Research Journal of Applied Sciences, Engineering and Technology 5(23): 5469-5473, 2013
DOI:10.19026/rjaset.5.4221
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: November 29, 2012 Accepted: January 17, 2013 Published: May 28, 2013

Research Article

High-Speed Recognition Algorithm Based on BRISK and Saliency Detection for Aerial Images

Teng-Jiao Xiao, Dan-Pei Zhao, Jun Shi and Ming Lu
Image Processing Center, School of Astronautics, Beihang University, Beijing, China

Abstract: A fast ground object recognition method for aerial images, taking airports, oil depots, harbors, etc., as research objects, is proposed in this study based on BRISK and the visual saliency detection. According to the characteristics of aerial images, such as high resolution and complex background interference, saliency detection is applied to select the candidate object region where the target may exist. Therefore, it can reduce the searching range effectively. And then, BRISK matching method is used to recognize the object efficiently. A variety of experiments under different interference factors are carried out based on the typical object database of aerial images in this study. Experimental results show that the proposed algorithm can not only maintain the validity of BRISK features under the conditions of rotation, scale, illumination and viewpoint changes, but also shorten the matching time, satisfying real-time demand.

Keywords: Aerial images, BRISK matching, object recognition, saliency detection

INTRODUCTION

Image matching technology is one of the important factors restricting the development of aerial reconnaissance and precision-guided system, which plays a very important role in modern warfare. The aerial images obtained real-time were matched to the pre-stored target image to achieve the target accurate recognition and precise position. Due to the high resolution and large size of the aerial images, the timeliness and accuracy of the matching algorithm become very important performance indicators. At the same time, not only the great variation of illumination, scale, rotation, viewpoint between the target image and the real-time image, but also many interference such as cloud shelter, noise, low contrast, image blur may affect the matching result. Therefore, looking for an accurate and fast matching algorithm becomes very important in order to meet the needs for practical application.

Over the past decade, for the good robustness of matching algorithm based on the interest point, this kind of matching methods becomes the research focus in matching field. One of the most classical matching algorithms is SIFT (Scale Invariant Feature Transform) (Lowe, 1999), which is based on the Gaussian scale space image pyramid and can be invariant to illumination changes and affine or 3D projection. In 2004, Lowe improved the algorithm (Lowe, 2004), making it further robust to affine distortion, 3D viewpoint change, noise and illumination changes. But the large calculation and slow running speed restrict its engineering practical application. Herbert Bay etc., proposed SURF (Speeded Up Robust Features) (Bay *et al*., 2008) using the integral image to accelerate the matching speed. But compared to SIFT, there exists a big gap of the matching performance. From 2010, a series of matching algorithm based on the FAST corner detection such as BRIEF (Calonder *et al*., 2010), BRISK (Leutenegger *et al*., 2011) and the ORB (Rublee *et al*., 2011) appear. These algorithms use binary strings to describe the features and Hamming distance to measure the feature distance, enhancing the speed of the matching largely. But searching in the whole real-time image still consumes much time and waste a lot of redundant computation. So far, this kind of matching methods has already peaked for the aspect of speed. To outperform this kind of matching methods in terms of speed is extremely difficult. Recently, saliency detection has attracted a lot of attention in many fields. It imitates human visual system to seek interesting regions in images to reduce the search effort in tasks such as object detection and recognition. Inspired by the above discussion, a high-speed recognition algorithm is proposed in this study, which combines BRISK and saliency detection. The algorithm further accelerates the speed of BRISK and achieves accuracy and real-time requirements.

SALIENCY DETECTION AND BRISK

Saliency detection: Visual attention analysis which imitates the human visual system can detect the salient area in an image automatically. In object detection or recognition task, just searching the interested target in

Corresponding Author: Teng-Jiao Xiao, Image Processing Center, School of Astronautics, Beihang University, Beijing, China

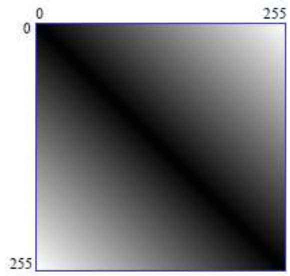
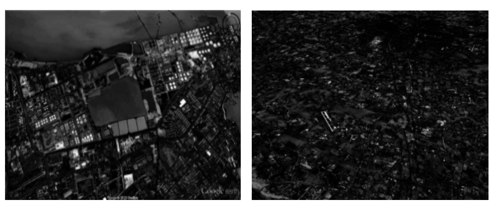


Fig. 1: Distance map

(a) Original images



(b) Saliency map

Fig. 2: Saliency detection result

the salient areas, can greatly reduce the amount of computation. Saliency evaluation can be classified into three folds: biologically inspired methods, computationally oriented models and methods combined both. Most stimuli-driven attention analyses are based on the biological model utilizing contrast as the influential factor. While the contrast are based on various types of image features, such as color, edges, gradients, spatial frequencies, histogram, or combinations thereof. Through the observation and analysis of real-time aerial images, the color contrast of the airport and oil depots are salient. Motivated by the spatiotemporal cues saliency detection (Zhai and Shah, 2006), an attention model based on the color contrast was proposed in this study. The saliency value of a pixel Ik of an image I is defined as:

$$SalS(I_k) = \sum_{\forall I_n \in I} \|I_k - I_n\| \tag{1}$$

where, In is in the range of [0, 255] and Eq. (1) can be expand into the following form:

$$SalS(I_k) = \|I_k - I_1\| + \|I_k - I_2\| + \cdots + \|I_k - I_N\| \tag{2}$$

where, N is the pixel number of an image. Given an image, the value of each pixel is known. Equation (1) can be further structured in the way that pixels with the same value are arranged to be together:

$$SalS(I_k) = \sum_{n=0}^{255} f_n \|I_k - I_n\| \tag{3}$$

where, f_n is the frequency of the same value I_n appears in the image. The frequency can be expressed in the form of histogram. For $I_n \in [0, 255]$, the color distance $\|I_k\text{-}I_n\|$ is also bounded in [0, 255]. Since it is a fixed range, a distance map can be constructed before the saliency map computation. In the distance map, element $D(x, y) = \|I_x\text{-}I_y\|$ is the color difference between the pixel value I_x and I_y, as shown in Fig. 1.

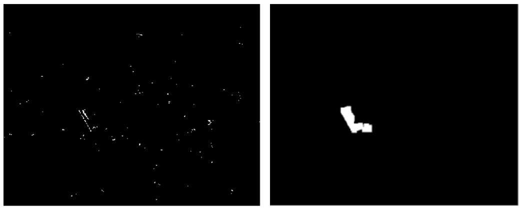
Given a frequency histogram and a color distance map, the saliency value of pixel Ik is computed as:

$$SalS(I_k) = \sum_{n=0}^{255} f_n D(k, n) \tag{4}$$

Thereby, there is no need to use Eq. (1) to calculate the value of all the pixels, only need to calculate the value of the colors $\{i_n, n = 0... 255\}$ to generate the final saliency map. The saliency detection result of the method is shown in Fig. 2.

BRISK: BRISK feature detection is based on the AGAST (Mairm *et al*., 2010), which is essentially an extension for accelerated performance of FAST. For invariance to scale is an important indicator for high-quality key points, BRISK algorithm extends FAST algorithm to image plane as well as scale-space and gives the optimal solution of the continuous scale space. The scale-space pyramid layers consist of n octaves c_i and n intra-octaves d_i, for {i = 0, 1… n-1} and n = 4 typically. Then the FAST 9-16 detector is applied on each octave and intra-octave with the same threshold T to detect the potential interested regions. The points belonging to the regions should be non-maxima suppressed in scale-space to get the key points. For the entire detected maximum, a sub-pixel and continuous scale refinement are performed. Through the 2D quadratic function fitting and 1D parabola fitting along the scale axis, the final optimal estimate can be determined.

The description of feature plays an important role in the matching process, which affects the matching efficiency greatly. The sampling pattern of BRISK is based on the neighborhood of the key point, which defines N locations equally spaced on circles concentric with the key point. BRISK descriptor is composed as binary string by catenating the results of simple brightness comparison tests between the key point and the sampling points. A characteristic direction of each key point is identified to allow for orientation-normalization, so thus to achieve rotation invariance which is a key to general robustness.



(a) Binary image (b) Morphological processing

(c) Salient patch

Fig. 3: The result of extracting salient patch

HIGH-SPEED RECOGNITION ALGORITHM

In this study, a novel method combining saliency detection and BRISK matching for target recognition is proposed. Firstly, the real-time image is performed by saliency detection processing, thus the whole image is cut into several salient patches. Then, for each patch, BRISK is used to recognize the target.

Extracting salient patches: After saliency detection, we obtain a saliency map as shown in Fig. 2b. In the saliency map, if the area is brighter, it will be more salient. We use simple threshold segmentation to detect objects in a saliency map, with the threshold 2/3. Due to the complex background, the resulting binary image always contains many small bright spots. Besides, due to the texture of the target, the region of the target is not entirely bright, as shown in Fig. 3a. For this reason, we apply a morphological processing. Firstly, expansion processing, through this we can combine the bright spot together to obtain connected regions. For airport or oil depot usually occupy a large area of the aerial image, they should locate in the largest connected region. Based on this assumption, in the binary image, we reduce the non-object patch by removing smaller connected area, thus to further reduce the number of salient patch significantly. The results of processing are illustrated in Fig. 3b. In each connected region, the x and y directions of the maximum and minimum coordinate are recorded to extract the candidate object patches in the original real-time image, namely getting the candidate object patches, as presented in Fig. 3c.

Matching and recognition: For each salient patch, we perform BRISK matching to recognize the object. Matching two BRISK descriptors is a simple computation done by their Hamming distance. The number of bits different in the two descriptors is the measure of their feature distance. The operations can be performed in a bitwise XOR followed by a bit count, computing quite efficiently.

(a) Beijing capital international airport

(b) Pudong airport

(c) Huangdao oil ports

Fig. 4: Examples for real-time images of each object

EXPERIMENTAL RESULTS

In order to verify the validity of the proposed algorithm, we use BCIA (Beijing Capital International Airport), PA (Pudong Airport), HOD (Huangdao Oil Depots) as objects to validate the high-speed target recognition experiments. We use the screenshot from Google Earth as real time image, with 10 images for each object. The image size is all about 1000×600 pixels. The target image is captured from Google Earth stochastically. Examples for real-time images of each object are show in Fig. 4.

The number of salient patches extracted from each real-time image is as show in Table 1.

For each salient patch, we perform BRISK to recognize the interested target. All the process of the proposed algorithm includes saliency detection, extracting salient patches and BRISK matching. We add the time of saliency detection and extracting salient patches together, called detection and extraction. We compare the running time of the proposed method to the original BRISK. It is worthwhile to note that BRISK makes use of some SSE2 and SSSE3 commands, which leading the high-speed of BRISK. In

Table 1: The number of salient patches

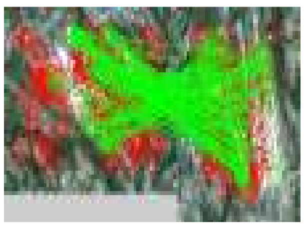
	1	2	3	4	5	6	7	8	9	10
BCIA	1	1	2	2	1	2	2	1	1	2
PA	1	1	1	1	1	1	1	1	1	1
HOD	4	6	3	4	4	2	3	2	4	5

Table 2: Time of comparison with accelerate commands

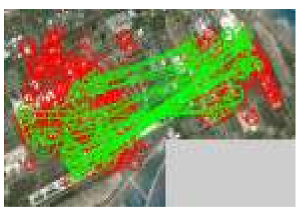
	Detection and extraction	BRISK matching	Total time	Original BRISK
BCIA	132.90 ms	349.350 ms	482.25 ms	557.36 ms
PA	152.01 ms	332.850 ms	484.87 ms	552.58 ms
HOD	151.00 ms	334.346 ms	485.45 ms	550.82 ms

Table 3: Time of comparison without accelerate commands

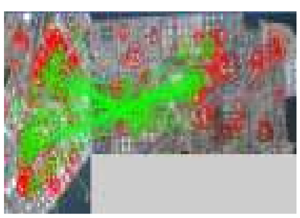
	Detection and extraction	BRISK matching	Total time	Original BRISK
BCIA	0.318s	9.06s	9.378s	13.270s
PA	0.322s	8.93s	9.252s	10.419s
HOD	0.324s	8.97s	9.294s	10.535s



(a) Recognition result of BCIA



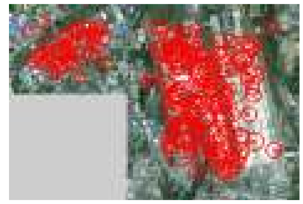
(b) Recognition result of PA



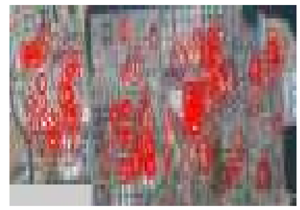
(c) Recognition result of HOD

Fig. 5: Examples of the recognition results

this study, we perform two groups of experiments, BRISK with accelerate commands and BRISK without accelerate commands, to see the function of saliency detection framework. We perform the first group experiment on the computer with 3.40 GHz CPU, 4.00 GB RAM and 64-bit operating system, while the second group experiment is on the computer with 1.60 GHz CPU, 2.50 GB RAM and 32-bit operating system. We arrange the experiments in this way because the SSSE3 command appears in computer from the year



(a) Result of BCIA



(b) Result of HOD

Fig. 6: Recognition results of non-object salient patches

2007. Computer produced before 2007 are in such configuration. The time of the comparison with accelerate commands is show in Table 2 and the one without accelerate commands is show in Table 3.

From the two tables above, we can see the function of the proposed framework, especially cooperating with the relative slow matching method, the accelerating effect is distinct. Examples of the recognition results are shown in Fig. 5. The images on the left are the salient patches extracted from the real-time images, while the images on the right are target images.

About the non-object salient patches, no matching result can be obtained, as shown in Fig. 6.

CONCLUSION

This study studied on the ground object recognition issue of aerial images, taking airports and oil ports as interested targets. The experimental results show that the proposed method not only maintains the accuracy of BRISK matching algorithm, but also further accelerates the speed of matching. Especially cooperating with some relative slow matching methods, the advantage of the proposed framework is obvious.

ACKNOWLEDGMENT

This study is supported by a grant from the Natural Science Foundation of China (No. 61071137) and National Basic Research Program of China (No. 2010CB327900).

REFERENCES

Bay, H., A. Ess, T. Tuytelaars and L.V. Gool, 2008. SURF: Speeded up robust features. Comput. Vision Image Understanding, 110(3): 346-359.

Calonder, M., V. Lepetit, C. Strecha and P. Fua, 2010. Brief: Binaru robust independent elementary features. Proceeding of European Conference on Computer Vision (ECCV). Crete, Greece, 6314: 778-792.

Leutenegger, S., M. Chli and R. Siegwart, 2011. BRISK: Binary robust invariant scalable key points. Proceeding of the International Conference on Computer Vision (ICCV).

Lowe, D.G., 1999. Object recognition from local scale-invariant features. Proceeding of International Conference on Computer Vision (ICCV). Corfu, Greece, pp: 1150-1157.

Lowe, D.G., 2004. Distinctive image features from scale-invariant key points. Int. J. Comput. Vision, 60(2): 91-110.

Mairm, E., G.D. Hager, D. Burschka, M. Suppa and G. Hirzinger, 2010. Adaptive and generic corner detection based on the accelerated segment test. Proceedings of the European Conference on Computer Vision (ECCV), pp: 2-5.

Rublee, E., C. Rabaud, K. Konolige and G. Bradski, 2011. ORB: An efficient alternative to SIFT or SURF. Proceeding of International Conference on Computer Vision (ICCV).

Zhai, Y. and M. Shah, 2006. Visual attention detection in video sequences using spatiotemporal cues. Proceeding of the 14th ACM Multimedia. Santa Barbara, CA, USA, pp: 815-824.