# ROBUST AIRPLANE DETECTION IN SATELLITE IMAGES

†*Wei Li,*†*Shiming Xiang,* †*Haibo Wang,* †*Chunhong Pan*

†National Laboratory of Pattern Recognition, Institute of Automation
Chinese Academy of Sciences, Beijing, 100190, China
{wli, smxiang, hbwang, chpan}@nlpr.ia.ac.cn

## ABSTRACT

Automatic target detection in satellite images remains a challenging problem. The main difficulties lie in the co-occurrence of variations of target type, pose, and size in huge satellite image. In this paper, we propose a new airplane detection approach based on visual saliency computation and symmetry detection. The advantages are twofold. First, saliency and symmetry detection perform stably in obtaining target location and orientation information. Second, independent of target type, pose and size, saliency map and symmetry detection are computed only once. This saves a large amount of computational time but does not miss any targets. Experiments show that our method provides a promising way to detect airplanes in complex airport scenes.

***Index Terms***— airplane detection, saliency, symmetry detection

## 1. INTRODUCTION

Automatic target detection in high resolution satellite images is a fundamental yet difficult task in computer vision. Although it has been studied for years [1][2][3], it remains difficult to develop a generic framework that can automatically detect all kinds of interesting targets. In this work, we focus on airplane detection in a complex airport scene, which is one of the most important yet most challenging problems.

The standard approach to target detection is template matching. Two different types exist in literature: sliding window [4, 5] and voting method [6]. Sliding window performs an exhaustive search over discrete image domain and transform space to estimate object and its pose. By contrast, in a voting scheme, a target is first decomposed into a set of parts. Each part is then separately localized and votes for the possible location and pose of this target.

However, these algorithms can not be directly applied for airplane detection in remote sensing images. Since sliding window searches over entire image domain, in a huge satellite image, the computational cost becomes undesirable. Furthermore, when considering the variations of size, orientation and airplane type, the time turns to unaffordable. Contrarily, the success of voting largely relies on how well each part can be localized. However, in airport scene, parts of an airplane appear to resembling to some manmade objects, such as little buildings, vehicles and embarkation. This fact leads to many erroneous part localizations, making voting fail frequently.

A promising solution is to combine the usage of sliding window and voting. If searched image domain and transform space can be greatly reduced, the time cost of sliding window will become reasonable. In practice, if we have well-defined attributes of airplane, the success possibility of voting can increase. In this paper, we show that these two problems can be simultaneously solved. In particular, we observe two basic facts about airplane and utilize them in such a way that they can hierarchically solve these two problems.

First, we observe that in airport scene, airplane is likely to be salient. The intuition is that airplanes can attract our attention at the first glance of a satellite image. This implies that we can apply existing saliency techniques to generate airplane candidates. The practical benefits are twofold. For one thing, it removes most background regions, which greatly reduces image domain for sliding window to search. For another thing, it provides many ROIs on which we can extract reliable attributes of an airplane. In this work, we employ the saliency method proposed by [7] since it is fast yet efficient.

Second, we observe that airplane is symmetric about the center line of its own body. If well measured, this symmetry enables us to localize its axis of symmetry precisely. From this axis, we can know the principal orientation and the rough position of the airplane. As a result, the search space of sliding window is guided to be small. For detecting this symmetry, we rely on the shape cues of an airplane and predict the axis of symmetry via efficient voting.

## 2. SALIENCY BASED ROI EXTRACTION

In this section, the main task is to segment the input image into ROIs. The proposed ROI extraction method is mainly based on visual saliency. Generally, saliency based ROI extraction approach is built on human attention mechanism. Our motivation of utilizing saliency arises from two aspects. First, we do not need to specify the prior of airplane target. Second,

This research is sponsored by Natural Science Foundation of China(NSFC No. 60873161, 61005013,61075016).

we can locate ROIs quickly. The input image is first analyzed by constructing a saliency map. Then, the saliency map is segmented into ROIs.

For a given image, the saliency map SM is constructed in terms of phase spectrum of Fourier transform [7],

$$\mathbf{SM} = \left\| F^{-1}\left(e^{i \cdot P(F(\mathbf{I}))}\right) \right\| \otimes g, \tag{1}$$

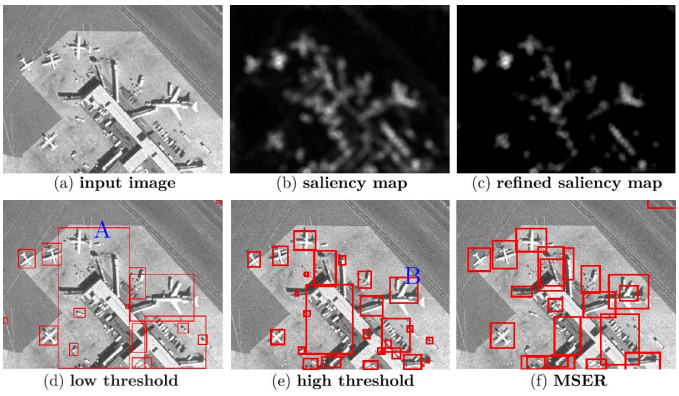
where $g$ is the two dimensional Gaussian kernel, $F^{-1}(\cdot)$, $F(\cdot)$, and $P(\cdot)$ are the Fourier transform, inverse Fourier transform, and image phase spectrum, respectively. Fig. 1 gives an example of saliency map. From this figure, we see that the saliency value of target is higher than that of the background. Thus, the saliency map provides very useful information for ROI extraction.

On the segmentation stage, classical algorithms assume that the regions with saliency values higher than a threshold are regarded as salient, while other areas become background. This simple segmentation strategy has a problem in practice. That is, it is difficult to choose a proper threshold. If selected improperly, target may be discarded (see ROI **B** of Fig. 1(e)) or adjacent salient region may be merged together (see ROI **A** of Fig. 1(d)).

Here, we solve the above problem in two steps. First, we improve the saliency map by spatial competition algorithm [8]. This spatial competition is similar to a local 'winner-take-all' procedure. It enhances isolated targets while suppressing the boundary regions. Hence, the segmentation on modified saliency map is less sensitive to the threshold selection. In comparison with Fig. 1(b), the target regions Fig. 1(c) are enhanced. In particular, the target regions are well separated in the refined saliency map.

Second, a marker based Maximally Stable Extremal Region (MSER) [9] is employed to segment saliency map into small regions. Local saliency maxima points are detected as seeds, around which MSER extracts extremal regions, whose boundaries are stable under a range of threshold. Marker based MSER can be regarded as a local threshold-adaptive segmentation approach. For some markers, multiple stable thresholds exist, accordingly several nested regions are detected, which makes the result more reliable. From Fig. 1, we see that ROIs of the proposed approach are more precise in coving airplane targets.

As interpreted above, the proposed ROI extraction method has the following advantages. First, the computation cost of our method is lower than the classical methods, which extract ROIs from the segmented image. Second, compared with other saliency based ROI extraction methods, the proposed method has higher ROI accuracy. Moreover, it is also insensitive to target type and size. Hence, our saliency based ROI extraction can be a good support for the next detection step.




(a) input image (b) saliency map (c) refined saliency map

(d) low threshold (e) high threshold (f) MSER

**Fig. 1**. An example of the **Saliency Based ROI Extraction** stage. (b) gives the initial saliency map, (c) depicts the saliency map refined by spatial competition. (d) and (e)show **ROI**s extracted from directly thresholding the saliency map using threshold 0.2, and 0.4, respectively. Comparatively, (f) illustrates the **ROI**s extracted with the proposed method.

## 3. AIRPLANE DETECTION IN ROIS

Once ROIs have been extracted, we can detect airplanes efficiently. To do this, we first detect the symmetry of an airplane to roughly localize its orientation and position. Then we refine the localization via template matching.

### 3.1. Detecting airplane symmetry

For symmetry detection, the classical methods use local invariant feature points [10]. However, since airplanes may appear very small, local invariant features are often few. Instead, we rely on airplane shape, which is relatively invariant to airplane size. Suppose $\mathbf{E}$ represent the shape obtained in an arbitrary ROI, and $\mathbf{E}_{\mathcal{T}}$ the reflective shape of $\mathbf{E}$ under the reflectional transform $\mathcal{T}$. By its definition, ideally we have $\mathbf{E} = \mathbf{E}_{\mathcal{T}}$. Thus, for the given $\mathbf{E}$ and $\mathbf{E}_{\mathcal{T}}$, we can minimize the following distance to estimate $\mathcal{T}$:

$$\min_{\mathcal{T}} d_{DCM}(\mathbf{E}_{\mathcal{T}}, \mathbf{E}), \tag{2}$$

where $d_{DCM}$ is defined as the directional chamfer matching (DCM) distance [5]. This distance is a fast and robust metric to measure shape similarity, which extends the original $2D$ chamfer matching to $3D$ by considering the additional edge orientation. In order to compute the DCM distance, the shape $\mathbf{E}$ is represented as a set of line segments. For airplane ROI, its shape $\mathbf{E}$ consists of a set of line segments, which are extracted from the edge map of the ROI, using the edge linking and line fitting algorithms [1].

Suppose that $\mathbf{E}$ is composed of $m$ line segments, $\mathbf{E} = \{L_j\}_{j=1}^{m}$, and $L_j$ is composed of $n_j$ edge points, $L_j =$

[1] http://www.csse.uwa.edu.au/ pk/Research/MatlabFns/

$\{\mathbf{e}_i\}_{i=1}^{n_j}$. Similarly, the reflective shape $\mathbf{E}$ is represented as $\mathbf{E}_{\mathcal{T}} = \{\widehat{L}_j\}_{j=1}^{m}$, and $\widehat{L}_j = \{\mathbf{t}_k\}_{k=1}^{n_j}$. Thus, the DCM distance $d_{DCM}$ is defined as

$$\begin{aligned} & d_{DCM}(\mathbf{E}_{\mathcal{T}}, \mathbf{E}) \\ =& \frac{1}{n} \sum_{\widehat{L}_j \in \mathbf{E}_{\mathcal{T}}} \sum_{\mathbf{t}_k \in \widehat{L}_j} \min_{\mathbf{e}_i} (|\mathbf{t}_k - \mathbf{e}_i| + \lambda \, |o(\mathbf{t}_k) - o(\mathbf{e}_i)|), \end{aligned} \quad (3)$$

where $o(.)$ is the orientation of an edge point, which is quantized into $K$ discrete channels evenly in the $[0, \pi)$ range. $\lambda$ is a weighting factor that make the orientation different comparable with location difference. And we have $n = \sum_j n_j$. This DCM distance is actually calculated by the efficient $3D$ distance transform [5].

Due to self-shadows and other salient structures, the obtained shape $\mathbf{E}$ not only contains target edge but also a lot of clutter. Since $\mathbf{E}$ is not strictly symmetric, it is unreliable to use the global DCM distance to detect symmetry. To solve this problem, we use Hough vote to estimate symmetry axis. All the straight line segments in ROIs are considered as potential symmetry axes, each line segment can be represented in polar coordinate as $(\rho, \theta)$, where $\theta$ is the line direction and $\rho$ is the distance from image coordinate origin to the line. Reflectional transform around line $(\rho, \theta)$ is denoted as $\mathcal{T}(\rho, \theta)$. For line segment $L_i$, if its mirror segment has a low distance to the image edge map, it is very likely that $L_i$ is part of a symmetry shape. Thus it casts a vote to the corresponding axis. Symmetry magnitude $M(\rho, \theta)$ quantifies the amount of symmetry of axis $(\rho, \theta)$, which is defined as:

$$M(\rho, \theta) = \sum_j |\widehat{L}_j| \exp\left(\frac{-d_{DCM}(\widehat{L}_j, \mathbf{E})}{\sigma}\right), \quad (4)$$

where $\sigma$ presents the tolerance to shape deformation and $|\widehat{L}_j|$ denotes the length of $\widehat{L}_j$.

The resulting Hough space $M(\rho, \theta)$ is blurred with a Gaussian. Then two strongest local maxima are retained as candidate symmetry axes. Fig. 2 shows some results, on which the red and blue lines represent the two most probable axes, respectively. As demonstrated in this figure, influenced by shadows, the position of detected axes often deviate from the true symmetry axes slightly, but the orientation is always accurate. Although line segments are less distinctive than local invariant feature points, they can also result in reliable pose estimation, when the vote space is restricted in ROI.

### 3.2. airplane detection by sliding window

We use a template-based shape matching algorithm to detect airplanes. An airplane edge template is first mapped on each ROI by aligning its symmetry axis with the detected axis. Then, we slide the template window in ROI to find the minimum of the DCM distance between image edge map and edge template.

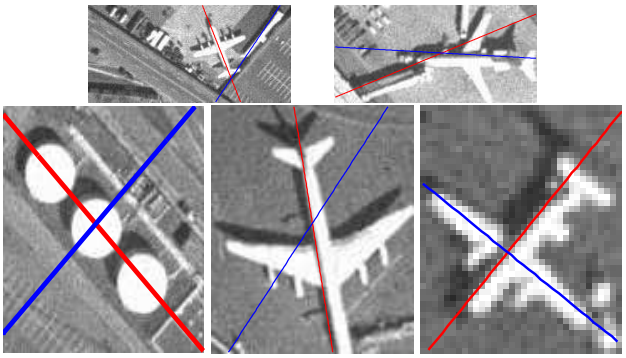


**Fig. 2**. symmetry axis detection results. Red line: most likely axis; Blue line: second most likely axis.

Since the orientation of detected symmetry axis is reliable, template window needs only to be slided along the direction perpendicular to the axis. For all possible target templates, symmetry detection is only computed once. Thus, search space is reduced in both the orientational and positional dimension. In addition, this template matching scheme is efficiently implemented, since it can use the $3D$ distance transform map already computed for the symmetry detection.

## 4. EXPERIMENTAL RESULT

Both qualitative and quantitative experiments were conducted to evaluate our method. The test dataset includes $40$ satellite images of complex airport scenes. The image size varies from $600 \times 600$ to $2000 \times 2000$. The involved parameters are given as follows: for the DCM computation, we set $K = 60$ and $\lambda = 0.5$ as in [5]; for the symmetry voting, we set $\sigma = 1$.

The goal of ROI extraction is to find ROIs covering all airplane targets. To evaluate its performance, we use two different metrics. First is the focus rate, which is the ratio of the area of all ROIs to the area of the entire satellite image. By using it, we intend to measure the efficiency of ROI extraction in reducing image domain for further airplane search. The main reason of using it instead of the traditional false alarms lie in that the computational cost of symmetry detection and shape matching is only linear to the sum of the areas of all ROIs . Second is the detection rate, which is the ratio of the number of correctly extracted airplane ROIs to the number of ground-truthed airplanes. In our experiments, the average focus rate is $12\%$ while the detection rate reaches up to $100\%$. Thus, all the removed $\approx 90\%$ regions are backgrounds. Especially, most highly textured backgrounds are successfully removed. We mention this because the next shape matching is very sensitive to these textures. As shown in Fig.3, the forests form highly textured backgrounds of the scene. Although they have high responses in the saliency map, our spatial competition strategy suppresses them.

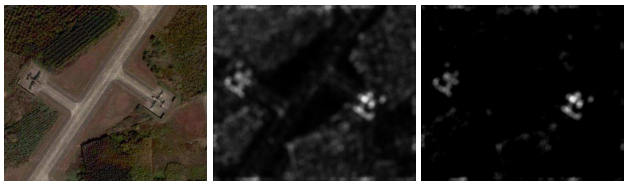


**Fig. 3**. Suppressing highly textured backgrounds. From left: input satellite image, saliency map, and saliency map after spatial competition.

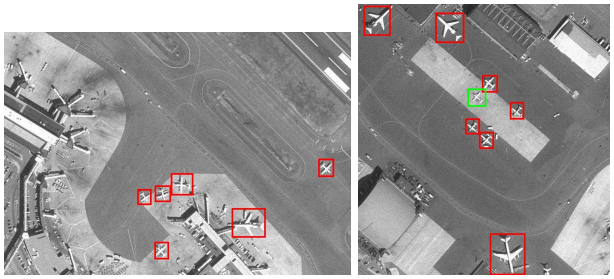


**Fig. 4**. Illustrating airplane detection results. Red rectangle: true positive; Green rectangle: true negative.

In the second experiment, we extensively compared our approach against the DCM-based sliding window method (DCM) on our dataset. Tab 1 shows the results. Since DCM relies on exhaustive sliding window, the detection rate of our approach is a little lower. However, we easily beat DCM in false alarm rate and time cost. In our approach, saliency based ROI extraction greatly reduce image space to search, and removes most backgrounds as well. Symmetry detection then reduces transform space by providing initial orientation and positions, and meanwhile rejects some other manmade targets. Fig. 4 gives two detection results of our approach. In the scene, the airplane types, orientations and sizes are all having large variations. The sizes of the 14 airplanes in the two images vary from 40 to 100 pixels. Their orientations look very inconsistent. Many confusing manmade objects, such as little buildings and vehicles, are also scattered in the scene. In spite of these challenges, our approach shows its power. Among the 14 targets, only one small airplane (see green bounding box) is missed, and no false alarms are reported. In our analysis, two factors attribute to its missing. The small size of this missed airplane causes some errors in fitting line segments and the ambiguous part of the airplane's fuselage makes DCM matching inaccurate.

**Table 1**. Detection performance comparison.

| | Detection rate | False alarm rate | Time |
|---|---|---|---|
| **DCM** | 0.95 | 0.3 | 3.54min |
| our method | 0.93 | 0.05 | 41.7s |

The most computationally expensive part of our approach is computing $3D$ distance transform maps for all the ROIs. Its time complexity is $O(NK)$, where $N$ is the number of the pixels of all ROIs and $K$ is number of discrete orientations. Constructing a $3D$ distance transform map for a $120 \times 120$ ROI will take about 1.5 seconds on a PC with 2.8GHz CPU and 1.5Gb RAM by using Matlab.

## 5. CONCLUSION

This paper proposes a new airplane detection algorithm by combining saliency and symmetry detection. We use a low-cost bottom-up saliency computation and MSER detection algorithm to yield initial ROIs. In each ROI, target orientation is estimated by symmetry detection, and then airplane is detected via shape matching. Experimental results show that this approach is robust and fast to detect airplanes in complex airport scenes.

## 6. REFERENCES


[1] A. Filippidis, L.C. Jain, and N. Martin, "Fusion of intelligent agents for the detection of aircraft in sar images," *IEEE T-PAMI*, vol. 22, no. 4, pp. 378–384, April 2000.

[2] Bernard Chalmond, Benjamin Francesconi, and Stphane Herbin, "Using hidden scale for salient object detection," *IEEE T-IP*, vol. 15, pp. 2644–2656, 2006.

[3] Jian Yao and Zhongfei (Mark) Zhang, "Semi-supervised learning based object detection in aerial imagery," *CVPR*, vol. 1, pp. 1011–1016, 2005.

[4] Clark F. Olson and Daniel P. Huttenlocher, "Automatic target recognition by matching oriented edge pixels," *IEEE T-IP*, vol. 6, pp. 103–113, 1997.

[5] Ming-Yu Liu, Oncel Tuzel, Ashok Veeraraghavan, and Rama Chellappa, "Fast directional chamfer matching," in *CVPR*, 2010, pp. 1696–1703.

[6] M. Villamizar, F. Moreno Noguer, J. Andrade Cetto, and A. Sanfeliu, "Efficient rotation invariant object detection using boosted random ferns," in *CVPR*, 2010, pp. 1038–1045.

[7] C.L. Guo, Q. Ma, and L.M. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," in *CVPR*, 2008, pp. 1–8.

[8] Laurent Itti and Christof Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision Research*, vol. 40, pp. 1489–1506, 2000.

[9] J. Matas, O. Chum, M. Urban, and T.Pajala, "Robust wide baseline stereo from maximally stable extremal regions," in *BMVC*, 2002, pp. 384–393.

[10] Gareth Loy and Jan olof Eklundh, "Detecting symmetry and symmetric constellations of features," in *ECCV*, 2006, pp. 508–521.