is one of the applications of high resolution remote-sensing images. Conventional method of aircraft detection use deep and complicated network, which needs highly computing cost. In this letter, an efficient and effective aircraft detection framework based on BOVW features and cascade AdaBoost classifier has been proposed. A variety of affine invariant features, which represent the complicate structure of aircraft, are extracted from sliding windows, and the direction estimation of aircraft is introduced to align the aircraft before detection. Besides, an accurate NMS algorithm is designed to make the target location more accurate. The proposed method is evaluated with recent advanced detection methods. The superior experimental outcome indicates that our framework achieves better accuracy as well as efficiency.

1. Introduction

Automatic interpretation of remote-sensing image aims at analysing the remote-sensing image by computer, and then output the semantic description of the corresponding region in the scene. In last decade, with the development of high-resolution remote-sensing (HRRS) technology and its sensors, researchers can obtain HRRS images with rich details and spatial information. So it provides a new opportunity to promote the automatic interpretation of HRRS images. Target detection of HRRS image is one of the tasks of automatic interpretation. Its purpose is to analyse whether the objects of interest contain in an HRRS image. If so, to locate the target in the image is required. Target detection has wide applications in military and economic fields. This paper is focus on aircraft detection, which is widely used in image reconnaissance, condition monitoring, and other military or civil fields. As the aircraft in the HRRS image usually has different directions parallel to the ground, and has a variety of scales and colours, it is vulnerable to be disturbed from the surrounding terrain, which makes aircraft detection a challenging problem.

Traditional solutions regard target detection as binary classification problem, namely to identify the target from the background. The common target detection methods can be divided into 4 categories, template-based matching models, knowledge-based models, object-based image analysis (OBIA) models and machine-learning based models [1]. Template-based matching models include two steps: building template followed by similarity measurement. Main disadvantage of template matching model is that the template should be very precise, so the result of target detection is sensitive to shape or rotation changes of target. Knowledge-based target detection model transforms the problem of target detection into a hypothesis testing problem, which uses some prior knowledge such as geometric information and context information. Object detection method based on OBIA divides the
HRRS image into relatively uniform regions by choosing appropriate scale, shape and compactness criteria, and then classifies these regions. Different segmentation scales have a great influence on the results, so the suitable segmentation scale is important in OBIA methods. Machine-learning based methods dominate the field of aircraft detecting. Researchers usually extract the key feature of the target with machine-learning methods, fuse them or reduce the dimension if needed, and input them to the classifier for training and testing. In the feature extraction stage, handcraft features like SIFT or gradient histogram (HoG) are extracted. These features are suitable for target detection in conventional optical images, however the targets in remote sensing images are sensitive to background and illumination changes. Many methods are proposed to solve the problem, [2] use circular frequency filters to locate regions of interest (RoI) and extracts HOG features. Finally these features are used for classifying these RoIs. In order to adjust the leading direction, [3] propose a novel rotation invariant feature called polar normalized directional gradient histogram. [4] use aggregated channel feature (ACF) to describe the aircraft in RoI, which has more abundant representation and speeds up the calculation. However, the parameters of handcraft features need to be carefully adjusted, and detection results are vulnerable to the various scales and rotation of targets. In these years, deep model, such as convolutional neural network (CNN) has been introduced into aircraft detection due to its powerful feature representation ability. However, deep learning methods usually depend heavily on labelled samples [5], which cannot be obtained abundantly. Besides, the high computational cost also hindered the development of quickly detection. So handcraft feature still has great potential for aircraft detection.

At present, most handcraft methods use contour information to detect targets, but contour information is not easy to obtain under complex background. At the same time, the existing handcraft features are sensitive to rotation and scale, which have a great impact on the detection results. Based on the above considerations, in this work, a novel target detection method based on affine invariance is proposed. Specifically, in each sliding windows, a variety of affine invariant features are calculated. The aircraft direction estimation is introduced to align the aircraft before detection, which makes the method rotate and scale invariant. In addition, an accurate NMS algorithm is designed to make the target location more accurate. The performance of our method is evaluated by detecting aircraft targets in optical HRRS images. Our obtained experimental results indicate the superior accuracy and efficiency of the proposed method.

The proposed structure has three advantages over the other detecting methods. Firstly, a variant of affine invariant feature is adopted to form the features rapidly. Secondly, an effective and efficient aircraft automatic detection model is proposed. The model comprise three components: (1) a novel affine invariant feature extractor; (2) a cascade AdaBoost classifier trained as detector and (3) post processing component based on non-maximum suppression (NMS). Thirdly, we proposed a novel NMS method with a bounding-box regression algorithm which effectively solves the problem of too many bounding boxes and the location of bounding box is further optimized.

2. Method Description

![Flowchart of the proposed framework.](image-url)
An effective and efficient aircraft detection model is proposed, Figure 1 shows the detail. We use sliding-windows with fix step to split the HRRS image into chips. Affine invariant interest points are extracted. SIFT descriptors of the elliptical regions of the interest points are extracted and encoded into BOVW representation. Then, an AdaBoost algorithm is adopted for classifying the chip into aircraft target or non-aircraft target. Lastly, an effective and accurate NMS method is proposed to fuse the outputs.

2.1. Feature Description
In order to represent the aircraft with affine invariant points, we need to find out the interest points. Due to the fact that aircraft structure contains corner-like or blob-like parts, which are different from other targets or background. Here, Hessian-Laplace and Harris-Laplace are adopted to detect the interest point. The Hessian-Laplace and Harris-Laplace which detects blob-like and corner-like structures respectively can be implemented by VLFeat [6] toolbox. Then the location and neighbourhood of each interest point are modified and converges to affine invariant point[7]. In order to represent each invariant region, each elliptical region is transformed into a circular region. Then the canonical directions are found based on an orientation histogram formed on the image gradients as SIFT [8] feature extraction. The dimension of the SIFT feature is represented by $d_s$, which is a constant value of 128. Let $N_b$ represent the number of sub patches sampled from the image. The structure features of the input HRRS image can be written as $f^i = (f^i_1, f^i_2, ..., f^i_{N_b}) \in \mathbb{R}^{d_s \times N_b}$. In BOVW coding stage, the structural features of each scene image are encoded respectively by vector quantization [9]. The vector quantization solves the following problem of constrained least squares fitting:

$$\arg \min_{\alpha} \sum_{i=1}^{m} \| f - D \alpha_i \|^2 \quad \text{subject to} \quad \| \alpha_i \|_1 = 1, \quad \| \alpha_i \|_1 = 1, \quad \alpha_i \geq 0 \quad (1)$$

Where $D$ is the codebook, $X = (x_1, x_2, ..., x_N)$ is the coding coefficient vector for $f$, and $x_i = (\alpha_1, \alpha_2, ..., \alpha_m), \alpha_i \in \mathbb{R}^{d_s}$. We get $X^i = (x^i_1, x^i_2, ..., x^i_{N_b}), x \in \mathbb{R}^{d_s}$ as the coding vectors of structure feature. In the BOVW pooling step, summation pooling and L1 normalization are carried out to get the representation of the image feature considering both compact and robust. The pooled feature coding vector $\tilde{x}^i \in \mathbb{R}^m$ reflects the co-occurrence relationship appeared in structure features.

2.2. Training Design
After feature extraction, AdaBoost classifier [10] is adopted to classify these candidate patches. The core idea of AdaBoost is to iteratively train weak classifiers with the original training dataset. The subset of the training dataset is chosen for the classification in iteration procedure. Then these weak classifiers are combined into a strong classifier. The AdaBoost classifier provides a simple but effective approach to achieve high classification accuracy, and can avoid over-fitting. In this work, Cascade AdaBoost [11] is adopted to combine different weak classifiers in iteration procedure.

Due to the fact that there are more easy examples and less hard examples in training dataset, while hard examples are more important than easy examples. The hard example has a great influence on the detection results due to its diversity. The discrimination ability of the network can be obviously enhanced by these hard examples. This idea of data bootstrapping is also called hard example mining (HNM) [12]. In this work, HNM is adopted to iteratively train the cascade AdaBoost classifier. Firstly, negative samples whose number is equal to that of the positive samples are randomly chosen to initialize the classifier. Then, the samples which are misclassified in the last iteration are put into the original negative set. In the next iteration these misclassified samples are used to retrain the classifier. The training process ended when accuracy stops increasing and become stable.
2.3. Postprocessing
After obtaining a set of detection windows, we need to output un-oriented bounding boxes and corresponding scores. NMS is to find exactly one bounding box for each ground truth with high localization accuracy from these highly inter-overlapped bounding boxes belonging to the same object, and remove overlapping detections by non-maximum suppression (NMS) algorithm. In original NMS, each input image chip can obtain a score from the AdaBoost classifier, and the score denotes the degree of similarity of the image chip and ground truth. All of these scores are sorted in descending order, the highest scored bounding box is accepted and the other bounding boxes whose overlap rates are greater than a threshold are rejected. While the there is still an issue that the selected bounding boxes are not accurately localized yet, because it doesn’t utilize the remaining higher quality boxes, which contains quite accurate localization information. We propose an accurate regression method, which can optimize the selected bounding boxes. The enhanced regression method includes two steps: firstly, the boxes which intersect each other are grouped. Then let the \( G_n \) denotes the \( n \)th group of the boxes, and \( B_{ni} \) denotes the the \( i \)th region of \( G_n \). For each \( B_{ni} \in G_n \), \( B_{ni} = (x_{ni}, y_{ni}, x_{ni}^2, y_{ni}^2) \), and \( x_{ni}^1, y_{ni}^1, x_{ni}^2, y_{ni}^2 \) corresponding to its left, top, right and bottom location in \( G_n \). Suppose that \( W_n = (x_{ni}, y_{ni}, x_{ni}^2, y_{ni}^2) \) is our objective coordinate, then let \( d_i = W_n - B_{ni} \), we can obtain the \( W_n \) by the following equation:

\[
F(W_n) = \arg \min \sum_{i=1} c_i d_i^2
\]  

Then, the coordinate for the optimized location can be written as:

\[
x_n^e = \frac{\sum c_i x_{ni}^e}{\sum c_i}
\]  

After the end of the iteration, the number of bounding boxes decreases significantly, and the positions we acquired become more accurate.

3. Experiment and analysis
We use Python 2.7.13 as the simulation platform. Our experimental hardware environment is Lenovo K500 Gragh Workstation with Intel E5 CPU, 16GB RAM, and a GeForce K2200 GPU with 4GB RAM.

3.1. Data and Experimental setup
All images which contain aircraft were collected from HRRS images of worlds’ famous airports. The spatial resolutions ranged from 1 to 10 m. In our method, bootstrap is adopted to enhance the detection result. In the classifier training step, we segmented the 455 HRRS airport images into 637 positive samples and 835 negative samples. The resolution of these samples is from 800 × 600 to 1200 × 800. Additionally, due to the fact that aircrafts appeared in HRRS images can be at any angle, but the acquired HRRS images at hand are insufficient. To address the difficulties in data collection of aircraft, positive aircraft samples are augmented by scale and rotation transforms. This data augmentation can also increase the diversity of the HRRS images and make the framework learn rotational features sufficiently. The positive samples are rotated at every 5°, and each rotate image is added into the positive sample dataset.

3.2. Comparisons of the performance with and without HNM
To verify the detection performance, the detection precision ratio and recall ratio are adopted. The precision rate and recall rate can be denoted as following:

\[
Recall = \frac{N_{TP}}{N_{PS}}
\]
where $N_{TS}$ represents the total number of real aircrafts, $N_{DS}$ represents the number of real detected aircrafts. The number of falsely identified aircrafts are denoted as $N_{DF}$. Figure 2 shows the comparison results of the Harris-Laplace and Hessian-Laplace with and without HNM. From figure we can see that, the interest points which are selected by Hessian-Laplace and Harris-Laplace both work well in the feature extraction. Compare to the Harris-Laplace, the Hessian-Laplace can bring higher recall rate, which is important when the background of HRRS is complicated. Besides, the HNM strategy can significantly increase the detection precision, because the samples which are misclassified in every iteration step will be trained again as negative data, this operation enhance the discriminate power of the classifier.

\[ \text{Precision} = \frac{N_{DS}}{N_{DS} + N_{DF}} \]  \hspace{1cm} (5)

**Figure 2.** Comparison of the Harris-Laplace and Hessian-Laplace methods with and without HNM. (a) Precision-Recall curve of Harris-Laplace (b) Precision-Recall curve of Hessian-Laplace.

### 3.3. Comparisons with other detecting models

**Table 1.** Comparisons with other aircraft detecting models.

| Method in [2] | Method in [4] | Method in [13] | Proposed method |
|---------------|---------------|----------------|-----------------|
| Precision     | 85.1%         | 58.8%          | 88.2%           | 89.7%           |
| Recall        | 59.6%         | 70.1%          | 92.4%           | 94.0%           |
| F1-measure    | 70.1%         | 63.9%          | 90.2%           | 92.1%           |
| Running time(s)| 19.78        | 1.54           | 0.91            | 0.81            |

In this part, performance of our proposed method is evaluated and compared with the other state-of-the-art methods.

We compare our detection model with the models in [2], [4] and [13]. Table 1 shows the Precision rate, Recall rate F1-measure and Running time(s). From the table we can see that our method achieves high detection performance and with less running time. We also make a comparison of detection result of aircraft location, as shown in Figure 3. In the figure of detection result, compare to the other detection model. Our model can detect the most of aircrafts without being influenced by other land use and land cover. Furthermore the aircraft targets can be accurately positioned, whether the background is complicated or not. Specifically, as shown in the second or third row of Figure 3, when there are many different kinds of aircraft in the HRRS images, our model can also distinguish the right aircraft targets from the background accurately.
Figure 3. Comparison of the detection result based on the accurate NMS and three competing methods. (a) Original images, (b) proposed in [2], (c) proposed in [4], (d) proposed in [13], and (e) Ours.

4. Conclusion
In this work, we propose a novel model for aircraft detection in HRRS images. Our model provides a rapid but efficient way to detect aircraft in the HRRS images. We adopt affine invariant feature to extract the interest points, and encode these low-level features with BOVW. Then we train the detector with cascade AdaBoost classifier, and merge the result based on a new NMS algorithm. The comprehensive experiment is conducted to verify the performance. Competitive results prove the simplicity and effectiveness of the proposed model.

Acknowledgments
The authors would like to acknowledge the support from the Aeronautical Science Foundation of China (201555U8010).

References
[1] G. Cheng, J. Han, A survey on object detection in optical remote sensing images, ISPRS J. Photogramm. Remote Sens. 117 (2016) 11-28.
[2] Z. An, Z. Shi, X. Teng, X. Yu, and W. Tang, An automated airplane detection system for large panchromatic image with high spatial resolution, Optik 125 (2014) 2768-2775.
[3] W. Zhang, X. Sun, H. Wang and K. Fu, A generic discriminative part-based model for geospatial object detection in optical remote sensing images, ISPRS J. Photogramm. Remote Sens. 99 (2015) 30-44.
[4] A. Zhao, K. Fu, H. Sun, X. Sun, F. Li, D. Zhang and H. Wang, An effective method based on ACF for aircraft detection in remote sensing images, IEEE Geosci. Remote Sens. Lett. 14 (2017) 744-748.
[5] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus and Y. LeCun, Overfeat: Integrated recognition, localization and detection using convolutional networks, (2013) arXiv preprint arXiv:1312.6229.
[6] A. Vedaldi and B. Fulkerson, “VLFeat: An open and portable library of computer vision algorithms,” in Proceedings of the 18th ACM international conference on Multimedia ACM 2010 pp. 1469-1472.
[7] K. Mikolajczyk and C. Schmid, Scale & affine invariant interest point detectors, Int. J. Comput. Vision 60 (2004) 63-86.
[8] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision 60 (2004) 91-110.

[9] S. Lazebnik, C. Schmid, J. Ponce, “Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories,” in Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, Volume 2 (2006) pp. 2169-2178.

[10] R.E. Schapire and Y. Singer, Improved boosting algorithms using confidence-rated predictions, Machine Learning, 37 (1999) 297-336.

[11] J. Friedman, T. Hastie and R. Tibshirani, Additive logistic regression: a statistical view ofboosting (with discussion and a rejoinder by the authors). Ann. Statist. 28 (2000) 337-407.

[12] J.F. Henriques, J. Carreira, R. Caseiro and J. Batista, “Beyond hard negative mining: Efficient detector learning via block-circulant decomposition,” in proceedings of the IEEE International Conference on Computer Vision (ICCV) (2013) pp. 2760-2767.

[13] X. Han, Y. Zhong and L. Zhang, An efficient and robust integrated geospatial object detection framework for high spatial resolution remote sensing imagery, Remote Sens. 9 (2017) 666.