Fast SIFT scene matching algorithm based on saliency detection and frequency segmentation for downward-viewing images

ZHAO Dan-pei
Image processing center
School of Astronautics
Beihang University
Beijing, China
zhaodanpei@buaa.edu.cn

WANG Jia-jia
Image processing center
School of Astronautics
Beihang University
Beijing, China
gaga214@126.com

WAN Jie-yuan
Image processing center
School of Astronautics
Beihang University
Beijing, China

XIAO Teng-jiao
Image processing center
School of Astronautics
Beihang University
Beijing, China
jiaojiao_8877@163.com

***Abstract*—A fast downward-viewing scene matching method, taking airports, oil depots, harbors and so on as research objects, is proposed in this article which is based on the visual saliency detection and the segmentation of frequency domain. According to the characteristics of downward-viewing images, such as high resolution and complex background texture, saliency detection is used to determine the candidate region where the target may exist to reduce the searching range effectively. And then, the segmentation of frequency domain is used to eliminate the frequency component except the frequency of the target to reduce the redundant information, thereby saving the computation of SIFT feature extraction and matching. A variety of experiments under different interference factors are carried out base on the typical object database of downward-viewing images in this paper. Experimental results show that the fast matching algorithm proposed in this paper can not only maintain the validity of SIFT features under the condition of rotation, scale, illumination and viewpoint changes, but also shorten the matching time largely and improve the matching efficiency, laying the foundation for further practical application.**

Keywords-Spectral residual; Saliency detection; SIFT matching; Segmentation of frequency domain

I. INTRODUCTION

The technology of downward-viewing scene matching, which is one of the most important factors that restrain the development of aerial reconnaissance and precision guidance system, plays a vital role in the modern war. Downward-viewing scene matching is performed by matching the real-time images with the pre-stored target images to accurately detect and locate the targets of interest. According to the characteristics of downward-viewing images, such as high resolution and complex background, the speed and accuracy of matching algorithms are becoming extremely important. Meanwhile, due to the special practical environment, images are not only influenced by the variance of illumination, rotation, viewpoint and scale, but also interfered by the similar objects, cloud and fog occlusion, noise, low contrast and fuzzy caused by flight vibration. Therefore, in order to satisfy the practical application demands, creating a matching algorithm which is accurate and fast has become a very urgent application bottleneck.

In the last decade, matching algorithms based on the feature points become a hotspot due to its good robust performance. The most classic algorithm, Scale Invariant Feature Transform (SIFT) [1], which is based on Gaussian scale space, was proposed by Professor Lowe in 1999. It has been proved to be accurate and robust. After that Herbert Bay put forward Speeded Up Robust Features (SURF) [3] by using integral image concept, which had greater acceleration, but a large gap in terms of matching accuracy compared with SIFT. In the latest two years, there occurs a series of matching algorithms based on FAST corner detection, such as BRIEF [4], BRISK [5] and ORB [6]. But they are still difficult to guarantee the reliability of the matching results due to the complexity of downward-viewing scene. So far, SIFT is generally recognized as the best matching algorithm, but its large computation, especially for the images with high resolution, restricts its applications in the engineering practice.

Recently, saliency detection gradually becomes an important instrument in many image processing and computer vision fields. Saliency evaluation can be classified into three folds: biologically inspired methods, computationally oriented models and methods combined both. Among them, the famous approach is proposed by Itti [7] whose method is based on reasonable biologically inspired framework. Algorithm SR [8] and algorithm PQ [8] take advantage of computation of the amplitude and phase of Fourier spectrum respectively to get the saliency region. Goferman put forward the CA [9] algorithm, Therefore, according to the characteristics of the down-viewing scene images, which is typical aerial reconnaissance and attack targets, such as airports, oil depots and harbors, we propose a fast SIFT scene matching algorithm based on visual saliency detection and frequency segmentation. This method is fast and robust towards the interference like rotation, scale, illumination and viewpoint changes. The main idea, by using the saliency detection, aims at eliminating the non-target area. The next step is to further remove redundant information and reduce the time of features extraction by using frequency segmentation, so the matching speed can be improved. Our experimental results show the efficiency of our algorithm that the speed of SIFT features matching algorithm is obviously to be meliorated, as well as its matching accuracy.

Published by Atlantis Press, Paris, France.

II. VISUAL SALIENCY DETECTION

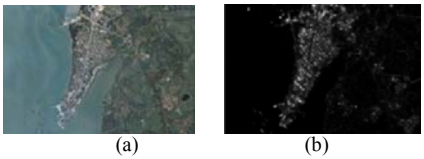
Visual attention mechanism imitates human visual system to detect the saliency region automatically by generating the saliency map. According to the characteristics of down-viewing terrain background, we try to extract feature points of interest in the visual saliency probable area, so the computation of redundant information and matching are likely to be reduced, and the matching efficiency can be improved.

A. Spectral residual model

SR (spectral residual) saliency detection [8] is a relatively simple visual saliency computational model, which is based on Fourier transform. It divides image information into two parts: redundant part and novelty part. The novelty part of the image is corresponding to the target region. After Fourier transform, the phase spectrum and the amplitude spectrum of a image can be got, denoting as $P(f)$ and $A(f)$ respectively. The amplitude spectrum $A(f)$ is used to get the novelty part.

Scale invariance, known as $1/f$ law, which means that the distribution of amplitude of the averaged Fourier spectrum is stated as $E\{A(f)\}\propto 1/f$. As the study found that the averaged log spectrum of the ensemble of natural images presents the partial linear characteristic, the log scale is adopted in our paper. We define $L(f)$ as the amplitude spectrum in log scale, it can be obtained by the transformation $L(f)=\log(A(f))$. The averaged amplitude spectrum $L(f)*h_n(f)$ can be got through mean filtering to $L(f)$, where $h_n(f)$ is a mean filterer with an n*n convolution nucleus(n=3). The spectral residual $R(f)$, which is the novelty part of log spectrum, can be got by subtracting the averaged amplitude spectrum $L(f)*h_n(f)$ from the log amplitude spectrum $L(f)$.

Finally, we do inverse Fourier transform to the phase spectrum $P(f)$ and the obtained spectral residual $R(f)$. In order to get better visual effect, Gaussian filtering $g(x)$ with $\sigma=8$ is used to process the image we got. In the end, the SR Saliency map is obtained, in which the magnitude of the brightness represents the degree of saliency, as shown in Fig.1.



(a) (b)

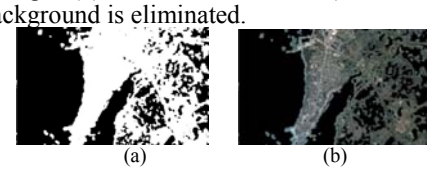
Figure 1. The result of SR saliency detection for a downward-viewing image:(a) Original image.(b) SR saliency map.

B. Saliency region extraction

Visual saliency region needs to be extracted from the original images after obtaining the SR saliency map. Assume that, SR saliency map is gained as $S(x)$ by saliency detection, so the binary mask $O(x)$ can be obtained:

$$O(x)=\begin{cases}1 & if\ S(x)>threshold \\ 0 & otherwise\end{cases}. \tag{1}$$

Where threshold=$E(S(x))\times 3$, and $E(S(x))$ denotes the averaged intensity value of the saliency map. Binary mask can help to extract the interest region from the corresponding position of the input images. It can be seen from Fig.2 (b) that the smooth area (ocean in this picture) of background is eliminated.



(a) (b)

Figure 2. The result of extracting salient regions from a downward-viewing image: (a) Binary mask. (b) Visual saliency regions.

III. IMPROVED FAST MATCHING ALGORITHM OF DOWN-VIEWING IMAGE

A novel matching algorithm is proposed in this paper based on SIFT combined with saliency detection. It can improve matching efficiency, as well as its accuracy. At the same time, in order to reduce matching time, we suggest to do information mapping and segmentation in frequency domain between the real-time images and the reference image, and then mapping the prior frequency information of the reference image into the frequency domain of the real-time images.

A. Frequency segmentation

Different regions of the image correspond to different parts of the frequency domain. Specifically, the low frequency part is corresponding to the smooth area of the image, while the high frequency part is corresponding to the dramatic change areas. We normalize the real part R(x, y) and the imaginary part I(x, y) of Fourier spectrum to 0~255 for histogram statistics. We remain the intervals $[\alpha_R\ \beta_R]$ and $[\alpha_I\ \beta_I]$ of statistics which is less than 2. Those intervals contain the main information of the real part $R(x, y)$ and the imaginary part $I(x, y)$ of the reference image. Then the intervals were de-normalized to get the actual intervals $[R_\alpha\ R_\beta]$ and $[I_\alpha\ I_\beta]$ corresponding to the real range of the Fourier transform. The denormalization equation is as follows:

$$actual\ interval = Maximum\times\frac{original\ \ interval}{255}. \tag{2}$$

After Fourier transform for real-time image, we can get the maximum values of the real part $Max[R'(x,y)]$ and imaginary part $Max[I'(x,y)]$. Then, just as stated in (2), the two maximums and the intervals $[\alpha_R\ \beta_R]$ and $[\alpha_I\ \beta_I]$ are use to get the actual intervals $[R_\alpha'\ \ R_\beta']$ and $[I_\alpha'\ \ I_\beta']$ respectively corresponding to the real part $R'(x,y)$ and imaginary part $I'(x,y)$of the real-time image. Comparing the interval parameter $[R_\alpha'\ \ R_\beta']$ and $[I_\alpha'\ \ I_\beta']$ of the real-time image respectively with interval $[R_\alpha\ R_\beta]$ and $[I_\alpha\ I_\beta]$ of the reference image, the minimum parameters are chosen. Therefore, we can obtain practical segmentation threshold $[R_\alpha''\ \ R_\beta'']$ and $[I_\alpha''\ \ I_\beta'']$. The main frequency information in the practical intervals $[R_\alpha''\ \ R_\beta'']$ and $[I_\alpha''$

$I_{\beta}^{''}$] is used to do inverse Fourier transform. Finally we get the segmented image which is ready to be matched.

Through frequency segmentation, it can remain the information of target region of an image, in the meantime, the details of the other region will be reduced, and a large number of redundant information can be discarded.

B. SIFT feature matching

SIFT feature matching is completed by comparing the reference images to the processed real-time images. The feature extraction and generation based on SIFT local invariant include four steps:

1) The establishment of Gaussian pyramid based on the extreme value detection of scale space.

2) Locating the position and scale of the key points accurately by fitting three dimensional quadratic functions.

3) The distribution of the direction for histogram peak is the main direction of the feature points.

4) The generation of the descriptor by rotating the coordinate axis according to the direction of the key points to keep its rotation invariance.

C. Implement of fast matching algorithm

After Fourier transform of the real-time images, the computation should be done in two aspects: one is to get the binary mask of the saliency region by SR saliency detection towards the result of the Fourier transform; the other is to do frequency segmentation to the real part and imaginary part of the Fourier spectrum which aims at keeping the target frequency information. Then using the binary mask of saliency map to extract the corresponding saliency region from the segmented images, we can get the images that prepared to be matched. The result of each step is shown as follows:

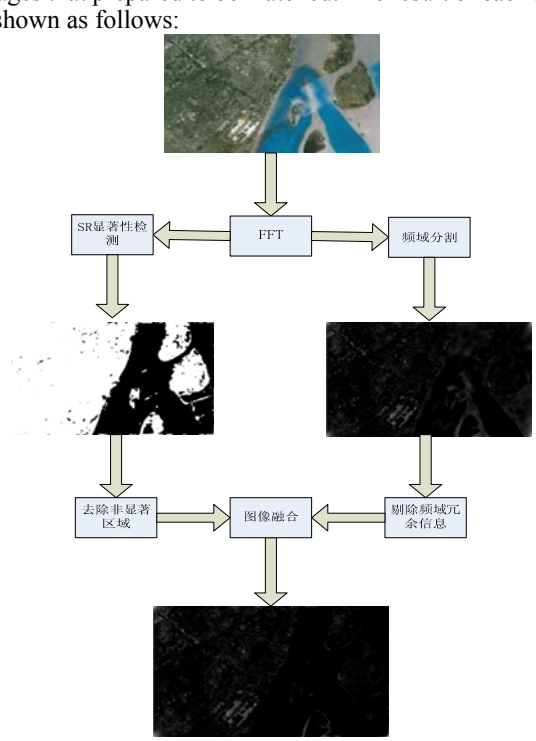


Figure 3.The processing result combined saliency detection and frequency segmentation

After the process above, the information details of target region can be well kept, meanwhile, the details of the non-salient areas and non-target areas are likely to be discarded. On this basis, after SIFT matching, we adopt the KD-tree search algorithm and the RANSAC to eliminate false match points. With this method, it can greatly improve the matching efficiency.

IV. EXPERIMENTAL RESULT

In this paper, we capture remote sensing images from Google Earth by random. Our algorithm can accurately detect and recognize the targets on the condition of the complex environment, such as the interference of the illumination, viewpoint, rotation and scale changes. In order to validate the efficiency of our algorithm, we take the typical ground objects as examples, such as airports, oil depots and harbors, to design two groups of experiments. The first group matches the target image with 50 real-time images that are in different size, whose length and width are approximately 450×700 pixels under the condition of different interference. Comparing our algorithm with the former algorithms, such as SIFT down-viewing scene matching algorithm and SIFT matching algorithm. combined with saliency detection， we can prove the robustness of our algorithm. The second group is to testify the effect of the image size on the algorithm performance. In order to be able to evaluate the time that matching algorithm performance takes, here, we define the improved time parameter ω according to algorithm A and algorithm B, just as following:

$$\omega=\frac{T(B)-T(A)}{T(B)}\times 100\%. \tag{3}$$

Where $T(B)$ and $T(A)$ respectively denotes the averaged time of algorithm B and A.

All of our experiments are performed under the software of MATLAB, which operates in a computer with windows XP operating system and 1.86 GHz processor.

A. Experiment 1: The performance of matching algorithms under illumination, viewpoint, rotation and scale variation conditions

The experiment of this group attempts to compare three matching algorithms including our method under illumination, viewpoint, rotation and scale variation conditions. Through the experimental data, we can get the results shown as table 1 and table 2 separately.

Comparing with the other two algorithms, we can draw the conclusion from the experimental results that our algorithm has greatly improved matching speed, at the same time, it can keep its accuracy. During 50 experiments, when illumination, rotation and scale changes, matching speed have great improvement while match ratio descends only 2% in table 1 and 4% in table 2. It is because that our algorithm would lose some details of the targets. As for matching speed, our algorithm performance time

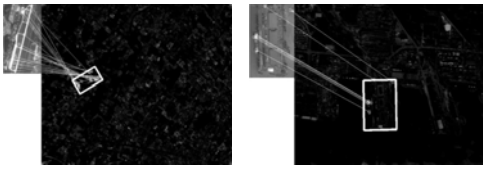
increases ω=24.07% in table 1 and ω=24.50% in table 2 compared with SIFT matching algorithm combined with saliency detection, and ω=27.76% in table 1 and ω=27.94% in table 2 compared with SIFT matching algorithm. Thus, it can be seen, our algorithm has a great superiority in speed performance. The experimental results are shown in Fig. 4.

TABLE I. THE PERFORMANCE COMPARISON OF THREE MATCHING ALGORITHMS UNDER ILLUMINATION, ROTATION AND SCALE VARIATION CONDITIONS

Algorithm	Experiment times	Mean matching time	Correct matching times	Match ratio
Our algorithm	50	7.13s	47	94%
SIFT algorithm combined with saliency	50	9.39s	48	96%
SIFT	50	9.87s	48	96%

TABLE II. THE PERFORMANCE COMPARISON OF THREE MATCHING ALGORITHMS UNDER VIEWPOINT, SCALE AND ROTATION VARIATION CONDITIONS

Algorithm	Experiment times	Mean matching time	Correct matching times	Match ratio
Our algorithm	50	6.81s	42	84%
SIFT algorithm combined with saliency	50	9.02s	44	88%
SIFT	50	9.45s	44	88%



(a) (b)

Figure 4. Examples of the proposed algorithm results under illumination, viewpoint,rotation and scale variation conditions:(a) Matching result of airport. (b) Matching result of harbor.

B. *Experiment 2: Effect of the image size on the algorithm performance*

In order to validate that our algorithm has good robustness for the different size images under the same view, we take four groups of images, which respectively are 512×512、1024×1024、1536×1536 and 2048×2048, to do the experiment by comparing the proposed method with SIFT matching algorithm. Then we can get the time performance parameter ω that defines in (3) through respectively computing the averaged matching time that two algorithms take. From the results shown in table 3, we can see that, when the real-time images size change, comparing with SIFT matching algorithm, the improvement of time performance parameter of our algorithm remains stable. It shows that our algorithm has good robustness for images of different size.

TABLE III. TIME PERFORMANCE PARAMETERS UNDER THE CONDITIONS OF DIFFERENT IMAGE SIZE

	512×512	1024×1024	1536×1536	2048×2048
Average SIFT matching time/s	9.17	30.69	81.37	283.22
Average matching time of our algorithm/s	6.79	22.87	63.37	227.94
Time performance parameter ω	25.97%	25.48%	28.41%	24.25%

V. CONCLUSION

We have presented a novel SIFT matching algorithm which is fast and accurate. In order to improve matching performance, our algorithm combine SIFT matching algorithm with saliency detection and frequency domain segmentation in terms of that saliency detection can discard the redundant information and frequency segmentation can reduce matching time. To evaluate the performance of our algorithm, we have designed two groups of experiments aiming at verifying the performance of our algorithm on complex environment and different image size. The experimental results prove that the matching speed has been greatly improved, at the same time the matching accuracy can be well preserved. Our algorithm show its efficiency compared with SIFT matching algorithm. In addition, it has good robustness on the condition of the variation of illumination, viewpoint, rotation and scale.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (No.61071137, No.60802043), and a project from 973 Program of China (No.2010CB327900).

REFERENCES

[1] D.G. Lowe, “Object recognition from local scale-invariant features[C],” Proc. International Conference on Computer Vision (ICCV), Corfu, Greece, 1999:1150-1157.

[2] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," Proc. International Journal of Computer Vision (IJCV), 60(2):91–110, 2004.

[3] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," Proc. Computer Vision and Image Understanding (CVIU), 110(3):346–359, 2008.

[4] M. Calonder, V. Lepetit, C. Strecha, and P. Fua "BRIEF: Binary robust independent elementary features," Proc. Europe Conference of Computer Vision(ECCV), 2010.

[5] S. Leutenegger, M. Chli, and R. Siegwart, "BRISK: Binary robust invariant scalable keypoints," Proc. International Conference on Computer Vision (ICCV), 2011.

[6] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," Proc. International Conference on Computer Vision (ICCV), 2011.

[7] Itti L,Koch C, "Feature combination strategies for saliency based visual attention system," Proc. Journal of Electronic Imaging, 2001.

[8] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach,", Proc. Computer Vision and Patten Recognition(CVPR), pages 1–8, 2007.

[9] Goferman S, Zelnik-Manor L, "Context-Aware Saliency Detection," Proc. Computer Vision and Patten Recognition (CVPR), 2010.