Journal of
Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

Robust aircraft segmentation from very high-resolution images based on bottom-up and top-down cue integration

Feng Gao
Qizhi Xu
Bo Li

SPIE.

Robust aircraft segmentation from very high-resolution images based on bottom-up and top-down cue integration

Feng Gao,[a,b] Qizhi Xu,[a,*] and Bo Li[a]
[a]Beihang University, School of Computer Science and Engineering, Beijing Key Laboratory of Digital Media, 37 Xueyuan Road, Beijing 100191, China
[b]Ocean University of China, Department of Computer Science and Technology, 238 Songling Road, Qingdao 266100, China

Abstract. Existing segmentation methods require manual interventions to optimally extract objects from cluttered background, so that they can hardly work well in automated surveillance systems. In order to automatically extract aircrafts from very high-resolution images, we proposed a segmentation method that combines bottom-up and top-down cues. Three essential principles from local contrast, global contrast, and center bias are involved to compute bottom-up cue. In addition, top-down cue is computed by incorporating aircraft shape priors, and it is achieved by training a classifier from a rich set of visual features. Iterative operations and adaptive fitting are designed to get refined results. Experimental results demonstrated that the proposed method can provide significant improvements on the segmentation accuracy. [DOI: 10.1117/1.JRS.10.016003]

Keywords: aircraft segmentation; bottom-up model; top-down model; GrabCut.

Paper 15660L received Oct. 1, 2015; accepted for publication Dec. 21, 2015; published online Jan. 19, 2016.

1 Introduction

Very high resolution (VHR) remote sensing images have been available at an astonishing ground sampling distance of 0.5 m from Google Earth. The applications of VHR images for ground object detection[1,2] and segmentation[3–6] have become attractive fields of research. Among that research, aircraft segmentation plays an important role in remote sensing image interpretation. For instance, it provides an important clue for aircraft recognition, since the colors of military aircraft and civil aircraft are generally different. In addition, aircraft segmentation can even contribute a lot to aircraft velocity estimation, considering that the contours of moving targets are deformed because of the relative movement between the sensor and the target.[7] Therefore, more attention should be paid to automatic aircraft segmentation.

In this paper, we concentrate on the problem of aircraft segmentation from VHR aerial images. To be specific, we particularly consider an airport dynamic surveillance environment, from which an aircraft detector would provide us with foreground windows. In these windows, the pose of an aircraft has been estimated, and then we can automatically perform aircraft segmentation.

Research in visual attention has demonstrated that both low-level visual stimulus and high-level conceptual knowledge related to the desired object determines the segmentation results.[8] Inspired by the visual attention studies, bottom-up and top-down cues have been applied for object segmentation. Many object segmentation methods are based on bottom-up cues, and these methods identifying object regions using low-level visual stimulus extracted from the image, such as intensity, color, texture, etc. Conversely, some object segmentation methods are based on top-down cues, and these methods use learnt statistical high-level knowledge of the desired object to identify object regions.

*Author all correspondence to: Qizhi Xu, E-mail: qizhi@buaa.edu.cn

1931-3195/2016/$25.00 © 2016 SPIE

Inspired by visual attention studies, bottom-up cues are widely applied for object segmentation from remote sensing images. Xu et al.[3] presented a level-set-based multiscale method by using wavelet feature for object segmentation. Zhang and Yang[4] proposed a target segmentation method by using frequency domain analysis. Wu and An[5] proposed an object segmentation method based on texture distribution from aerial images. In addition, other methods[9,10–13] based on color and contrast analysis are proposed to segment salient objects. These methods have demonstrated impressive results. However, it is widely acknowledged that bottom-up process alone can hardly extract object real boundaries, since object is often comprised of heterogeneous textures and colors. Especially in some cases when backgrounds are highly cluttered, these methods may respond to many unrelated visual stimuli, and therefore miss the desired objects.

On the other hand, research has explored top-down cues to incorporate prior knowledge for specific object segmentation. Liu et al.[6] presented a shape-based global minimization model for aircraft segmentation. Xu et al.[14] proposed a method for inshore ship segmentation by incorporating shape priors. In addition, in order to build appropriate shape priors of interested objects, Rother et al.[15] proposed GrabCut (GC) for interactive object segmentation. Incorporating prior knowledge can significantly improve the accuracy and reliability of the segmentation result. However, there remains an open problem of how to define an appropriate prior appearance model. In addition, manual interventions of GC-based methods are inconvenient for large scale image processing, and then limit applications of these methods.

To address the inconvenience of manual intervention, we proposed a two-stage automatic aircraft segmentation method for high resolution remote sensing images. In the first stage, three essential principles from local contrast, global contrast, and center prior are considered to compute bottom-up cues. Top-down cue is computed by incorporating aircraft shape priors, and it is achieved by training a classifier from a set of visual features. In the second stage, both bottom-up and top-down cues are combined to generate an integrated cue for GC initialization. Iteration operation and adaptive fitting are applied to get refined segmentation results.

The proposed method can effectively suppress background clutter. In this paper, background clutter includes shadows, oil spots, boarding bridges, and catering vehicles. Shadows and oil spots are high contrast regions. We use supervised visual feature learning (top-down cue computation) to remove shadows and oil spots. On the other hand, boarding bridges and catering vehicles are with similar visual features with the desired aircraft regions. Top-down process alone can hardly get satisfactory segmentation results. It needs the assistance of contrast analysis (bottom-up cue) to enhance the segmentation accuracy.

There are mainly two advantages of the proposed method: First, the combination of bottom-up and top-down cues provides good performance. Bottom-up cue favors contrast, whereas the top-down cue prefers shape priors. The integrated cue takes advantage of both cues and therefore is robust to shadow and background clutters. Second, it is of high practical value since it can extract aircraft automatically and accurately. Existing segmentation methods require manual intervention to optimally segment objects from cluttered background. The proposed method can automatically compute an integrated cue for iteratively GC initialization, and therefore is more convenient. Experimental results demonstrate that our method performs favorably against other closely related methods.

2 Methodology

The overview of our method is shown in Fig. 1. This method includes two steps, i.e., the bottom-up and top-down cue computation, cue integration, and refinement.

2.1 *Bottom-Up Cue Computation*

Perceptual research has revealed that the most influential factor in bottom-up visual stimulus is contrast. Therefore, we define a function to get an aircraft's bottom-up cue based on multiscale contrast. Three simple but essential principles are used: local contrast, global contrast, and center prior. To be specific, local contrast principle means a region of higher contrast with its neighbors

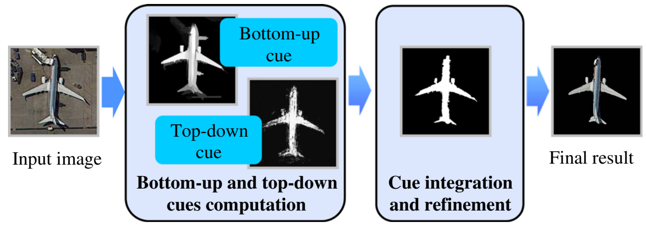

Fig. 1 Framework of the proposed aircraft segmentation method. First, bottom-up cues are computed based on multiscale contrast, top-down cues are computed based on supervised learning of visual features. Then, both cues are integrated for GC initialization, and iterative operation is applied to get refined segmentation results.

should have higher value. Global contrast principle means a region of higher contrast with all other regions should have higher value. Center prior principle means a region closer to the image center is more likely to belong to aircraft and should be emphasized.

An input image is oversegmented into superpixels at M scales. Graph-based multiscale segmentation algorithm[16] is used. On the scale m, an image is fragmented into regions $\{r_i^m\}$, $i = 1, \ldots, N^m$, where N^m is the number of regions. Given a region r_i^m and its neighboring regions $\{n_j^m\}$, $j = 1, \ldots, N_i^m$, where N_i^m is the number of its neighbors. Inspired by the saliency function in Ref. 17, local contrast measure of r_i^m is defined as

$$\mathrm{lc}(r_i^m) = \sum_{j=1}^{N_i^m} \omega_{ij}^m h[\mathrm{DR}(r_i^m, n_j^m)], \tag{1}$$

where ω_{ij}^m is the ratio between the area of n_j^m and the total area of the neighbors of $\{n_j^m\}$. $\mathrm{DR}(r_i^m, n_j^m)$ is the color histogram distance between two regions. The function $h(\rho)$ is defined as

$$h(\rho) = -\log(1-\rho). \tag{2}$$

It should be noted that we use logarithm functions to emphasize regions with higher color contrast. However, when two regions are with very similar colors, logarithm functions may generate the negative values. The function $h(\rho)$ is used to ensure a positive output. Through the computation of $\mathrm{lc}(r_i^m)$, regions of high contrast with their neighbors are emphasized.

In addition, in order to emphasize regions of high global contrast, function $\mathrm{gc}(r_i^m)$ is defined as

$$\mathrm{gc}(r_i^m) = \sum_{j=1}^{N^m} |r_i^m| \cdot \left\| \mathrm{mc}_i^m - \mathrm{mc}_j^m \right\|_2, \tag{3}$$

where mc_i^m and mc_j^m denote the mean color of r_i^m and r_j^m, respectively. $|r_i^m|$ denotes the number of pixels in r_i^m, and bigger regions with high global color contrast are emphasized.

In order to emphasize regions closer to the image center, the computation of center prior is defined as

$$g(x,y) = \exp[-(x-x_c)^2/(2\sigma_x^2) - (y-y_c)^2/(2\sigma_y^2)], \tag{4}$$

where σ_x and σ_y are set as one third of the width and height of the image, respectively. (x_c, y_c) is the image center. Hence, regions closer to the image center are assigned higher values.

As mentioned above, we use local contrast, global contrast, center prior as three essential principles for aircraft's bottom-up cue computation. After the computation of three local contrast, global contrast, and center prior measures, they are integrated as follows:

$$f(r_i^m) = \mathrm{lc}(r_i^m) \times \mathrm{gc}(r_i^m) \times g(x,y). \tag{5}$$

Finally, we get bottom-up cue by propagate contrast values from multiple regions to pixels. Here we use the function defined in Ref. 17, and the bottom-up cue is defined as

$$C_b(p) = \frac{\sum_{m=1}^{M} \sum_{i=1}^{N^m} f(r_i^m)(\left\|I_p - mc_i^m\right\|_2 + \varepsilon)^{-1} \delta(p \in r_i^m)}{\sum_{m=1}^{M} \sum_{i=1}^{N^m} (\left\|I_p - mc_i^m\right\|_2 + \varepsilon)^{-1} \delta(p \in r_i^m)}, \tag{6}$$

where ε is a small constant and it is set as 0.1 in our implementation. i is the index of the region, n is the index of the superpixel scale. mc_i^m is the mean color of region r_i^m, $\|I_p - \mathrm{mc}_i^m\|_2$ is the color distance from the pixel p to the color center of r_i^m, and $\delta(\cdot)$ is the indicator function.

Then, integrity of coarse scales and precision of fine scales are involved. Examples for bottom-up cue are shown in the second column of Fig. 2. Due to the application of multiscale segmentation, perceptually homogeneous regions are assigned the same value. We can observe that highly contrast regions are emphasized and highlighted.

2.2 *Top-Down Cue Computation*

In the fourth row of Fig. 2, aircraft's left wing and empennage regions are lost in the bottom-up cue, since these regions are with lower contrast compared with shadows and cluttered background. It is a hint that without prior knowledge, contrast analysis alone can hardly capture the appearance of interested objects. Therefore, we build top-down models by incorporating aircraft shape priors.

As shown in Fig. 3, the proposed top-down cue computation process has training and testing phases. The learning phase involves training a classifier from training images with ground truth. Once the classifier is trained, the top-down cue of a given input image can be estimated by the classifier. It is carried out by extracting features from each image pixel and then feeding the corresponding feature vector into the classifier.

Features used to train the classifier are based on color, location, and contrast. These features provide a top-down modulation of aircraft shape priors and are described in more detail as follows:

Fig. 2 Illustration of main phases of the proposed method. We first compute the bottom-up and top-down cues, respectively. Then both cues are combined to produce an integrated cue to assist in aircraft segmentation. The bottom-up cue prefers contrast, whereas the top-down cue favors shape priors. The combination of both cues provides robust predictions of aircraft regions.

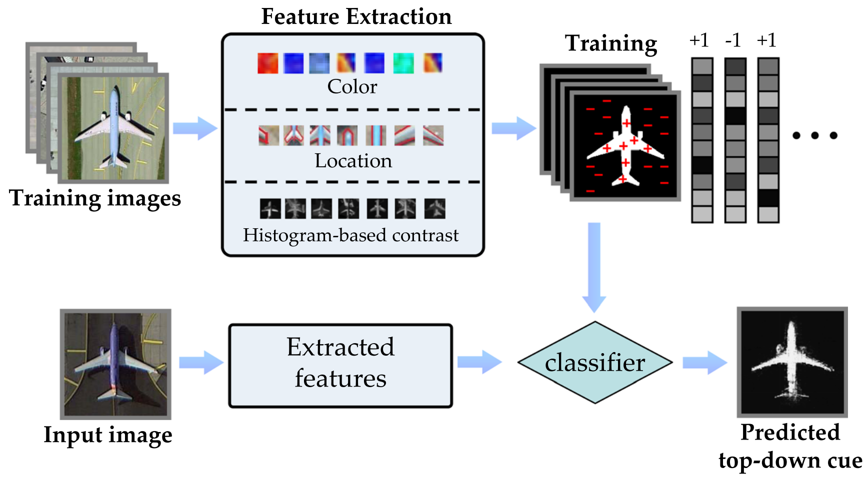

Fig. 3 Illustration of top-down cue computation. A set of visual features are extracted from some training images. Feature vectors extracted from aircraft regions (background regions) are assigned $+1(-1)$ labels. Then a classifier is trained and is used to predict top-down cues of an input image.

Color: The color features extracted from each pixel include RGB values. The wider a color is distributed in the image, the less possible the aircraft contains that color. In addition, aircraft colors have regularities to some extent. Fox example, military aircrafts are painted light gray for camouflage, while civilian aircrafts are painted bright color for conspicuous and noticeable. Hence, color features provide an important clue to predict aircraft regions.

Location: Pixel location is a surprisingly good feature. Indeed, pixels belong to aircraft are usually located at specified locations in the image. Also, pixels near the image boundary are highly likely to belong to the background. Therefore, the normalized x and y coordinates of a pixel are computed as location feature.

Contrast: As mentioned before, contrast is one of the most influential prior of aircraft, and specifically, aircraft pixels present high contrast with background pixels. We use histogram-based method[12] to measure pixel contrast by using color statistics of the input image. Specifically, the contrast of a pixel is defined using its color contrast to all other pixels in the image. Since pixels with the same color have the same contrast, the contrast value of a pixel I in image is defined as

$$\mathrm{hc}(I) = \mathrm{hc}(c_l) = \sum_{j=1}^{n} P(c_j)\mathrm{DC}(c_l, c_j), \tag{7}$$

where c_l is the color value of pixel I, n is the number of distinct pixel colors, $P(c_j)$ is the probability of pixel color c_j, and $\mathrm{DC}(c_l, c_j)$ is the color distance between c_l and c_j.

We extract these visual features from some training images in which aircraft regions are manually labeled. Then, feature vectors extracted from aircraft regions are assigned $+1$ label, while feature vectors extracted from background regions are assigned -1 label. Linear support vector machine is used to train a model. We used the linear model because we found it performed as well as models with multiple kernel learning for the specific task. In addition, linear model is easy to handle and runs with high computational efficiency. Examples for top-down cue are shown in the third column of Fig. 2. It can be seen that the obtained top-down cues are robust to shadows and background clutters.

2.3 *Cue Integration and Refinement*

The bottom-up model and top-down model produce confidence maps C_b and C_t, respectively. Each map is complementary to the other. We normalize both maps to the range [0, 1] and combine two maps as follows to compute an integrated cue:

$$C = C_b \cdot \exp(k \cdot C_t). \tag{8}$$

Through extensive experiments, we observed that the top-down cue C_t is of more significance and discriminative power. Therefore, we use an exponential function to emphasize C_t. In our experiments, we use $k = 3$ as the scaling factor for the exponential.

Figure 2 shows the performance of the bottom-up cue C_b (the second column), top-down cue C_t (the third column), and the integrated cue C (the fourth column). The bottom-up cue prefers contrast, but it is sensitive to shadows and background clutters. The top-down cue favors aircraft shape priors, whereas it always generates some noise pixels. The combination of both cues provides better performance.

We use the integrated cue to assist in aircraft segmentation. Inspired by Cheng's work,[12] iterative GC and adaptive fitting are applied to get final segmentation results. GC is initialized by thresholding the integrated cue. Then, we iteratively run GC to refine the segmentation result (three iterations are employed). Adaptive fitting is used on each new initializations of GC. Specifically, nearby aircraft candidate regions are incorporated, and background regions are excluded according to color dissimilarity.

3 Experiments

3.1 *Data Set Description*

We evaluate the performance of the proposed method on large aerial image datasets collected from Google Earth. For each image, manually segmented mask is generated as the ground truth. The images are 160×160 in size and with the spatial resolution approximately from 0.4 to 0.6 m. The dataset contains both civilian and military aircrafts, such as A320, A340, B737, B747, C5, C17, B1B, KC135, KC130, etc. Therefore, the dataset can be regarded as a representative sample for the validation of the proposed method. The total number of images in the dataset is 550. We randomly chose 150 images as the training set, and the remaining 400 images as the testing set. The dataset and segmentation results are available at Ref. 18.

3.2 *Performance of the Integrated Cue*

The average F-measure score is computed to evaluate the performance of different methods. Indeed, the F-measure is the harmonic mean of precision and recall, i.e.,

$$F = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}}. \tag{9}$$

The F-measure is close to 1 if and only if the segmentation results are close to the ground truth, and the F-measure declines when the segmentation results deviate from the ground truth.

In order to demonstrate the performance of the integrated cue, we use the bottom-up, top-down, and integrated cues for aircraft segmentation, separately. Table 1 shows the segmentation results by using only the bottom-up, top-down, and integrated cues, separately. The F-measure score of segmentation results by using integrated cues outperforms the results by using other cues. It means the integrated cues can take advantage of both cues, and suppress the background clutters.

The visual comparisons of the segmentation results by using different cues are shown in Fig. 4. In the first row, bottom-up cues are distracted by shadow regions, since shadows are high-contrast regions. In the bottom row, top-down cues generate some noisy regions, since

Table 1 *F*-measures of comparing the bottom-up, top-down and integrated cues.

Methods	Bottom-up cue	Top-down cue	Integrated cue
F-measure Score	0.786 ± 0.005	0.876 ± 0.004	0.898 ± 0.002

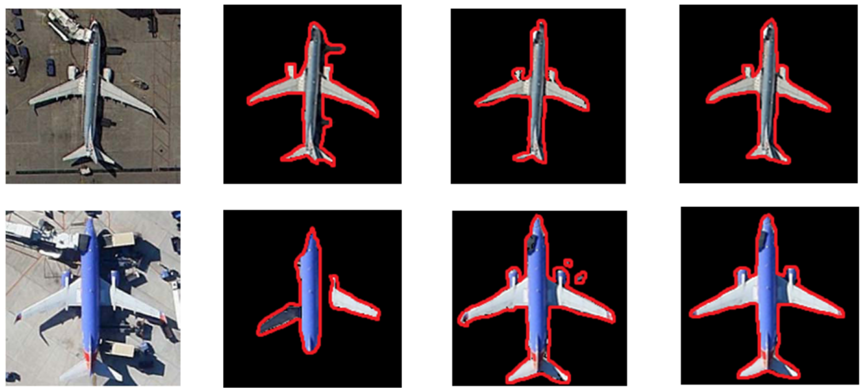

Fig. 4 Visual comparisons of segmentation results using only bottom-up, top-down, and integrated cues. The segmentation results using only bottom-up cue respond to some unrelated visual stimulus. The segmentation results using only top-down cue generate some noise regions. The integrated cue can provide robust segmentation results.

it is based on pixel-level features learning. As mentioned in Sec. 2.3, the bottom-up cue and top-down cue are complementary to the other. The proposed method can take advantage of contrast (bottom-up cue) and shape priors (top-down cue). To be specific, the proposed method can remove noisy regions generated by top-down cues, since these regions are mainly low-contrast regions. In the same time, the proposed method can remove shadow regions generate by bottom-up cues, since these regions are removed by supervised feature learning.

It should be noted that the aircraft in the second row of Fig. 4 is in highly cluttered background. The existence of shadows, boarding bridge, and catering vehicles makes the segmentation of desired aircraft rather challenging. By only using the bottom-up cue, the aircraft's left wing and empennage regions are lost. It is because that these regions are with lower contrast values compared with shadows. The proposed method can suppress the background clutters by supervised feature learning in top-down cue computation. On the other hand, top-down cue computation alone can hardly extract aircraft real boundaries, since some background regions are often comprised of similar visual features. Supervised visual feature learning may respond to some noisy regions, and therefore top-down cues need the assistance of bottom-up cues to get satisfactory segmentation results.

3.3 *Experimental Results and Analysis*

The proposed method is compared with six closely related methods. Results of these methods are obtained by running the authors' publicly available source code. These methods are normalized cut (NC),[9] local binary fitting (LBF),[10] contextual hypergraph modeling (CHM),[11] region-based contrast (RC),[12] level-set (LS),[3] and GC.[15] It should be noted that the CHM method only generate a confidence map of the input image as the final result. Therefore, we applied different thresholds on the confidence map and selected the threshold that gives the best segmentation results. In addition, we use GC[15] implementations from OpenCV.[19] Users are required to give some strokes to mark interested object regions and background regions, respectively.

Table 2 summarized the F-measures on the dataset for all the competing methods. The F-measure score of the proposed method is 0.898 ± 0.002, which significantly outperforms all other methods tested. NC,[9] LBF,[10] CHM,[11] RC,[12] and LS[3] uses bottom-up cues such as textures and colors for segmentation. The proposed method outperforms these methods, demonstrating that incorporating top-down cues is effective for aircraft segmentation. Furthermore, the proposed automatic method compares favorably with GC[15] which is interactive. By taking advantage of both bottom-up and top-down cues, the proposed method can automatically provide better initializations for GC.[15]

Visual comparisons of different methods are provided in Fig. 5. We can see that the proposed method deal well with the challenging cases where the background is highly cluttered. For example, in the top four rows, other methods may be distracted by the shadows while our method almost successfully extracts the whole aircraft. Shadows are high contrast regions in

Table 2 *F*-measures of comparing the proposed method with six closely related methods.

Methods	*F*-measure score	Remarks
NC[9]	0.623 ± 0.002	Automatic
LBF[10]	0.675 ± 0.008	Automatic
CHM[11]	0.795 ± 0.004	Automatic
RC[12]	0.809 ± 0.008	Automatic
LS[3]	0.784 ± 0.004	Automatic
GC[15]	0.839 ± 0.006	Interactive
Proposed method	0.898 ± 0.002	Automatic

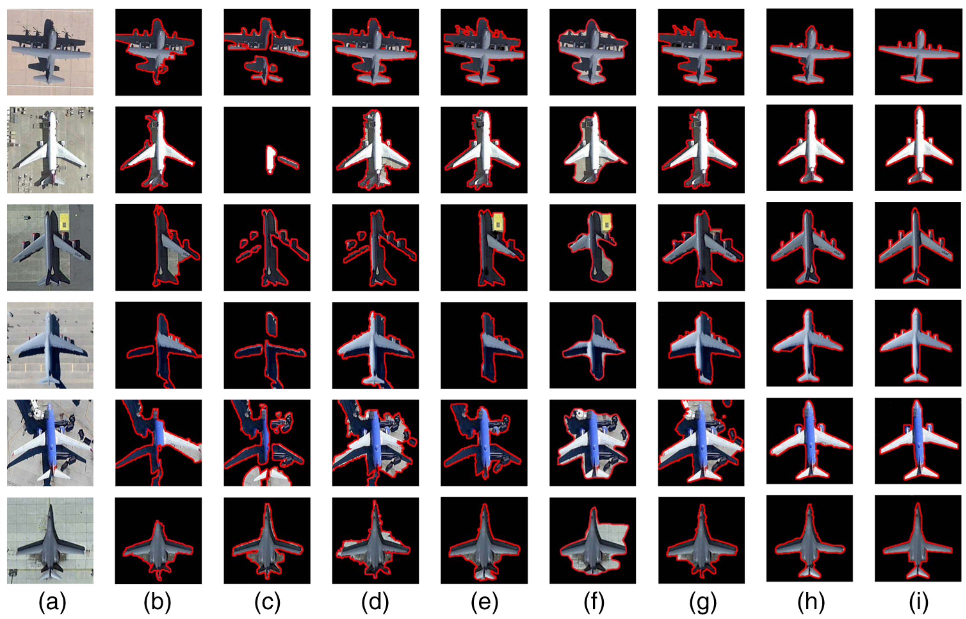

Fig. 5 Visual comparison of the proposed method with several closely related methods: NC, LBF, CHM, RC, LS, and GC. Input image and ground truth (G-Truth) are shown in (a) and (i) for reference. Our results are generally closer to the ground truth compared with the results generated by other methods.

remote sensing images, and they often incur false alarms. The proposed method excludes many shadow pixels by supervised learning of visual features in top-down cue computation, and therefore the segmentation results are robust to shadows to some extent. Moreover, it is also worth pointing out that our method performs well when background is cluttered with oil spots, boarding bridge and catering vehicles, e.g., the fifth and sixth row. In such challenging cases, NC,[9] LBF,[10] CHM,[11] RC,[12] and LS[3] can hardly extract aircraft's actual boundaries, since the bottom-up methods are distracted by unrelated background visual stimulus. GC[15] also fails because it needs more specified aircraft shape information. The proposed method initializes segmentation by combining both bottom-up and top-down cues. Specified aircraft information are provided by contrast (bottom-up cue) and shape priors (top-down cue), respectively. Therefore, by integrating both cues, background clutters are suppressed and better segmentation results are obtained.

4 Discussion

We presented an aircraft segmentation method that combines bottom-up and top-down cues. The proposed method was evaluated on VHR images, and the experimental results show that our method consistently outperforms other methods in terms of F-measure score. We believe that the bottom-up and top-down cue integration helps to achieve this performance. Taking RC[12] method, e.g., it is based on regional contrast. The results of RC[12] in the sixth row of Fig. 5 shows that oil spots are easily annotated as aircraft regions when solely considering the bottom-up cue. The LBF[10] method follows the hypothesis of the boundary prior, and the result of LBF[10] in the second row of Fig. 5 shows that the aircraft fuselage is lost. It is due to the fact that color features and aircraft shape priors are not taken into account. The GC[15] method incorporates shape priors by manual interventions. However, some strokes which mark desired object regions may hardly provide enough shape priors. The result of GC[15] method in the first row of Fig. 5 shows that shadow regions are incorrectly annotated as aircraft regions. That is because these shadow regions are adhering to aircraft fuselage, more detailed shape prior information needs to be added.

The described failure cases show that, the RC[12] and LBF[10] methods using only bottom-up cues may produce unsatisfactory results in some cases. On the other hand, GC[15] method needs to incorporate more detailed shape prior information to produce satisfactory results. However, integration of the bottom-up and top-down cues enable the proposed method to obtain more precise segmentation results than the results of methods that are exemplarily analyzed earlier.

We particularly consider applying the proposed method in an airport dynamic surveillance environment. Foreground windows containing aircrafts need to be provided by an aircraft detector. In these windows, the pose of each aircraft needs to be estimated. Then we can automatically perform aircraft segmentation. To be specific, before aircraft segmentation, the input image should be well cropped, and aircraft should be center aligned. If the images are note well cropped or rotated, the proposed method may not always be effective. Currently, we only test the proposed method on well cropped images, leaving the largely unexplored area as future research. We believe that integrating more sophisticated methods such as aircraft pose estimation would be beneficial.

Although the dataset built in our research only contains well cropped images, we argue that the bottom-up and top-down cue integration for object segmentation with regular shapes is very important for remote sensing image application, especially for automatically processing large quantities of images containing desired objects.

5 Conclusion

This paper has presented an aircraft segmentation method that combines bottom-up and top-down cues. Existing segmentation methods usually require manual intervention to optimally segment objects from cluttered background, so that they can hardly work well in automated surveillance systems. The proposed method integrates both bottom-up and top-down cues for GC initialization, and therefore can provide robust segmentation results. We build a manually labeled aircraft image dataset which have been made available for other researchers. Experimental results on the dataset show that the proposed method consistently outperforms other closely related methods.

We recognize that there is still potential for accuracy improvement. From segmentation results in Fig. 5, we can observe that shadows influence the segmentation accuracy, removing shadows before bottom-up and top-down cue computation may contribute to performance improvement. In addition, it may be beneficial to take more high-level factors (such as aircraft knowledge and imaging time) into consideration. Hence, our future research will focus on these factors to enhance the segmentation accuracy.

Acknowledgments

The authors would like to thank Google Inc. for providing remote sensing data. The authors would also like to thank all the anonymous reviewers for their very competent comments and helpful suggestions. This work was partially supported by the National Science Fund for distinguished young scholars (No. 61125206) and the NSFC key project (No. 61331017).

References

1. M. N. Klaric, G. J. Scott, and C. R. Shyu, "Multi-index multi-object content-based retrieval," *IEEE Trans. Geosci. Remote Sens.* **50**(10), 4036–4049 (2012).
2. T. Moranduzzo and F. Melgani, "Automatic car counting method for unmanned aerial vehicle images," *IEEE Trans. Geosci. Remote Sens.* **52**(3), 1635–1647 (2014).
3. Q. Xu et al., "Multiscale contour extraction using a level set method in optical satellite images," *IEEE Geosci. Remote Sens. Lett.* **8**(5), 854–858 (2011).
4. L. Zhang and K. Yang, "Region-of-interest extraction based on frequency domain analysis and salient region detection for remote sensing image," *IEEE Geosci. Remote Sens. Lett.* **11**(5), 916–920 (2014).
5. Q. Wu and J. An, "An active contour model based on texture distribution for extracting inhomogeneous insulators from aerial images," *IEEE Trans. Geosci. Remote Sens.* **52**(6), 3613–3626 (2014).
6. G. Liu et al., "Interactive geospatial object extraction in high resolution remote sensing images using shape-based global minimization active contour model," *Pattern Recogni. Lett.* **34**(10), 1186–1195 (2013).
7. Z. Zhao et al., "Velocity estimation of an airplane through a single satellite image," *Chin. Opt. Lett.* **10**(3), 031001 (2012).
8. A. Borji, "Boosting bottom-up and top-down visual features for saliency estimation," in *Proc. IEEE Conf. CVPR*, pp. 438–445 (2012).
9. J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(8), 888–905 (2000).
10. C. Li et al., "Implicit active contours driven by local binary fitting energy," in *Proc. IEEE Conf. CVPR*, pp. 1–7 (2007).
11. X. Li et al., "Contextual hypergraph modeling for salient object detection," in *Proc. IEEE Conf. ICCV*, pp. 3328–3335 (2013).
12. M. M. Cheng et al., "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(3), 569–582 (2014).
13. N. Tong et al., "Saliency detection with multi-scale superpixels," *IEEE Signal Process. Lett.* **21**(9), 1035–1039 (2014).
14. J. Xu et al., "Automatic detection of inshore ships in high-resolution remote sensing images using robust invariant generalized Hough transform," *IEEE Geosci. Remote Sens. Lett.* **11**(12), 2070–2074 (2014).
15. C. Rother, V. Kolmogorov, and A. Blake, "Grabcut–interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.* **23**(3), 309–314 (2004).
16. P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.* **59**(2), 167–181 (2004).
17. H. Jiang et al., "Automatic salient object segmentation based on context and shape prior," in *Proc. BMVC*, pp. 1–12 (2011).
18. Feng Gao's Homepage, http://fenggao.sinaapp.com
19. OpenCV, "Open source computer vision library," http://www.opencv.org/

Feng Gao is currently a lecturer at the College of Information Science and Engineering, Ocean University of China. He acquired his BS degree in software engineering from Chongqing University, China, in 2008 and received his PhD in computer science from Beihang University, China, in 2015. His research interests include information fusion, computer vision, and image understanding.

Qizhi Xu received his BS degree in computer science from Jiangxi Normal University, Jiangxi, China, in 2005 and his PhD in computer science from Beihang University, Beijing, China, in 2012. He is a postdoctoral fellow with the Canada Research Chair Laboratory in Advanced Geomatics Image Processing, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, New Brunswick, Canada. His research interests include image registration, information fusion, computer vision, and image understanding.

Bo Li received his BS degree from Chongqing University, China, in 1986, his MS degree from Xi'an Jiaotong University, China, in 1989, and his PhD from Beihang University, China, in 1993, all in computer science. Currently, he is a professor of computer science and engineering and the director of the Beijing Key Laboratory of Digital Media in Beihang University.