© Indian Academy of Sciences

CrossMark

Detecting aircrafts from satellite images using saliency and conical pyramid based template representation

SAMIK BANERJEE[1], NITIN GUPTA[1], SUKHENDU DAS[1,*], PINAKI ROY CHOWDHURY[2] and L K SINHA[2]

[1] Department of Computer Science and Engineering, Indian Institute of Technology Madras, Chennai 600036, India
[2] Defence Terrain Research Laboratory, Defence Research and Development Organization, Delhi 110054, India
e-mail: samik@cse.iitm.ac.in; niting@cse.iitm.ac.in; sdas@cse.iitm.ac.in; pinaki@dtrl.drdo.in; lksinha@dtrl.drdo.in

Abstract. Automatic target localization in satellite images still remains as a challenging problem in the field of computer vision. The issues involved in locating targets in satellite images are viewpoint, spectral (intensity) and scale variations. Diversity in background texture and target clutter also adds up to the complexity of the problem of localizing aircrafts in satellite images. Failure of modern feature extraction and object detection methods highlight the complexity of the problem. In the proposed work, pre-processing techniques, viz. denoising and contrast enhancement, are first used to improve the quality of the images. Then, the concept of unsupervised saliency is used to detect the potential regions of interest, which reduces the search space. Parts from the salient regions are further processed using clustering and morphological processing to get the probable regions of isolated aircraft targets. Finally, a novel conical pyramid based framework for template representation of the target samples is proposed for matching. Experimental results shown on a few satellite images exhibit the superior performance of the proposed methods.

Keywords. Localization; aircraft; MOCP; saliency; matching; DBSCAN; normalized cross correlation.

1. Introduction

Over the past decade, scientists in computer vision have found it difficult to develop an efficient framework that can automatically detect all kinds of interesting targets in satellite images. In this paper, we try to address the challenging task of aircraft detection [1–4] from satellite images. To handle this task of automatic aircraft detection, the standard approach adopted by the vision scientists is template matching in its two different variants: sliding window [5, 6] and voting method [7]. An exhaustive search over the entire image domain is performed in case of the sliding window to estimate the object and the pose, while in a voting scheme, a target is first decomposed into a set of parts, which is separately localized and votes for the possible location and pose of this target.

Results of matching by using two state-of-the-art feature extraction techniques: (a) scale invariant feature transform (SIFT) proposed by Lowe [8], and (b) speeded up robust features (SURF) proposed by Bay *et al* [9] are shown in figure 1. From the results, we can see that only few correct

targets are detected in both the cases, and a large set of false alarms are produced. This example shows a typical case, where modern feature based matching techniques do not produce the desired results for automatic target localization task. To solve this problem of target localization, we resort to the mechanisms of saliency combined with modules for enhancing the templates/images.

In the satellite images having aircrafts as targets, the most likely salient part of the image which instantly attracts our attention are the aircrafts. Hence, we use a combination of the existing saliency and clustering based techniques to localize the aircrafts in the image. It helps to remove the background regions, which inturn reduces the search space significantly. The sliding window based techniques perform efficiently on the selected spatial regions of the scenes. A set of widely varying scales and discrete set of orientations (covering the entire range of 2π) are used to form a synthetically generated multi-oriented conical pyramid (MOCP) for each template of different aircrafts as targets.

In this work, we propose a framework with two variants: extensive scan and quick scan; to handle this problem of automatic target localization in satellite images, based on time and performance trade-offs. The common stages in the

---

*For correspondence
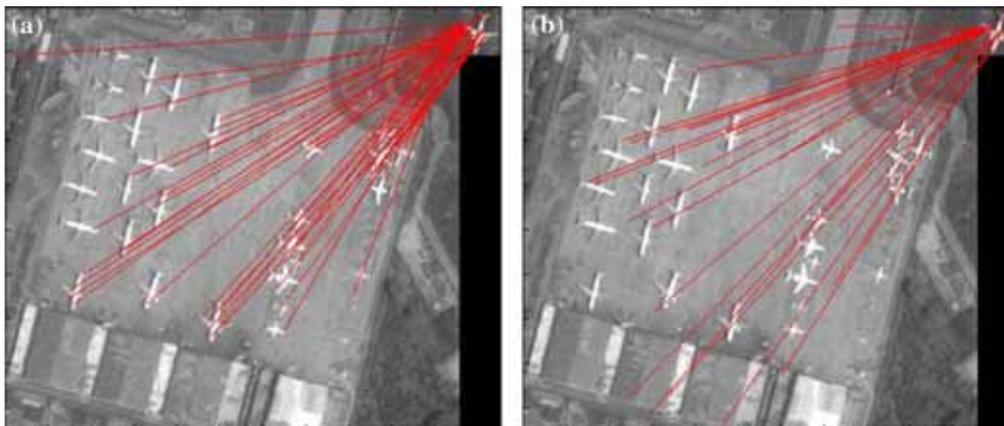
Figure 1. Results of feature based matching using (**a**) SIFT [8] and (**b**) SURF [9]. The *red color lines* show the correspondence of keypoints between query image and template images. Many targets are missed (false negatives) and many incorrect locations (false alarms) are obtained.

framework includes the saliency region detection and normalized cross-correlation [10] based matching in both the variants. In case of extensive scan, the images are preprocessed by deblurring, denoising and contrast enhancement followed by stages of matching. The quick scan variant surpasses the pre-processing of images and uses a simpler saliency based method for finding candidate target regions (CTR), with reduced computational time.

The major contributions of the work presented in this paper are: (1) Combining concepts of saliency and clustering for search space reduction; (2) Multi-oriented conical pyramid (MOCP) representation for efficient matching of aircraft templates, to deal with large variations of scale and poses; (3) Providing a better performance than a state-of-the-art method. Experimental results on real world satellite scenes exhibit the superior performance of the proposed method, over some similar attempts of aircraft localization. The rest of the paper is organized as follows: We discuss related work in section 2. Section 3 gives a detailed overview of proposed frameworks for target localization. Section 4 describes detailed discussion of experimental results and section 5 concludes the work.

2. Related work

In the past few decades, a large number of algorithms and systems [1, 2, 4, 11, 12] have been proposed in order to address target detection problem. Perrotton *et al* [3, 11] proposed a Haar like descriptor and adaboost based learning approach, which automatically learns a multi-view detector for object detection task. Results are shown on only two synthetically generated images of optical band, containing aircrafts on airports. Another approach of detecting aircrafts from high resolution satellite imagery is proposed by Si *et al* [12]. In this approach, a bilateral

filtering is first used to denoise the image, and then a detection operator based on the boundary shape information (BSI) is applied to detect enhanced objects. Results are shown on only 21 high resolution images of optical band. Cai *et al* [13] designed a man-made object detection methodology by which typical man-made objects in complex natural background, including airplanes, tanks and vehicles can be detected, by extracting texture and geometry structure features.

Various saliency based approaches like Li and Itti [2], Li *et al* [14] were also proposed in the past to handle the problem of automatic aircraft detection in satellite images. In the work by Li and Itti [2], high resolution satellite images were first divided into small images, from which saliency and gist features were extracted and concatenated to classify each small image as containing target(s) or not, using a support vector machine. Results were shown on three categories: boats, buildings and airplanes. Li *et al* [14] proposed an airplane detection approach which utilized the concept of visual saliency computation and symmetry detection. Recently, bag-of-words (BOW) based model, which is used for text and image categorization, has been also introduced for object classification in remote sensing images. In the work by Xu *et al* [15], a BOW model was used for land-use mapping of a high-spatial-resolution aerial photograph for object-based classification. Liu *et al* [16] proposed a spatial-sparse-coding-BOW (SSCBOW) model-based detection framework for targets with complex shapes in high resolution remote sensing images. Results were shown on remote sensing images from Google Earth for aircraft detection. Most of the images used were of high resolution, containing clear pictures of aircrafts as targets.

Several deep learning frameworks like Cheng *et al* [17], Wu *et al* [18] were also proposed in recent past to handle the problem of automatic target detection in satellite images. Cheng *et al* [17] used a deep learning framework for vehicle

(car) detection in satellite images, while Wu *et al* [18] proposed a framework which utilizes the concept from objectiveness detection techniques (e.g., BING Cheng *et al* [17]) and convolutional neural networks (CNN) for aircraft detection. In the work by Chen *et al* [19], deep belief nets has been used to detect the aircrafts in satellite images. Liu *et al* [16] proposed a coarse-to-fine shape method based on edge computing to recognize aircrafts. The framework proposed by Zhang *et al* [20] consists of a weakly supervised training procedure to yield the detectors and an effective scheme to detect targets from testing images. Results are shown on aircrafts on airports from hyperspectral imagery. A deformable part-based model (DPM) has been proposed by Felzenszwalb *et al* [21], to detect objects in complex scenes. Malisiewicz *et al* [22] use the concept of an ensemble of Exemplar-SVMs for object detection task. Mishne *et al* [23] proposed a detection algorithm based on data-driven target modeling which implicitly handles variations in target appearance, by constructing a low-dimensional embedding of test images using a metric based on local neighborhoods within the training set, in a supervised graph framework. Results are shown on 44 high-resolution images of optical band. Images had clear target aircrafts on airports, on a clear background.

The computer vision community has proposed a number of techniques to solve the problem of image denoising, deblurring and enhancement tasks. Basic approaches for denoising, such as Gaussian and median filtering, have a tendency to over-smooth edges and remove image detail. More sophisticated approaches use the properties of natural image statistics to enhance large intensity edges and suppress lower intensity edges. Zhang *et al* [24] proposed an efficient image denoising scheme which uses principal component analysis (PCA) with local pixel grouping (LPG). A generalized and mathematically sound L0 sparse expression is proposed by Xu *et al* [25], for motion deblurring. Dong *et al* [26] proposed a sparse representation technique for image restoration task. A set of autoregressive models are learned from the dataset of example patches and also non-local self-similarity is introduced as a regularization term. An advanced multiband satellite color, contrast improvement technique is proposed by Bhandari *et al* [27]. A new technique has been proposed by Bhandari *et al* [28], based on the singular value decomposition (SVD) and discrete cosine transform (DCT). The proposed technique of modified SVD-DCT converts an image into the SVD-DCT domain after normalizing the singular value matrix. Buades *et al* [29] proposed an algorithm for image denoising task, called non-local means denoising. Gu *et al* [30] propose a new spectral matching method for target detection in hyperspectral imagery, which utilizes a prewhitening procedure and defines a regularized spectral angle between the spectra of the test sample and the targets.

Detection of salient image regions is useful for applications like image segmentation, adaptive compression, and region-based image retrieval. In recent past, researchers are putting great effort to develop efficient saliency detection strategies [31–34] for partitioning an image into two segments: foreground object and background. The work proposed in Zhu *et al* [35] presents a novel method based on both bottom-up (BU) saliency and top-down saliency. Top-down, bottom-up, combination of both top-down and bottom-up saliency models are used by researchers. The approaches for determining low-level saliency can be based on biological models or purely computational ones. Some approaches consider saliency over several scales while others operate on a single scale. In general, all methods use some means of determining local contrast of image regions with their surroundings using one or more of the features of color, intensity, and orientation.

Quaternion based spectral saliency method [36, 37] uses Quaternion Fourier or cosine spectrum to calculate the visual saliency. It is useful for color image processing, especially for the color-sensitive smoothing, edge detection, and data compression. Graph based visual saliency (GBVS), a bottom up saliency method proposed by Harel *et al* [31], forms activation maps on certain feature channels, and then normalizes them in a way which highlights conspicuity and admits combination with other maps. Regional contrast based saliency extraction algorithm has been proposed by Cheng *et al* [38], which simultaneously evaluates global contrast differences and spatial coherence. Goferman *et al* [39] proposed a Context-aware saliency which aims at detecting the image regions that represent the scene. Mean shift segmentation algorithm was proposed by Comaniciu and Meer [40], which is non-parametric technique for the analysis of a complex multimodal feature space and to delineate arbitrary shaped clusters in it. Ester *et al* [41] proposed a density-based spatial clustering of applications with noise (DBSCAN) technique for data clustering. It is called a density based clustering technique because it finds a number of clusters starting from the estimated density distribution of corresponding nodes.

The work presented in this paper consists of rigorous experimentation to select optimal stages in a framework, designed for aircraft detection. Most methods of foreground or salient object detection are not scale and clutter (when target objects lie close and overlap) invariant. Shape feature based methods rely on accurate contour or object part detection, which perform unsatisfactorily under noisy and blurry condition, specifically when small object parts are not highlighted by feature detection operators. Hence most of these modern methods fail to deliver even satisfactory performance in our context/application of aircraft detection, where the foreground object is of low contrast and small in spatial extent (size). We have verified this using latest feature extraction algorithms (SURF, SIFT). The novelty of our proposed methods lies in combining modern and traditional methods in an optimal way for the design of a pipelined framework, which is mostly unsupervised. Training samples are used for MOCP representation, which is only used at the final stage of matching.
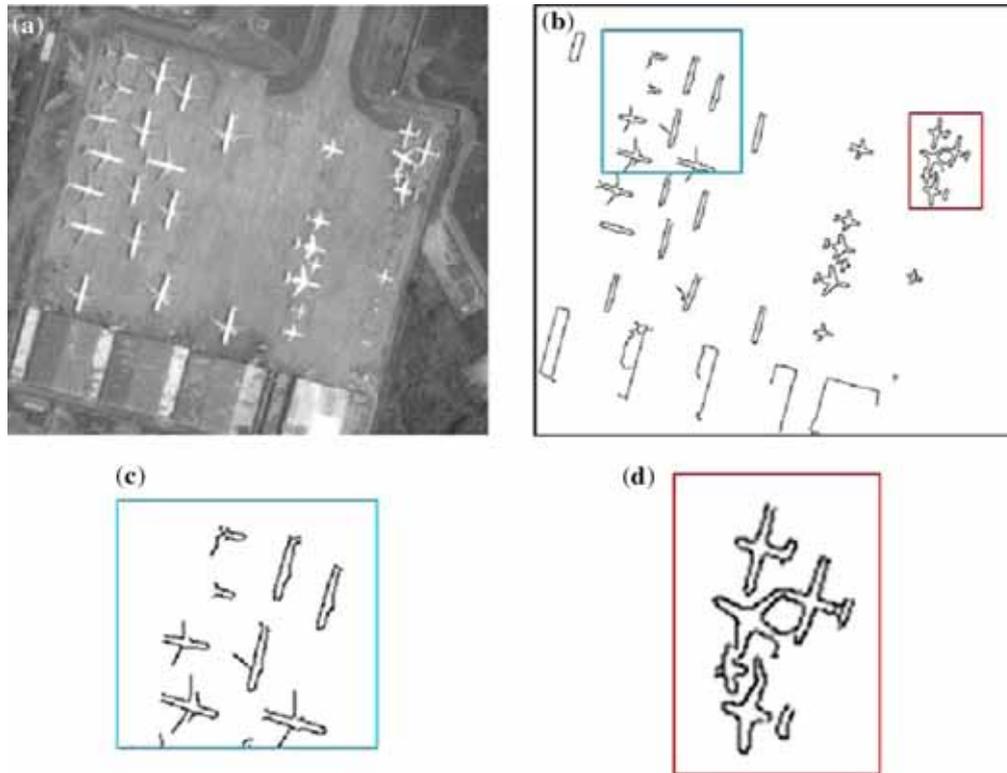
Figure 2. Illustrations to highlight the complexity of the problem being addressed: (**a**) satellite image; (**b**) edges detected using Canny's edge detector [42]; (**c**), (**d**) zoomed in parts of (b) to highlight the incompleteness and overlap in contours of the aircrafts. Contour based or object based methods cannot perform in such complex scenarios, to detect targets.

## 3. Proposed stages of the algorithm for localization of aircrafts

As shown in figure 1, modern feature-extraction techniques (SIFT and SURF) do not perform well, due to small size of objects (targets) to be matched. Modern object detection methods such as DPM [43] was also observed to fail in detecting the target objects. As edge or contour detectors [42] provide incomplete or discontinuous contours (see figure 2), shape contour based [44] or object-part based recognition methods [45] also fail to perform well. With proper pre-processing and pose and scale invariant representation of templates, tradition methods work well, which is described below. In this paper, two similar frameworks have been proposed to solve the problem of target localization in satellite images: extensive scan and quick scan. Extensive scan framework is more accurate as compared to the quick scan framework, but less efficient in terms of computational time.

3.1 *Brief description of the overall framework used*

In this paper, we have solved the problem of target recognition and localization from a high-resolution satellite image, using low-level features. Given a satellite scene, we aim to locate and identify the aircrafts as targets, present in the scene. Target aircrafts are available as a gallery of training samples with only a fixed size and pose. The overall framework consists of five stages (see section 3.2):

1. Generation of synthetic samples for targets, with different scales (resolution) and poses, for each target, to form a multi-oriented conical pyramid as (MOCP) gallery.
2. Pre-processing stage for enhancement and denoising, of the given scene to detect targets.
3. Detection of tentative foreground regions of targets using an unsupervised algorithm of saliency estimation. For any larger region detected as salient, partition to create small regions (or blobs) of comparable size as that of gallery templates. These blobs are considered to be those regions potentially containing target aircrafts. We term a blob produced by this process, as candidate target region (CTR).
4. Estimate dominant scale and orientation for each CTR (salient blob).
5. Feature matching using normalized cross correlation (NCC) or distance transform (DT) to detect presence of an aircraft in each blob.
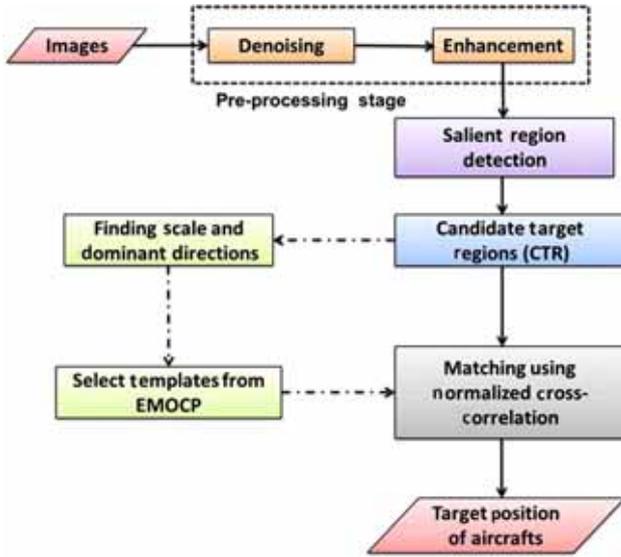
**Figure 3.** Overall extensive scan framework for aircraft localization task.

The flowchart of the overall process is shown in figure 3. It consists of pre-processing stages, creation of salient blobs as candidate target regions and matching blobs with templates. As the quality of the images is often low, they require denoising, deblurring and enhancement. Process of saliency helps in identifying potential targets. The next few sub-sections give details of the individual stages of the pipeline used for processing.

3.2 *Extensive scan*

The proposed extensive scan framework for aircraft localization task consists of two important stages: (a) multi-oriented conical pyramid (EMOCP) generation during training and (b) aircraft localization during testing, which are discussed below in detail:

3.2a *Enhanced multi-oriented conical pyramid (EMOCP) generation stage for extensive scan*: In this stage, a conical pyramid of templates at 7 different scales and 24 view-points is created for the purpose of matching. We term this as a enhanced multi-oriented conical pyramid (EMOCP) (to be discussed below). The overall framework for EMOCP generation is shown in figure 4, which consists of the following four phases:

(i) *Pre-processing phase*—This phase helps to remove the blur and noise from the image, followed by a contrast enhancement process. It consists of three parts:

- *Deblurring*—It is done using a state-of-the-art technique: the L0 sparse representation, proposed by Xu *et al* [25]. The objective function
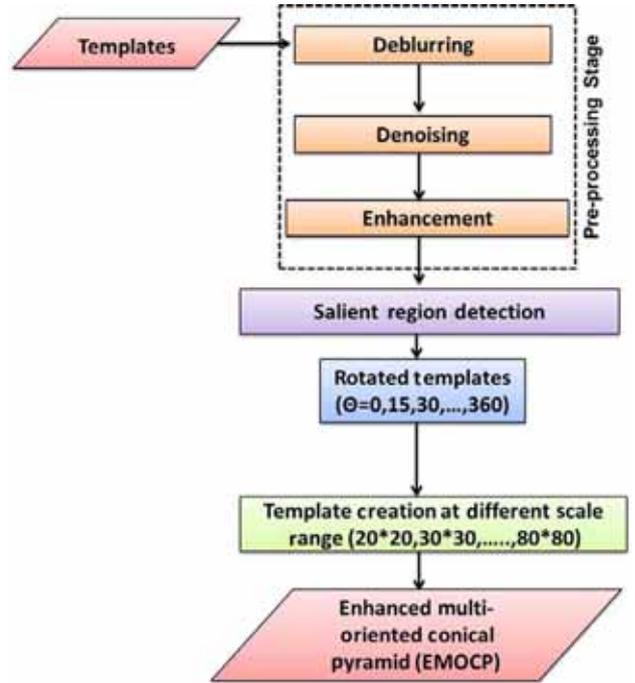
Figure 4. Enhanced multi-oriented conical pyramid (EMOCP) creation task, used in the extensive scan framework.

to estimate the blur kernel from the input image is

$$\min_{(\bar{x},k)}\left\{\left\|\sum_m k_m H_m \bar{x} - y\right\|^2 + \lambda \sum_{\cdot \in h,v}\phi_0(\partial \cdot \bar{x} + \gamma\|k\|^2)\right\}$$
(1)

where $\bar{x}$ is an intermediate sparse representation containing only necessary edges of the aircraft that are crisp and enhanced in the template, H_m is a $N \times N$ transformation matrix, which corresponds to either camera rotation or translation of the satellite for pose m of the aircraft, k_m denotes the time that camera pose m lasts and is a weight in this function. λ and γ are two regularization weights, ϕ_0 is a sparse loss function, y is the blurred observation, vector $k = [k_0, k_1, \ldots]^T$ contains all k_m. ∂ is the partial derivative operator. h, v represents the two neighborhood directions.

- *Denoising*—The deblurred templates are passed to the denoising module proposed by Zhang *et al* [24], which uses the concept of principal component analysis (PCA) with local pixel grouping (LPG). For an underlying pixel to be denoised, a $K \times K$ window centered on it, denoted by $\mathbf{x} = [x_1 \ldots x_m]^T$, $(m = K^2)$, is used to form a vector containing all pixel components within the window. An $L \times L$ $(L > K)$ training

block centered on x_v is selected to obtain the training samples for a noisy variable $x_v = x + v$. Thus, totally $(LK + 1)^2$ training samples for each component x_n^v of x_v is obtained. A block matching method is used for Local pixel grouping (LPG) to identify similar (to the central block) $K \times K$ blocks within the $L \times L$ window. The training dataset is formed by concatenating the selected matching blocks together. The denoised module is obtained by an orthonormal PCA transformation matrix for **x** formed from the training dataset vector.

- *Enhancement*—To further improve the quality of templates, the denoised templates are enhanced by using a DCT+SVD method (see Bhandari *et al* [28] for details). A singular value matrix represents the intensity information of the given image and any change of the singular values in effect alters the intensity of the input image. This technique converts the image into the SVD-DCT domain and after normalizing the singular value matrix, the enhanced image is reconstructed by using inverse DCT.

These three processes of deblurring [25], denoising [24] and contrast enhancement [28] are empirically selected by observing the performance of several other methods published in literature (see section 2), on satellite images of aircrafts.

(ii) *Saliency region detection phase*—The graph based visual saliency (GBVS) [31] is used to get the prominent regions of interest containing aircraft from the enhanced templates. Graph-based visual saliency (GBVS) is a bottom-up visual saliency model, which consists of two steps: first, forming activation maps on certain feature channels, and then normalizing them in a way which highlights conspicuity and admits combination with other maps. Rarity of pixel intensity is used as a criterion in an optimization framework, solved using graph-cut, is used to obtain the salient regions.

(iii) *Rotation phase*—In this phase the processed target templates, each with an aircraft as the foreground salient region, are rotated at 15 degree intervals, to create 24 different oriented (0, 15, 30,...360) templates. These templates are rotated as an image, which in turn rotates the foreground object. During the formation of the rotated templates, the vacant regions appearing in the background layer are filled using the intensity pixels from the boundary.

(iv) *Scale pyramid formation*—A scale pyramid of several levels of resolution (ranging from $20 * 20$ to $80 * 80$ pixels, at steps of 10 pixels), is created for
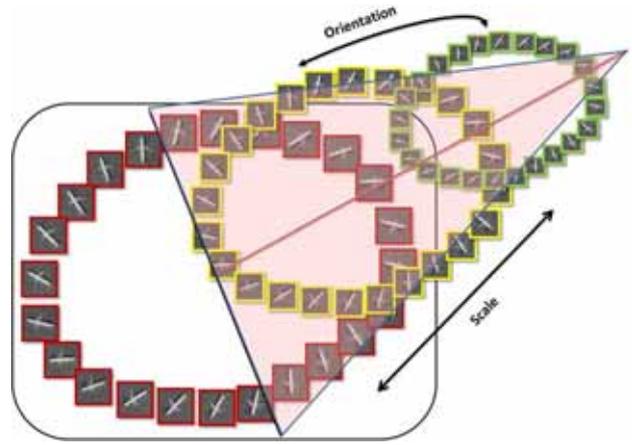
Figure 5. Enhanced multi-oriented conical pyramids (EMOCP) of templates, for a target aircraft.
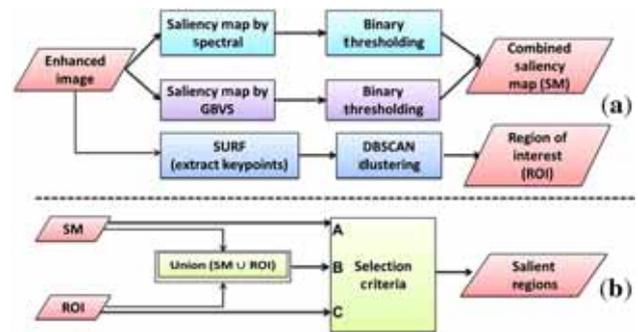
Figure 6. Salient region detection stage for extensive scan framework. The inputs to stage (**b**) are obtained from the outputs of stage (**a**). See figure 3 for the overall framework.

each rotated template obtained after the salient region detection phase. We thus have seven different levels of scales in EMOCP.

Finally, by combining all rotated and scaled templates, an EMOCP is formed as shown in figure 5, which helps in aircraft localization task. Figure 5 is a typical example of the EMOCP which contains all the rotated versions (synthetic templates) of a single template, shown only for three different scales. Each set of potential templates at a particular scale is shown in a circular-ring type layout. A stack of rings with gradually increasing/decreasing scales produces a cone, hence we term this generally as multi-oriented conical pyramid (MOCP). An EMOCP is created for each input template given as input gallery.

3.2b *Aircraft localization stage using extensive scan*: The overall framework for aircraft localization task (the testing phase) is shown in figure 3, which consists of the five phases, as discussed below:

(i) *Pre-processing*—This phase has divided into two parts:

- *Denoising*—The test image is denoised by the method proposed by Zhang *et al* [24], to remove noise from the image (see section 3.2a (i) for details).
- *Enhancement*—To further improve the quality of denoised image, the the process suggested [28] based on DCT + SVD is used.

(ii) *Salient region detection phase*—The concepts of GBVS [31], Spectral Saliency [37] and DBSCAN [41] are combined (see figure 6), and used to find the salient regions from the given satellite image. This phase consists of three modules:

- *Saliency map*—Saliency maps generated by GBVS and spectral saliency are thresholded and combined, using an intersection operation, to get a tentative saliency map (SM).
- *Clustering*—The SURF detector [9] is applied on the enhanced image to determine the keypoints. These keypoints are grouped using DBSCAN clustering to obtain the region of interests (ROI) as tentative foreground blobs.
- *Salient region selection*—If the percentage of the foreground area in SM is less than a threshold value (α), then the ROI and the SM are combined to get the final salient regions of interest. If the percentage of the foreground area (FGA) in SM is between a lower (α) and an upper thresholds (β), the SM is chosen as the final salient regions of interest. While, if the percentage is more than the upper threshold, the ROI gives the final salient regions of interest. Using the following heuristic, the final salient region of interests are determined as

$$SalientRegion = \begin{cases} SM \cup ROI & FGA < \alpha \\ SM & \alpha < FGA < \beta \\ ROI & FGA > \beta \end{cases}$$

(2)

where FGA denotes the foreground area of a blob (salient region).

(iii) *Candidate target region (CTR) detection*—The processed salient regions are further analyzed (see figure 7) to detect the presence of any clutter of aircrafts.

Insignificant (based on size) salient blobs are eliminated from further processing. For the rest of the blobs, in each case a minimal rectangular area covering the salient region of interest (RAOI—Rectangular Area of Interest) is formed to highlight the potential target(s). Then for segments with appreciably larger areas, Mean-shift segmentation [40] is performed on the regions containing a clutter
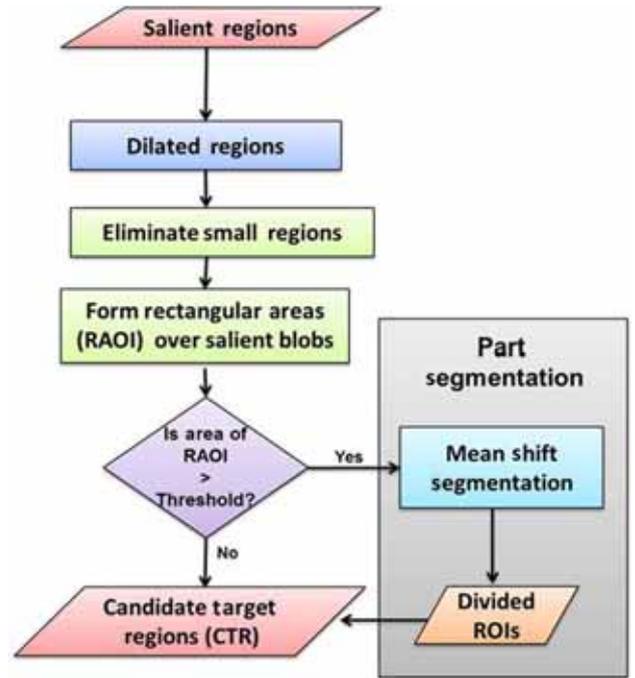

Figure 7. Detection of candidate target regions (CTR) for individual Aircraft Localization Task, using output from saliency region detection stage (see figure 6). See figure 3 for the overall framework.

Table 1. Optimal values of different heuristic parameters used for experimentation, estimated empirically, using the satellite images, to provide the best performance.

Process	Values of thresholds
GBVS	Intensity ≥ 40
Spectral saliency	Intensity ≥ 20
Removal of small blobs	Area (size) of blob ≤ 250pixels
Mean shift segmentation	Area of RAOI ≥ 100
Mean shift segmentation	Radius of spatial window $= 10$
Mean shift segmentation	Radius of color window $= 35$
Mean shift segmentation	Maximum level of pyramid $= 3$
Normalized cross correlation (NCC)	Threshold for matching $= 0.5$
Lower threshold (Eq. 2)	$\alpha = 25$
Upper threshold (Eq. 2)	$\beta = 45$
DBSCAN	No. of pixels ≤ 5000
Threshold based saliency	Intensity ≥ 190
K_m (Eq. 1)	9
λ (Eq. 1)	$6e^{-3}$
γ (Eq. 1)	$2e^3$
K [section 3.2a (i)]	20
L [section 3.2a (i)]	30

of aircrafts, to further subdivide the regions into smaller sub-regions containing a single aircraft. Parameters (spatial window radius, color window radius and maximum level of pyramid for segmentation) of mean-shift segmentation (see table 1, in
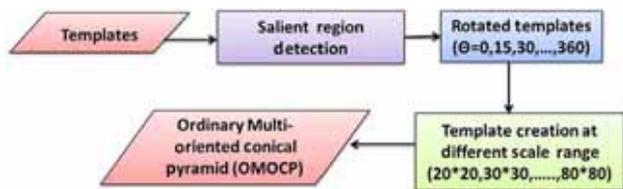
Figure 8. Ordinary multi-oriented conical pyramid (OMOCP) creation task, for the quick scan framework.
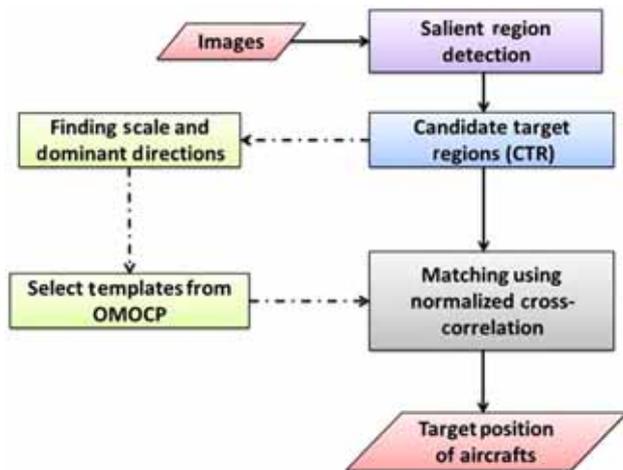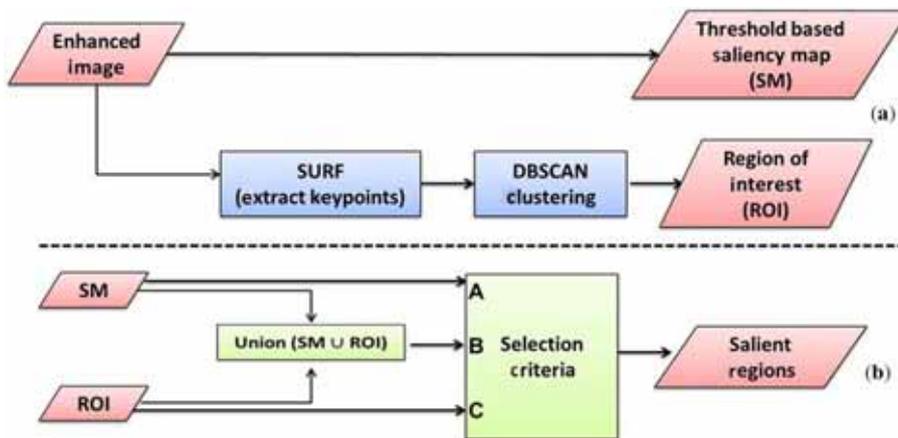

Figure 9. Overall quick scan framework for aircraft localization task.

section 4.2) are empirically set to produce smaller partitions of size (20 to 80 pixels wide) comparable to that in EMOCP. A rectangular area with minimal span covering any particular blob is termed as CTR (see figures 3 and 7). The aspect ratio (scale) and dominant orientation of each CTR produced by Mean-Shift segmentation, is used at the matching stage for aircraft detection using EMOCP.
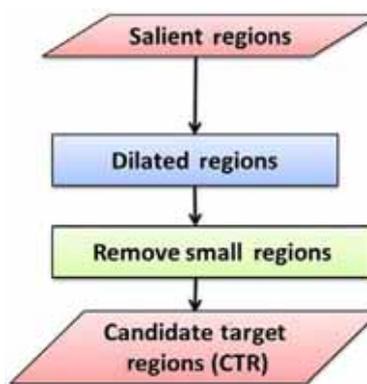

Figure 11. Detection of CTR for aircraft localization task for quick scan framework. See figure 9 for overall quick scan framework.

(iv) *Dominant direction and scale estimation*—The scale and orientation (dominant direction) for each candidate target region (CTR) is found using the following two steps: (a) Scale of each region is given by the 2-D spatial extent of that particular region, and (b) dominant direction (Orientation) is given by the direction of major axis of an ellipse, fit using Hough Transform on each CTR.

(v) *Matching Phase*—In this phase, for each CTR a subset of templates is selected from the EMOCP, based on the dominant direction and scale, estimated at the previous step. The templates selected for matching at this stage, range from the one scale below to one above the detected scale. Similarly, poses of templates are selected in EMOCP with one orientation step on either side of the detected orientation value. Thus for each salient region (CTR) detected as a potential candidate (in step (iii)) for a target aircraft, a set of nine templates (three orientations and three scales) are selected (out

Figure 10. Salient region detection stage for quick scan framework. The inputs to stage (**b**) are obtained from the outputs of stage (**a**). See figure 9 for overall Quick Scan framework.
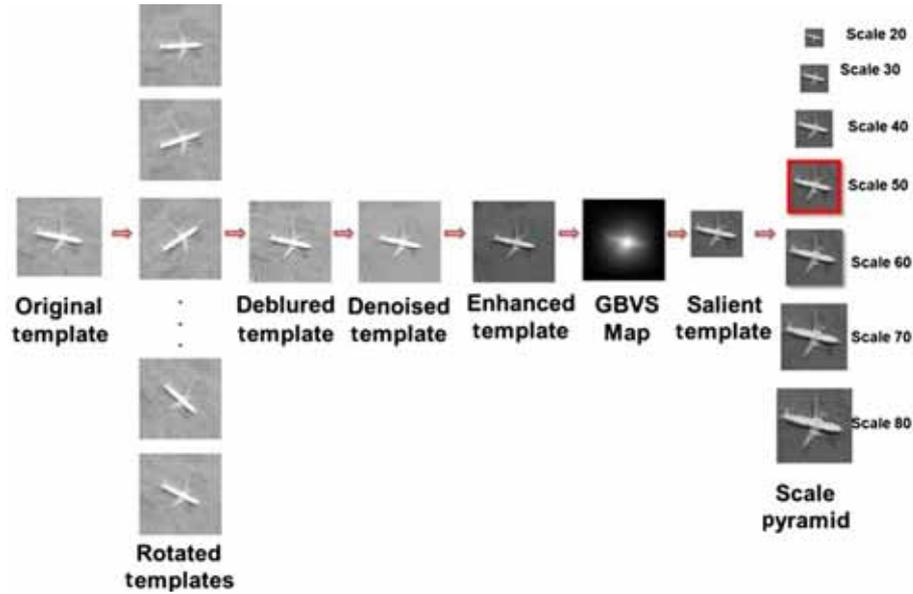
Figure 12. Processing of templates for generation of MOCP. The template highlighted in *red* is that with the same pose as the original template used in this task.
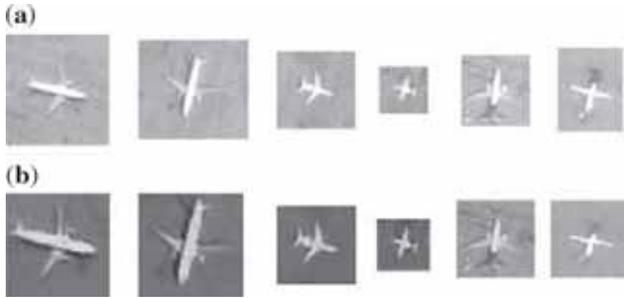
Figure 13. (**a**) A few samples of aircraft templates given as training gallery, to detect aircrafts in satellite scenes; (**b**) the post-processed respective templates of the same shown in (**a**), used for creating the EMOCP.

of 168 in synthetic training gallery, per template) from the EMOCP for matching. This speeds up the search process, as its space is reduced considerably. Finally, matching is done using the normalized cross correlation [10]. The NCC is computed as

$$\gamma(u,v) = \frac{\sum_{x,y}[f(x,y) - \bar{f}_{u,v}][t(x-u,y-v) - \bar{t}]}{\left\{\sum_{x,y}[f(x,y) - \bar{f}_{u,v}]^2 \sum_{x,y}[t(x-u,y-v) - \bar{t}]\right\}} \quad (3)$$

where f is the image, $\bar{f}_{u,v}$ is the mean of $f(x,y)$ in the region under the template, t is the template, and $\bar{t}$ is the mean of the template $t(x,y)$.

If there are T templates given as an input set of target aircrafts, each CTR is matched with only a set of $9 \times T$ templates using NCC (Eq. 3), and the best match is selected.

3.3 *Quick scan*

The proposed quick scan framework (computationally efficient and simplified) consists of two stages: (a) ordinary multi-oriented conical pyramid (OMOCP) generation, and (b) aircraft localization, which are discussed below:

3.3a *Ordinary multi-oriented conical pyramid (OMOCP) generation stage for quick scan:* In OMOCP generation stage, no pre-processing stage is used, as shown in figure 8. The overall framework for OMOCP generation consists of the following three phases: (a) saliency region detection phase, (b) rotation phase, and (c) scale pyramid formation phase. In salient region detection stage, the concept of threshold [46] based saliency is used instead of the GBVS and spectral saliency.

3.3b *Aircraft localization stage for quick scan:* This overall framework for aircraft localization task is shown in figure 9, which consists of the four phases, as discussed below:

(i) *Salient region detection phase*—In this phase, the image is thresholded to find the salient image map (SM) (see figure 10). Also, the grouping of the keypoints (detected using SURF detector) are performed using DBSCAN clustering technique, to obtain the ROIs. To select the final region of interests, a same heuristic is performed, as discussed in Extensive Scan framework [section 3.2b (ii)].

(ii) *Candidate target region detection*—The salient regions are processed furthr (as shown in figure 11) for further refinement purpose. This method (morphological process) produces the candidate target regions (CTR) for further processing.
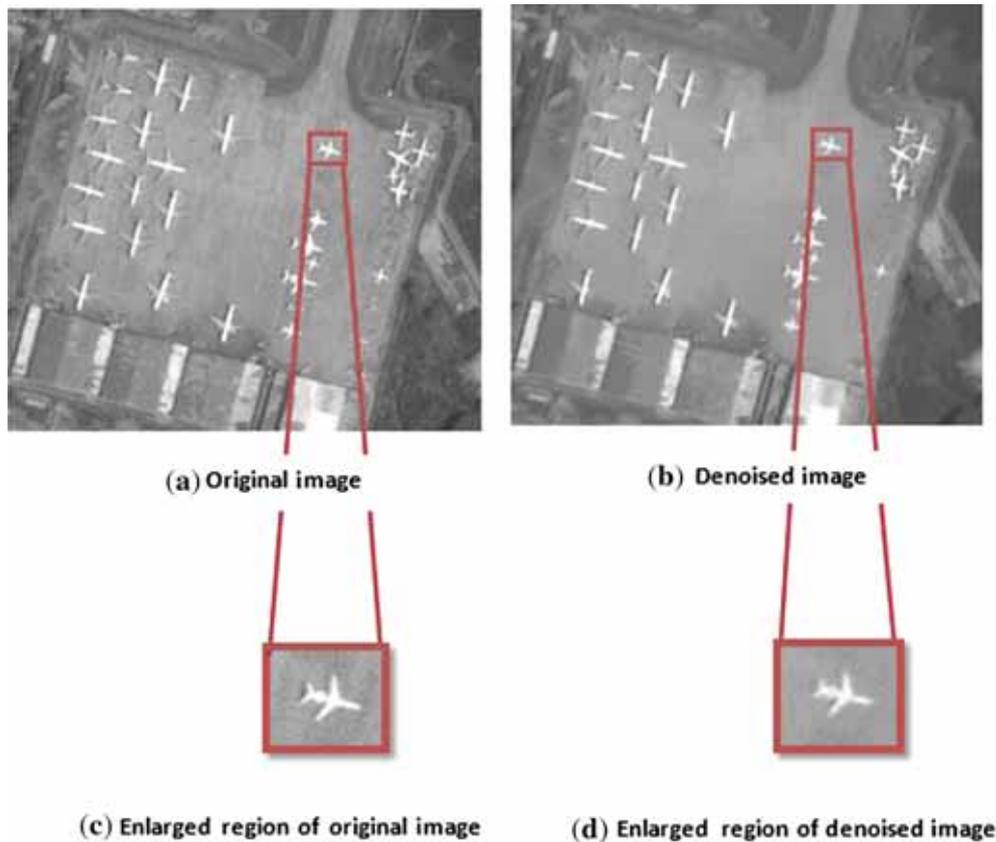
Figure 14. Intermediate result [in (**b**)] after applying the denoising technique on a given satellite image in (**a**). See figure 4 for the entire framework.

(iii) *Dominant direction and scale estimation*—The scale and orientation (dominant direction) for each CTRs are determined using the technique discussed in the case of Extensive Scan [section 3.2b (iv)].

(iv) *Matching phase*—In this phase, for each CTR the selected templates [see section 3.2b (v)] are matched with the detected CTRs using the normalized cross correlation (NCC) [10] (see Eq. 3). We have also experimented with distance transform DT, Ragnemalm [47]) based matching and found that the results give similar performance. Hence we recommend NCC as its complexity is less expensive.
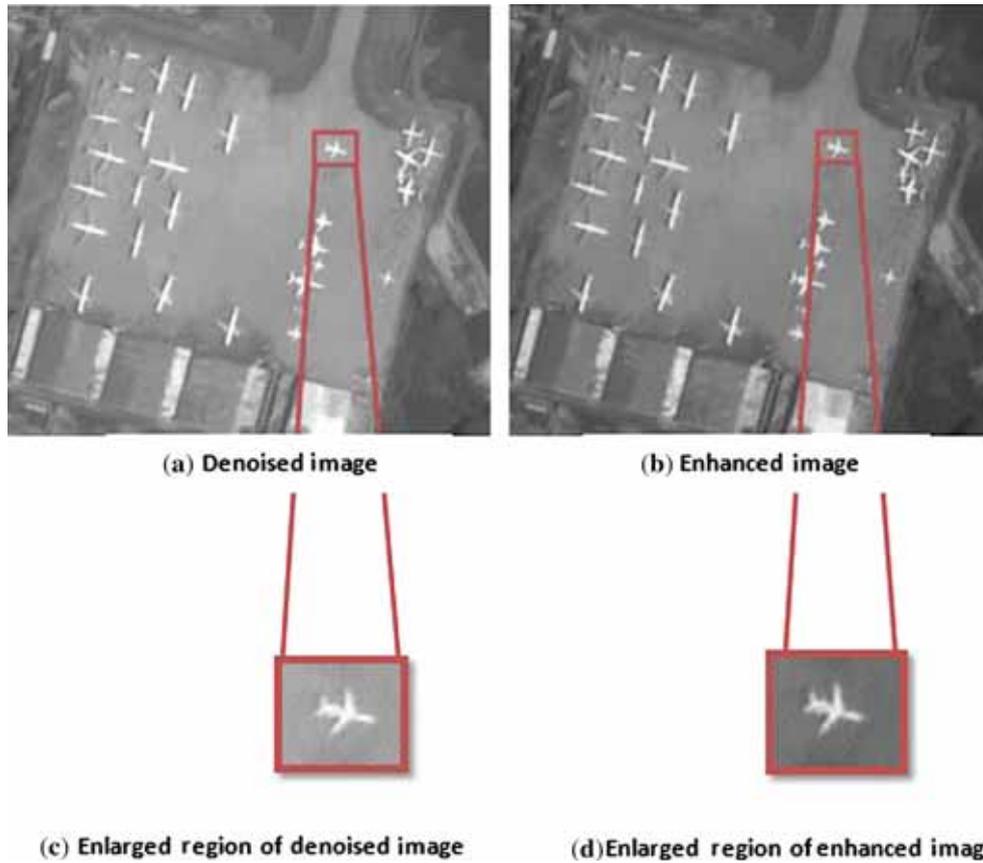
4. Intermediate results of processing

The experiments are conducted on a satellite image aircraft dataset, having a total of 30 images. Satellite images used as query, vary in size from 10 to 790 KB, containing 5–20 target samples. MOCP (EMOCP or OMOCP) generated for each of the 10 target samples (hand-cropped templates from query scenes), consists of synthetically generated 168 samples at seven different scales and 24 orientations. These templates also need pre-processing, as they are of low

contrast, similar to that in query scenes on which targets are identified. Else, we found by experimentation that the overall performance of the system degrades. Sections 4.1 and 4.2 discuss the results of intermediate steps results of MOCP generation and the aircraft localization task.

4.1 *Illustration of proposed MOCP*

Henceforth, we use MOCP, which refers to EMOCP for extensive scan and OMOCP for quick scan algorithms. MOCP is created for each training sample in gallery, and used later during the testing/matching stages. A template of a target is first rotated at intervals of 15 degrees to generate all the rotated templates (MOCP). Then, these rotated templates are processed using the deblurring module [see section 3.2a (ii)] to sharpen them. Further, the deblurred templates are denoised [see section 3.2a (ii)] and enhanced [see section 3.2a (ii)] to suppress the background information. Then, a salient region is found from the enhanced templates using GBVS. Based on these salient regions, new templates of the target (aircrafts) are created by cropping and masking of the background information. Finally, the scale pyramid is formed by stacking these salient templates. Figure 12 shows the intermediate results for the overall

(a) Denoised image (b) Enhanced image

(c) Enlarged region of denoised image (d) Enlarged region of enhanced image

Figure 15. Result for enhancement [in (**b**)] on the denoised image, as in figure 14(b). See figure 3 the entire framework.

MOCP generation process in extensive scan for one template. Rotated versions of the output shown in figure 12, are used to create the MOCP. The scale pyramid of rotated templates represents the multi oriented conical pyramid (MOCP) as shown in figure 5. Figure 13 shows a subset (6) of the 10 different samples of aircrafts, hand-cropped and selected for creating the MOCP. Figure 13(a) shows a few samples of templates obtained from satellite scenes, given as training gallery, to detect aircrafts in query satellite images, while figure 13(b) shows the respective post-processed templates of the same used for creating the MOCP (see figures 4, 5 and 8 for framework).

### 4.2 *Intermediate results of the aircraft localization stage*

For the aircraft localization task using extensive scan, the query or test image [see figure 14(a)] is first passed to the denoising module, to remove any noise present in an image. Figure 14(b) shows the output of the process of denoising. Figure 14(c) and (d) show the enlarged regions (using a target) of the images, to highlight the effect of denoising.
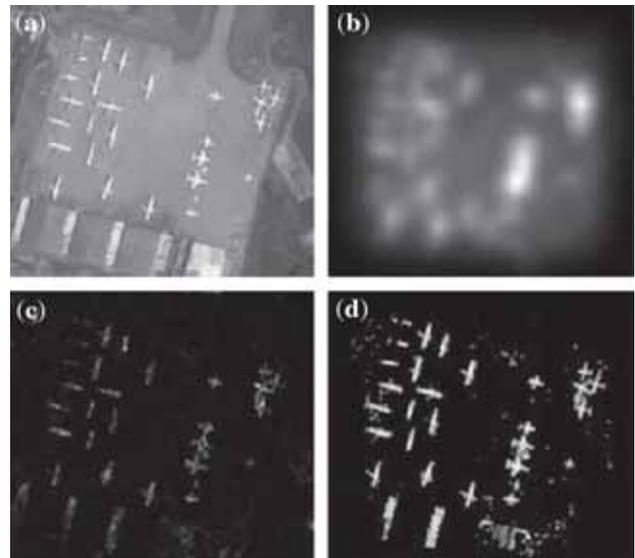

Figure 16. Results of saliency: (**a**) enhanced image as the output from figure 15(**b**). Stages after applying: (**b**) GBVS, (**c**) Spectral Saliency and (**d**) combined Saliency Map (SM). See figure 6(A) for the Saliency framework.
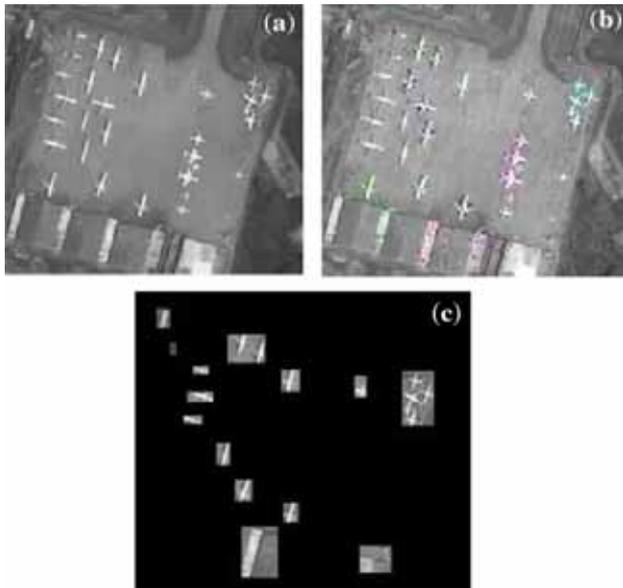
Figure 17. Results of saliency: (**a**) Enhanced image as the output from figure 15(**b**). Stages after applying: (**b**) DBSCAN Clustering (with 18 groups of clusters, each identified by a unique color), and (**c**) DBSCAN based ROI. See figure 6(A) for the saliency framework.
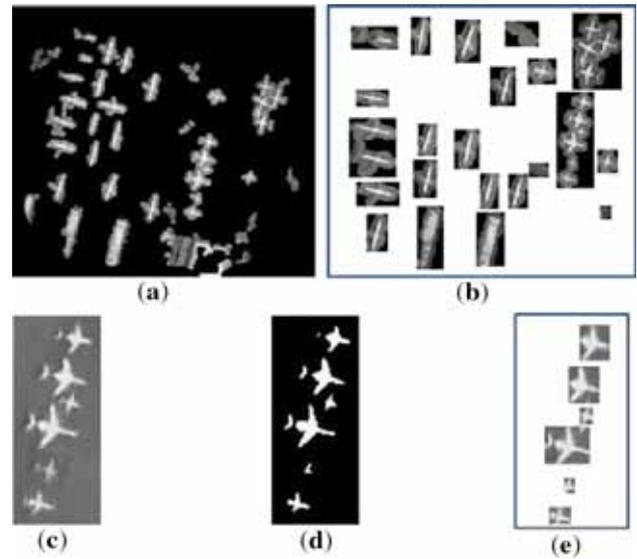
Figure 19. Post-processing of salient regions in (**a**) [as shown in figure 18(c)], to get the Candidate target regions (CTR). (**b**) Rectangular areas of interest (RAOI), formed over image in (**a**); (**c**) shows a zoomed in RAOI cropped (on the *right part*) from (**b**), being used as an input to the post-processing stage. (**d**) Results of the mean shift segmentation on the cropped scene shown in (**c**), and (**e**) candidate target regions (CTR), obtained from (**c**) and (**d**) using a minimum enclosed rectangular RAOI.
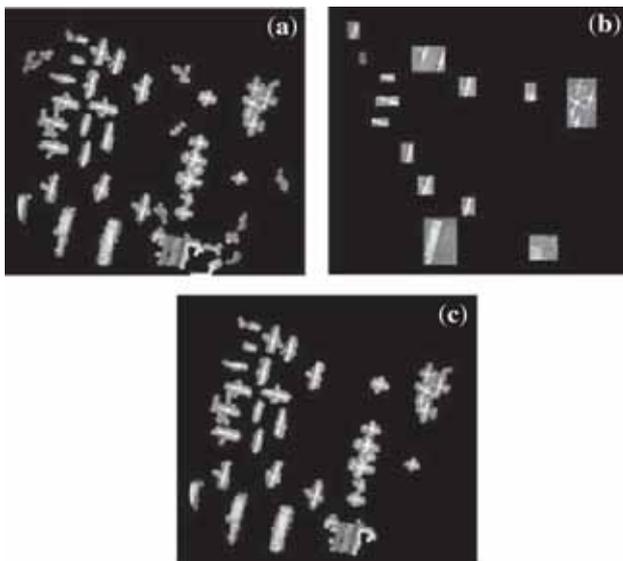
Figure 18. Combination of the outputs of the Saliency and Clustering methods [see figure 6(B)]. Results are shown as: (**a**) Combined Saliency Map [from figure 16(d)], (**b**) DBSCAN based ROI [from figure 17(c)], (**c**) Final Salient regions obtained from (a) and (b). See Figure 6(B) for detailed framework.

The denoised image [figure 15(a)] is then enhanced further to highlight the foreground targets with respect to background layer, followed by gamma correction. Figure 15(b) depicts the results of enhancement. Figures 15(c) and (d) show the enlarged regions (using a

target) of the images, to highlight the effect of enhancement after denoising (see section 3.2a (i) for details).

In the proposed extensive scan framework, the salient regions of interest are obtained by combining the outputs of two saliency (unsupervised) algorithms: graph based visual saliency [31] and spectral-based saliency [36], applied on the enhanced image [shown in figure 15(b), and same again in figure 16(a)]. The saliency maps obtained by graph based visual saliency [31] and spectral-based saliency [36] are shown in figures 16(b) and (c) respectively. Figure 16(d) shows the combined saliency map (SM) obtained from the images of figures 16(b) and (c), using the intersection operation [see flowchart in figure 6(A) and details in section 3.2b (ii)].

In a parallel process, DBSCAN is used for clustering the keypoints obtained using the SURF detector. For the same enhanced image shown again in figure 17(a), figure 17(b) shows the result of DBSCAN (unsupervised process), where each color denotes a group of pixels belonging to one particular cluster. Figure 17(c) shows the result of DBSCAN based ROI, obtained after thresholding the number of pixels in a clustered region. [see figure 6(A) for flowchart and details in section 3.2b (ii)]

Figures 18(a) and (b) show the results of the combined saliency map (SM) and DBSCAN based ROI, as shown earlier in figures 16(d) and 17(c) respectively. Note that the insignificantly small blobs appearing (as noise) in figure 16(d) are removed and these do not appear in figure 18(a). These two maps [saliency maps (SM) and DBSCAN based
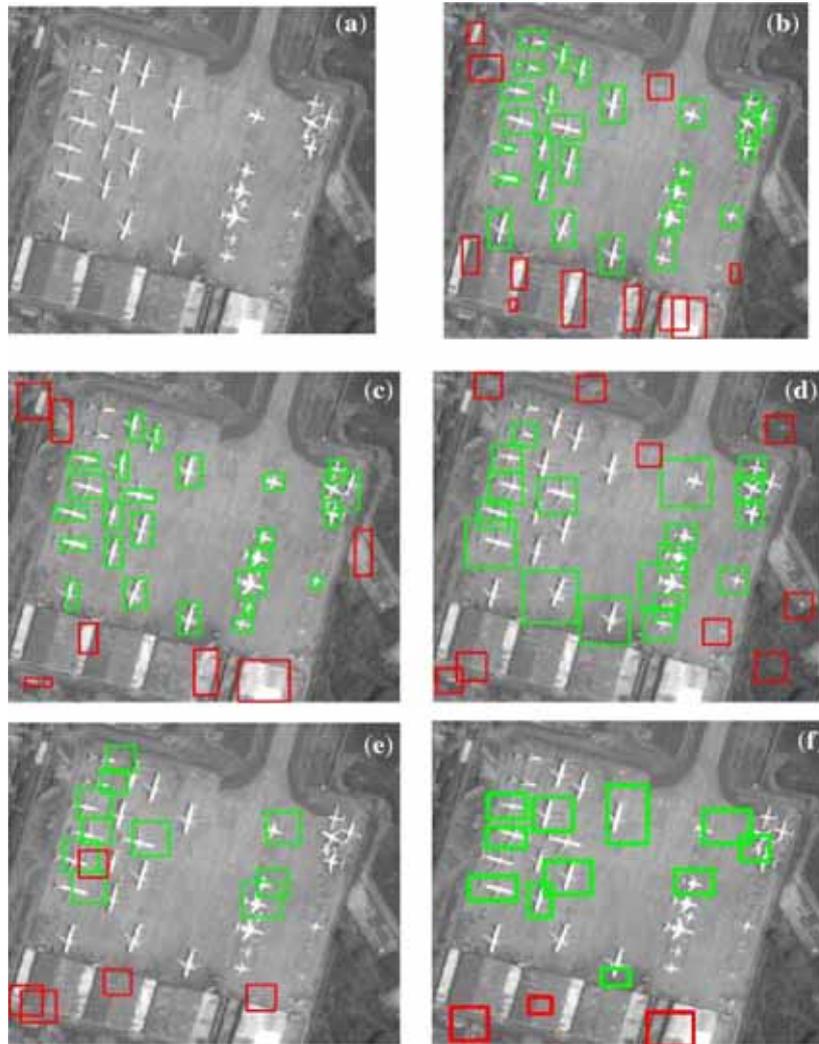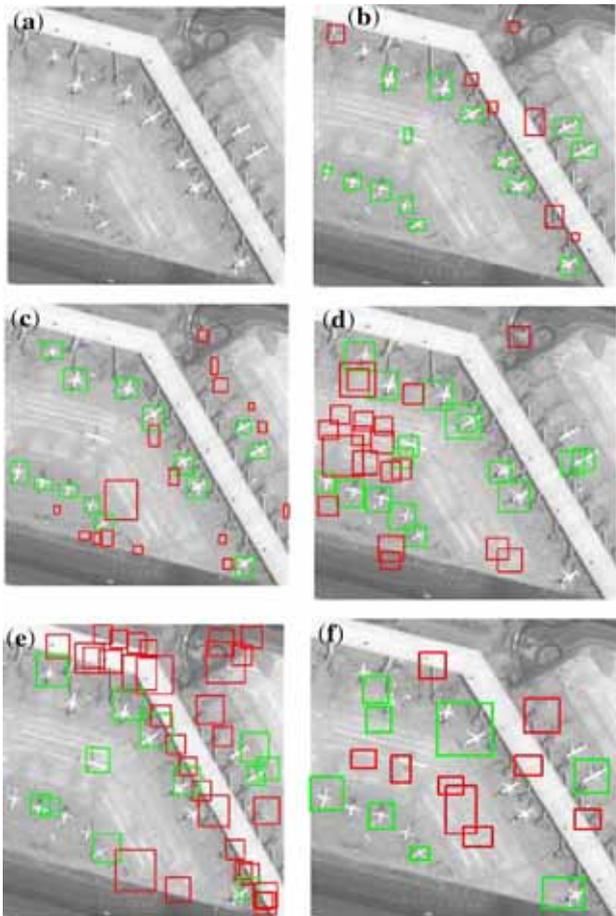
Figure 20. Results (I) for targets detected, shown for visual comparative study, (**a**) query satellite image, (**b**) proposed extensive scan method, (**c**) proposed quick scan method, (**d**) Haar + ADABOOST [11], (**e**) HOG + ADABOOST and (**f**) SURF [9]. Detected targets are shown using *green rectangular boxes*, while false alarms are shown using *red boxes*.

(ROI)] are combined [see figure 6(B) and Eq. 2] to obtain the final salient regions as shown in figure 18(c).

These salient regions (shown in figure 19(a) again) are further processed using morphological operations, to eliminate small spurious blobs appearing as false alarms. The rest of the process to create CTRs as isolated candidate target regions for aircrafts to be matched has been depicted earlier in figure 7 [details given in section 3.2b (iii)]. Connected components obtained from the salient regions shown in figure 19(a) are grouped to form tight rectangular bounding boxes, called the rectangular area of interest (RAOI). Figure 19(b) shows the RAOIs, formed using the salient region in figure 19(a) [see figure 7 and details in section 3.2b (iii)]. Some of these RAOIs shown in figure 19(b) may contain clutters of aircrafts. These individual aircrafts are separated using area as a feature, by mean shift segmentation operation [40] (see figure 7 and section 3.2b

(iii) for details) and a morphological process, as shown in figure 19(c). This processing is shown by highlighting a cropped area of the result in figure 19(b), which is shown at a larger (zoomed in) scale in figure 19(c). Figure 19(c) shows an example of an isolated candidate area [from figure 19(b)] containing a clutter of aircrafts, from which individual aircrafts must be isolated as separate blobs. Figure 19(d) shows the isolated segments of individual aircrafts separated by mean shift segmentation. Figures 19(e) shows the results of the final stage of processing, which create the candidate target regions (CTR) as potential aircraft targets (each in isolation). These isolated segments of potential aircraft targets are individually matched with a selected set of templates in MOCP, based on scale and dominant direction estimated from these segments [see section 3.2b (iv)–(v)]. The codes used for denoising and deblurring used in the pre-processing stage
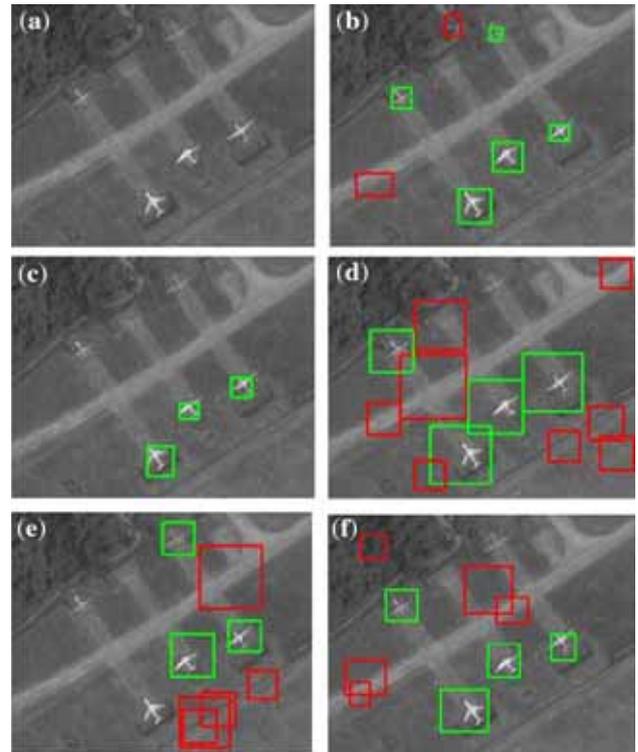
Figure 21. Results (II) for targets detected, shown for visual comparative study, (**a**) query satellite image, (**b**) proposed extensive scan method, (**c**) proposed quick scan method, (**d**) Haar + ADABOOST [11], (**e**) HOG + ADABOOST and (**f**) SURF [9]. Detected targets are shown using *green rectangular boxes*, while false alarms are shown using *red boxes*.

Figure 22. Results (III) for targets detected, shown for visual comparative study, (**a**) query satellite image, (**b**) proposed extensive scan method, (**c**) proposed quick scan method, (**d**) Haar + ADABOOST [11], (**e**) HOG + ADABOOST and (**f**) SURF [9]. Detected targets are shown using *green rectangular boxes*, while false alarms are shown using *red boxes*.

[see section 3.2a (i) and 3.2b (i)] are obtained from the authors of [24] and [25] respectively. The different values of the heuristic parameters (optimal values obtained empirically) used for experimentation are given in table 1.

5. Performance analysis

For performance analysis of the proposed frameworks (quick and extensive scan) for localization of aircrafts from satellite images, a set of 30 images are acquired from CARTOSAT-2A satellite (Sensor: PAN). The images are of 0.8 m/pixel ground resolution. The pixel resolution (size) of images scenes vary from (152×150) to (638×639), with all in monochrome. Set of library templates as targets are also retrieved from the scenes, as specified by the user (in total 12 aircrafts were specified). These images have been provided by DTRL (DRDO) Lab, Delhi, India, for research purposes. For qualitative analysis using visual

illustrations, results of detected aircrafts in satellite images using the proposed frameworks and other three state-of-the-art techniques: (Haar + Adaboost [11], HOG + Adaboost [48] and SURF based matching, are shown in figures 20–22). The code of [11] is a near implementation of the method proposed in [11]. Detected targets are highlighted using rectangular boxes. Green boxes denote the correct targets detected (true positives), while the erroneous results (false positives) are marked using red boxes. Our proposed framework produces few erroneous results and more accurate detections (see figures 20–22) as compared to other state-of-the-art techniques. The output of the proposed method in figures 20–22(b) misses only a very few targets (false negatives)—none in figure 20 and 22(b), while only one target is missed in the scene in figure 21(b). Based on empirical results we have verified that our method outperforms other competing methods. Since the targets are very small in size and have low-resolution, a proper preprocessing technique enhances the quality of the image which in turn enhances the performance. Detecting articulations and edges from these small targets to derive higher order features (SIFT, SURF) was also attempted. However, due to low image quality and noise these features do not give a robust, comprehensive representation of the target
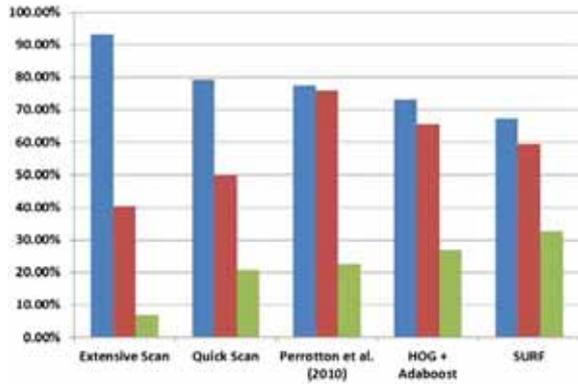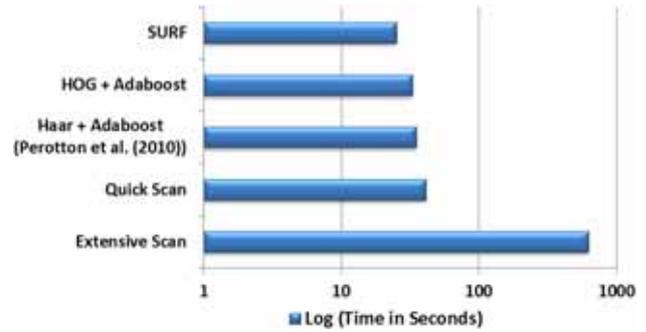
Figure 23. Quantitative comparison of the performance of proposed frameworks (extensive scan and quick scan), with Haar like features [11] and HOG feature with Adaboost (similar to [11]) and SURF [9], averaged over 30 satellite images. (*TP* true positives, *FP* false positives, *FN* false negatives).

Figure 24. Comparison of the execution time (in seconds, on LOG scale) of the proposed frameworks (extensive scan and quick scan), with Haar like feature [11], HOG feature with Adaboost (similar to [11]) and SURF [9]. Time is estimated by averaging over 30 images.

and performance fails. However, our method relies on the low-level enhancement (processing) of the images, and a template-matching process using a sliding window protocol using probable candidate templates. Other modules of saliency detection, scale and dominant direction selection reduces the search space and gives less opportunity for false matches. This boosts the performance of our proposed method over other competing methods.

To validate the effectiveness of the proposed frameworks for aircraft localization, we compared our results with a technique proposed by Perrotton *et al* [11], using true positives (TP) or correct identifications, false positives (FP) or false alarms and false negatives (FN) or missed targets. For details of this metric, refer Powers [49]. Also, we have used the HOG feature [48] instead of Haar like features as in the method proposed in Perrotton *et al* [11] for comparison. The results are further compared with a process based on SURF [9] based matching. The bar-charts of TP, FP and FN values given in figure 23 are averaged over 30 images, having a total of 116 targets. Proposed frameworks (quick scan and extensive scan) perform significantly better as compared to the other techniques. The proposed extensive scan framework performs much better than the proposed quick scan method in terms of TP, FP and FN values. Both the proposed frameworks have the largest TP and least FP and FN values, compared to others. It is due to the presence of efficient pre-processing modules, an efficient saliency module and MOCP representation, that the extensive scan framework performs the best.

However, in terms of time complexity the extensive scan framework takes around 500–600 s, while quick scan takes around 30–50 s (see figure 24). Results have been averaged over 30 images of resolution varying from 152×150 to 638×639 pixels. The time taken by quick scan framework is comparable with the other state-of-the art techniques (see figure 24). All the algorithms are processed on a system having a following configuration: Intel core i7-3770,

3.4 GHz processor, 32 GB RAM and Windows 7 (64 bit) Operating System.

6. Conclusion

This work proposes a novel aircraft detection framework for satellite images. The proposed techniques outperform all the recent techniques by a considerable margin. The data we have analyzed is in optical band and free of cloud cover, fog, etc. Hence the process will not work for data under cloud or fog cover (or those acquired from different sensors—SAR NIR, etc.). To address the problem of aircraft localization, saliency-based technique was used to reduce the search space. The proposed EMOCP or OMOCP representations of the target aircrafts help to improve the effectiveness of the automatic target localization system by combining saliency and clustering. Given the low processing time, the proposed quick scan method can be used as an online solution for detection of aircrafts in satellite images.

References

[1] Chen X, Xiang S, Liu C-L and Pan C-H 2014 Vehicle detection in satellite images by hybrid deep convolutional neural networks. *Geosci. Remote Sens.* 11(10): 1797–1801

[2] Li Z and Itti L 2011 Saliency and gist features for target detection in satellite images. *IEEE Trans. Image Process.* 20(7): 2017–2029

[3] Perrotton X, Sturzel M and Roux M 2008 *Automatic object detection on aerial images using local descriptors and image synthesis*. Berlin: Computer Vision Systems, Springer

[4] Sun H, Sun X, Wang H, Li Y and Li X 2012 Automatic target detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model. *Geosci. Remote Sens. Lett.* 9(1): 109–113

[5] Liu M-Y, Tuzel O, Veeraraghavan A and Chellappa R 2010 Fast directional chamfer matching. In: *Computer Vision and Pattern Recognition*

[6] Olson C F and Huttenlocher D P 1997 Automatic target recognition by matching oriented edge pixels. *IEEE Trans. Image Process.* 6(1): 103–113

[7] Villamizar M, Moreno-Noguer F, Andrade-Cetto J and Sanfeliu A 2010 Efficient rotation invariant object detection using boosted random ferns. In: *Computer Vision and Pattern Recognition*

[8] Lowe D G 2004 Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* 60(2): 91–110

[9] Bay H, Tuytelaars T and Van Gool L 2006 Surf: Speeded up robust features. In *European Confernce on Computer Vision*

[10] Briechle K and Hanebeck U D 2001 Template matching using fast normalized cross correlation. In: *Aerospace/Defense Sensing, Simulation, and Controls*. International Society for Optics and Photonics

[11] Perrotton X, Sturzel M and Roux M 2010 Implicit hierarchical boosting for multi-view object detection. In: *Computer Vision and Pattern Recognition*

[12] Si X, Xuemin H and Hong Z 2010 A novel object detection approach based on the boundary shape information from high resolution satellite imagery. In: *WSEAS Transactions on Computers*, vol. 9, pp. 455–464

[13] Cai F, Chen H and Ma J 2011 Man-made object detection based on texture clustering and geometric structure feature extracting. *Int. J. Inform. Technol. Comput. Sci.* 3(2): 9–16

[14] Li W, Xiang S, Wang H and Pan C 2011 Robust airplane detection in satellite images. In: *18th IEEE International Conference on Image Processing (ICIP), 2011*, pp. 2821–2824

[15] Xu S, Fang T, Li D and Wang S 2010 Object classification of aerial images with bag-of-visual words. *IEEE Geosci. Remote Sens. Lett.* 7(2): 366–370

[16] Liu G, Sun X, Fu K and Wang H 2013 Aircraft recognition in high-resolution satellite images using coarse-to-fine shape prior. *IEEE Geosci. Remote Sens. Lett.* 10(3): 573–577

[17] Cheng M-M, Zhang Z, Lin W-Y and Torr P 2014 Bing: Binarized normed gradients for objectness estimation at 300 fps. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3286–3293

[18] Wu H, Zhang H, Zhang J and Xu F 2015 Fast aircraft detection in satellite images based on convolutional neural networks. In: *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 4210–4214

[19] Chen X, Xiang S, Liu C-L and Pan C-H 2013 Aircraft detection by deep belief nets. In: *2013 2nd IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 54–58. IEEE

[20] Zhang D, Han J, Cheng G, Liu Z, Bu S and Guo L 2015 Weakly supervised learning for target detection in remote sensing images. *Geosci. Remote Sens. Lett.* 12(4): 701–705

[21] Felzenszwalb P F, Girshick R B, McAllester D A and Ramanan D 2010b Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.* 32: 1627–1645

[22] Malisiewicz T, Gupta A and Efros A A 2011 Ensemble of exemplar-svms for object detection and beyond. In: *International Conference on Computer Vision*

[23] Mishne G, Talmon R and Cohen I 2015 Graph-based supervised automatic target detection. *IEEE Trans. Geosci. Remote Sens.* 53(5): 2738–2754

[24] Zhang L, Dong W, Zhang D and Shi G 2010 Two-stage image denoising by principal component analysis with local pixel grouping. *Pattern Recognit.* 43(4): 1531–1549

[25] Xu L, Zheng S and Jia J 2013 Unnatural l0 sparse representation for natural image deblurring. In: *Computer Vision and Pattern Recognition*

[26] Dong W, Zhang D, Shi G and Wu X 2011 Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Trans. Image Process.* 20(7): 1838–1857

[27] Bhandari A, Kumar A and Singh G 2012 Svd based poor contrast improvement of blurred multispectral remote sensing satellite images. In: *International Conference on Computer and Communication Technology*

[28] Bhandari A, Kumar A and Padhy P 2011 Enhancement of low contrast satellite images using discrete cosine transform and singular value decomposition. In: *World Academy of Science, Engineering and Technology*

[29] Buades A, Coll B and Morel J-M 2005 A non-local algorithm for image denoising. In: *Computer Vision and Pattern Recognition*

[30] Gu Y, Wang C, Wang S and Zhang Y 2011 Kernel-based regularized-angle spectral matching for target detection in hyperspectral imagery. *Pattern Recognit. Lett.* 32(2): 114–119

[31] Harel J, Koch C and Perona P 2006 Graph-based visual saliency. In: *Advances in Neural Information Processing Systems*

[32] Rother C, Kolmogorov V and Blake A 2004 Grabcut: Interactive foreground extraction using iterated graph-cuts. *ACM Trans. Graph.* 23(3): 309–334

[33] Yang C, Zhang L, Liu H, Ruan X and Yang M-H 2013 Saliency detection via graph-based manifold ranking. In: *Computer Vision and Pattern Recognition*

[34] Zhu W, Liang S, Wei Y and Sun J 2014 Saliency optimization from robust background detection. In *Computer Vision and Pattern Recognition*

[35] Zhu D, Wang B and Zhang L 2015 Airport target detection in remote sensing images: A new method based on two-way saliency. *Geosci. Remote Sens. Lett.* 12(5): 1096–1100

[36] Schauerte B and Stiefelhagen R 2012a Predicting human gaze using quaternion dct image signature saliency and face detection. In: *IEEE Workshop on Applications of Computer Vision*

[37] Schauerte B and Stiefelhagen R 2012b Quaternion-based spectral saliency detection for eye fixation prediction. In: *European Confernce on Computer Vision*

[38] Cheng M-M, Zhang G-X, Mitra N J, Huang X and Hu S-M 2011 Global contrast based salient region detection. In: *Computer Vision and Pattern Recognition*

[39] Goferman S, Zelnik-Manor L and Tal A 2012 Context-aware saliency detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 34(10): 1915–1926

[40] Comaniciu D and Meer P 2002 Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 24(5): 603–619

[41] Ester M, Kriegel H-P, Sander J and Xu X 1996 A density-based algorithm for discovering clusters in large spatial databases with noise. In: *International Conference on Knowledge Discovery and Data Mining*

[42] Canny J 1986 A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* (6): 679–698

[43] Felzenszwalb P F, Girshick R B, McAllester D and Ramanan D 2010a Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.* 32(9): 1627–1645

[44] Belongie S, Malik J and Puzicha J 2002 Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* 24(4): 509–522

[45] Bennamoun M and Mamic G J 2012 *Object recognition: Fundamentals and case studies*. Berlin: Springer

[46] Gonzalez R C and Woods R E 2009 *Digital image processing*. Upper Saddle River, NJ: Pearson Education

[47] Ragnemalm I 1993 The euclidean distance transform in arbitrary dimensions. *Pattern Recognit. Lett.* 14(11): 883–888

[48] Dalal N and Triggs B 2005 Histograms of oriented gradients for human detection. In: *Computer Vision and Pattern Recognition*

[49] Powers D M 2011 Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation (tech. rep.). *Adelaide, Australia*