Automatic splicing algorithm for building super viewing field from disordered image sequence

Cong Chen (陈 聪)[1,2] and Guixi Liu (刘贵喜)[1*]

[1]*School of Mechano-Electronic Engineering, Xidian University, Xi'an 710071, China*

[2]*Chinese Aeronautical Radio Electronics Research Institute, Shanghai 200233, China*

[*]*Corresponding author: gxliu@xidian.edu.cn*

Received December 27, 2011; accepted February 22, 2012; posted online June 20, 2012

An effective image splicing algorithm based on phase correlation and speeded-UP robust features (SURF) operator is proposed which can sort the disordered sequence and stitch them into a super viewing field image without any human intervention. Phase correlation in frequency domain is used for images sorting and region of interest (ROI) estimation, and guiding features extracting and matching in spatial domain by SURF operator and bidirectional best bin first (BBF) strategy. The experimental results demonstrate that this algorithm not only can deal with the input images with translation, rotation and scale changes, but also outperforms the pre-existing methods on the aspect of repeatability, efficiency and accuracy.

OCIS codes: 100.2000, 110.3010, 110.6980, 350.5730.

doi: 10.3788/COL201210.S11009.

Image splicing refers to stitching a series of partially overlapped images and blending them to create a large, seamless and high-resolution one. The automatic construction of super viewing field image from a disordered image sequence involves two crucial steps: sequence sorting and image stitching. The present splicing algorithms usually need artificial intervention or have restrictions on the source images[1,2].

Image stitching methods fall into the following categories: graph-theoretic methods, frequency domain methods, and feature-based methods[3]. However, each of them has its own characteristic and limitation. A variety of local invariant descriptors recently have made remarkable progresses. Scale-invariant feature transform (SIFT)[4] is invariant to image scaling and rotation and partially invariant to illumination and viewpoint changes. However, SIFT has a huge computational burden and cannot meet the real-time requirements in some applications. In 2006, speeded-up robust features (SURF) algorithm[5] is proposed, which approximates or even outperforms SIFT with respect to repeatability, distinctiveness, and robustness, and yet can be evaluated at a very low computational cost.

In this letter, an automatic super viewing field image splicing algorithm based on phase correlation and SURF operator is presented, which improves the efficiency of features detecting and accuracy of features matching obviously and obtains a satisfactory super viewing field splicing image without any manual interference.

The automatic sorting scheme is put forward in frequency domain based on phase correlation technique[6] which has high accuracy at the image translation and the correlation output shows a highly peaked shape. Phase correlation is based on the Fourier shift property, which states that a shift in the coordinate frames of two functions is transformed in the Fourier domain as linear phase differences. Let $f_1(x, y)$ and $f_2(x, y)$ be the two images that differ only by a displacement $(\Delta x, \Delta y)$ i.e.,

$$f_1(x, y) = f_2(x - \Delta x, y - \Delta y). \tag{1}$$

An unordered sequence with N images would be sorted automatically by the following steps:

Step 1: compute the normalized cross-power spectrum for each image with another, and then work out the principle peak value and translation parameters.

Step 2: pick out the image pair with the maximum correlation degree and they can be identified as the adjacent images in the sequence. According to the sign of horizontal displacement (Δx) to rank the neighboring images (f_1 and f_2 is supposed as Eq. (1)):

a) $\Delta x > 0$, f_1 is on the left of f_2;

b) $\Delta x < 0$, f_1 is on the right of f_2.

In addition, if Δx is over half of the image's width, it should be subtracted from the image width to obtain the real displacement. Up to now, N ranked image pairs are obtained and partly of them might be same.

Step 3: make a serial connection among the N ranking results:

a) The N image pairs can be chained with each other, then they join together to form the final order.

b) Not all of them can be chained. Examine all the possible permutations at the disconnected joint and consider the principle peak value to find the true joint. Then connect them to get the correct order.

SURF operator always extracts abundant and intensive feature points in the area where texture features are mass. To avoid that, we use SURF to detect interest points only in region of interest (ROI), and adopt the bidirectional best bin first (BBF)[7] matching strategy. Hence, the algorithm could dramatically cut down the number of extracted features and speed up the matching process, and certainly will greatly reduce the computational burden in image stitching.

Let $F_1 = \{f_1^1, f_2^1, \cdots, f_M^1\}$ and $F_2 = \{f_1^2, f_2^2, \cdots, f_N^2\}$ represent the SURF feature vectors extracted from the ROIs of image 1 and image 2, respectively. f_{NN} and f_{SN} are used to denote the nearest neighbor and next nearest neighbor, δ is defined as a threshold, and the bidirectional matching method is as follows:

a) Forward matching: $F_1 \to F_2$. For an arbitrary feature f_K^1 from F_1, find its f_{NN} and f_{SN} in F_2. f_K^1 and f_{NN} are matching successful if $\frac{f_{\mathrm{NN}}}{f_{\mathrm{SN}}} < \delta$. Then repeat the step to find all the forward matching pairs, and stored them in set V_1 .

b) Backward matching: $F_2 \rightarrow F_1$. For an arbitrary feature f_K^2 from F_2, find its $f_{\rm NN}$ and $f_{\rm SN}$ in F_1. f_K^2 and $f_{\rm NN}$ are matching successful if $\frac{f_{\rm NN}}{f_{\rm SN}} < \delta$. Then repeat the step to find all the backward matching pairs, and stored them in set V_2 .

c) Calculate the set V which is defined as the intersection of V_1 and V_2.

Although the bidirectional BBF can acquire relatively high rate of correct matching, there are still some error matched pairs. Random sample consensus (RANSAC)[8] is adopted to purify the matching features set.

For stitching the sequence images into a large one, the motion parameters model of adjacent images is needed. This transform is given as

$$sm' = Hm. \tag{2}$$

Where m and m' are corresponding pixels in two images, s is a scale factor and H is a non-singular matrix. H has 8 degrees of freedom and can be estimated linearly given 4 or more matching points. Because of the errors in feature matching and the instability of the perspective projection, we use Levenberg-Marquardt (L-M) algorithm[9] to optimize the estimated homography H.

To minimize the accumulated errors in the stitching, we choose the middle one in the ordered sequence as the reference and use an expanding way to map its adjacent image to it. Let I_k as the reference, for an arbitrary image $I_j (1 \leqslant j \leqslant n)$, the transformation from I_j to I_k is

$$H_{k,j} = \begin{cases} I (j = k) \\ H_{k,j-1} H_{k-1,j} (j < k) \\ H_{k,j+1} H_{k+1,j} (j > k) \end{cases} . \tag{3}$$

where I denotes the identity matrix. Then the average weighted strategy is applied for blending to improve the non-uniform brightness in the overlapping regions.

All the experiments for this article have been done on PC AMD Athlon X2 240, 2.81 GHz and 2G RAM with Windows XP, using VS2008 and OpenCV library.

The flow diagram of the proposed algorithm is given in Fig. 1. The frequency method acts as a guide for the following operations in the spatial domain.

Figure 2 shows the set of disordered images from IPM Vision Group[10]. They were taken from different angles with resolution of 400×300 (pixel). Table 1 gives the maximum correlation degrees and translation parameters calculated by phase correlation method. According to the sorting strategy described in this study, we can get the serial connection that is 3→4 and 1→5→2. Then the correlation between 1 and 4 is compared to that between 2 and 3, and the former one is larger. In addition, the translation parameter of image 1 and 4 is negative which indicates that image 1 is on the left of 4. Hence, the final order of the image sequence is 3→4→1→5→2, as shown in Fig. 3.

The images in Fig. 4(a) are chosen from the ordered sequence in Fig. 3. Figures 4(b) and (c) are the results by using SURF and our method for features detecting. Figure 5 shows the rough matching and image stitching result. The concrete data is in Table 2. Table 3 shows

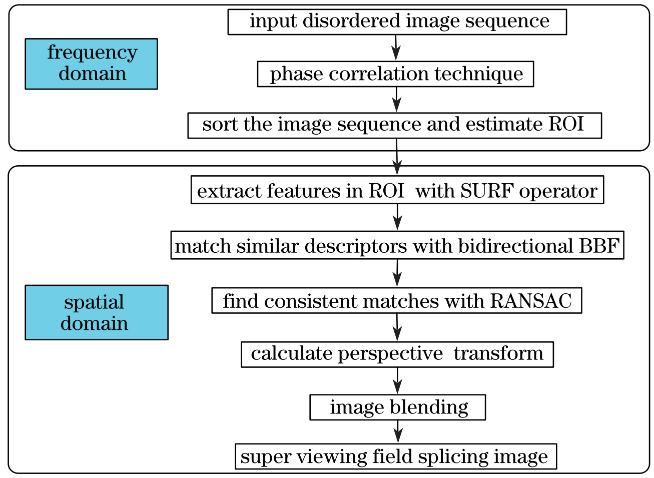
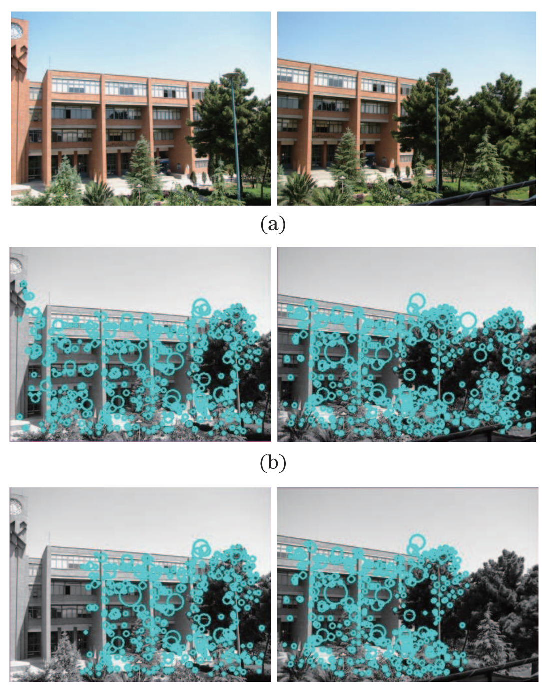


Fig. 1. Flow of the automatic image splicing algorithm.

Fig. 2. Disordered image sequence.

Fig. 3. Ordered image sequence.



(a)

(b)

(c)

Fig. 4. Original images and features extraction. (a) Original images; (b) detected by SURF; (c) detected by our method.

Table 1. Maximum Correlation Degrees and Translation Parameters

Image	1	2	3	4	5
1	/	0.0147	0.0155	0.0207 (−118)	0.0256 (−152)
2		/	0.0138	0.0155	0.0162 (+96)
3				0.0155 (−148)	0.0136
4					0.0139
5					

Table 2. Performance Comparison of Different Methods

Method	Feature Descriptors			Matched	Repeatability	Total
	Image1	Image2	Time (ms)	Pairs	Score	Time (ms)
SIFT	482	519	2 812	150	29.97%	3 125
SURF	472	511	580	191	38.86%	615
Our Method	366	389	423	189	53.72%	472

Table 3. Comparison of Various Matching Strategy

Stage	Forward Matching	Backward Matching	Bidirectional Matching
Rough Matching	199	202	187
Precise Matching	176	180	170
Accuracy Rate	88.44%	89.11%	90.91%

the different matching strategies in detail. The threshold of matching was set as 0.55.

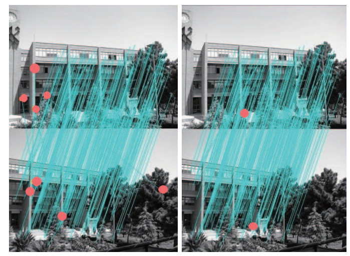
From Fig. 5(a), we can see 4 and 1 visible mismatched pairs marked with the red points respectively. Figure 5(b) proves the good applicability of our method. To evaluate the different methods at the same level, we modified the parameters of SIFT to obtain the similar features number that extracted by SURF, and also worked out the repeatability score. It can be inferred from Table 2 that SURF is about 5 times faster than SIFT. Moreover, our method has improved the computational efficiency by 27% comparing with SURF, and improved by 85% comparing with SIFT. Data recorded in Table 3 confirms that the bidirectional matching strategy improves the accuracy of correct matching to some extent.

Two examples are presented here to confirm the good performance of the proposed algorithm.

Experiment 1: figure 6 shows the super viewing field image splicing from image sequence in Fig. 3. Figure 6(a) illustrates the expanding way of splicing the small images into a large one, and Fig. 6(b) is the output super viewing field image constructed by our method. It can be seen that our splicing image is clarity, smooth transition, and no distortion. The super viewing field image has a good visual effect.

Experiment 2: the original disordered images in Fig. 7(a) were taken by us with digital camera in a cloudy weather, and there exist some displacements and rotations. The proposed algorithm sorts the source images and outputs a super viewing field image without any manual intervention, prior information or restrictions, as shown in Figs. 7(b) and (c).

In conclusion, this letter presents an effective image splicing algorithm which can sort the disordered sequence without human intervention and stitch them into a super viewing field image fully automatically. It is able to deal with the input images with translation, rotation and scale changes, and could cut down the calculated amount dramatically. The super viewing field image obtained by


(a)


(b)

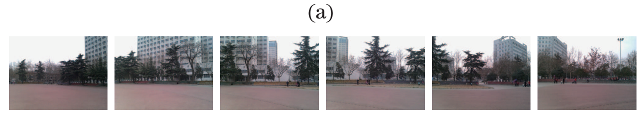
Fig. 5. Features matching and image stitching. (a) Matched by SURF and our method; (b) stitching result of our method.


(a)


(b)

Fig. 6. Automatic image splicing from disordered image sequence 1. (a) Set the middle image as reference and map the adjacent to it; (b) output 1 248×384 (pixel) super viewing field image.


(a)


(b)

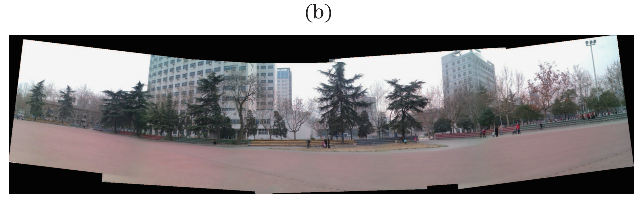

(c)

Fig. 7. Automatic image splicing from disordered image sequence 2. (a) Disordered 640×480 (pixel) image sequence; (b) ordered image sequence; (c) output 2 424×618 (pixel) super viewing field image.

this method shows the good visual effect and it is clarity, smooth transition, and no distortion.

This work was supported by the National Ministries Foundation of China (No. 9140A16050310DZ01).

References

1. H. Zhao, H. Chen, and H. Yu, J. Image and Graphics (in Chinese) **12,** 336 (2007).
2. X. Wu, B. Guo, and J. Wang, J. Optoelectronics Laser (in Chinese) **20,** 1114 (2009).
3. B. Zitová and J. Flusser, Image and Vis. Comput. **21,** 977 (2003).
4. G. Lowe, International J. Computer Vision **60,** 91 (2004).
5. H. Bay, A. Ess, T. Tuytelaars, and L. Vangool, Computer Vis. and Image Understanding **110,** 346 (2008).
6. B. S. Reddy and B. Chatterji, IEEE Tran. Image Processing **5,** 1266 (1996).
7. J. S. Beis and D. G. Lowe, in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1000 (1997).
8. M. A. Fischler and R. C. Bolles, Commun. ACM **24,** 381 (1981).
9. M. I. A. Lourakis, Matrix **3,** 2 (2005).
10. Image Stitching Database: http://math.ipm.ac.ir/vision/VDownloads.html.