Image Mosaic Using ORB Descriptor and Improved Blending Algorithm

TIAN Fang, SHI Ping
School of Information Engineering
Communication University of China
Beijing, China

ABSTRACT: *This paper presents an image mosaic technique using ORB (oriented FAST and rotated BRIEF) descriptor and an improved blending algorithm. Features of images with overlapping regions are first described by ORB descriptors. This binary descriptor outperforms SIFT and SURF in regard of processing speed. RANSAC algorithm is then applied to reject mismatches. Once the best projective deformation matrix is found, brightness equalization is conducted and an improved weighting function is utilized to blend images and reduce visible artifacts. Experimental results show this approach to be efficient in accounting for image mosaic, and the edge artifacts could be significantly reduced and even completely eliminated with the blending process designed in this paper.*

Keywords: Image Mosaic, ORB Descriptor, RANSAC, Image Blending

Received: 17 June 2014, Revised 25 July 2014, Accepted 29 July 2014

1. Introduction

The development of remote monitoring and video conference calls for higher resolution and wider optic angles of cameras. Special equipments like fisheye lens offer wide view angles yet generate image distortion, whose correction is of high complexity. An alternative way to achieve higher resolutions is the technique of image mosaic, which registers images with overlapping regions and composites them together into large mosaics of the whole scene. With such process the mosaic could theoretically achieve unlimited resolution and wideness.

As the core problem of image mosaic, image registration determines the accuracy of the final composite. Current registration approaches are based on intensity, wavelet or frequency domain, and image features [1]. Feature detection and extraction enjoy high popularity owing to its simplicity and robustness. Typical image features and feature descriptors include gray edge feature [2], Harris corner [3], SIFT feature descriptor [4] and SURF feature descriptor with color invariants [5]. SIFT algorithm detects and describes local features of images and is scale-, rotation- and affine-invariant. However, SIFT algorithm performs poorly in real time applications due to its complexity [6]. SURF is several times faster than SIFT and robust against image transformations, but remains time-consuming on account of its floating point eigenvectors [7].

ORB (oriented FAST and rotated BRIEF) is an efficient binary key point detector and descriptor proposed recently [8]. It

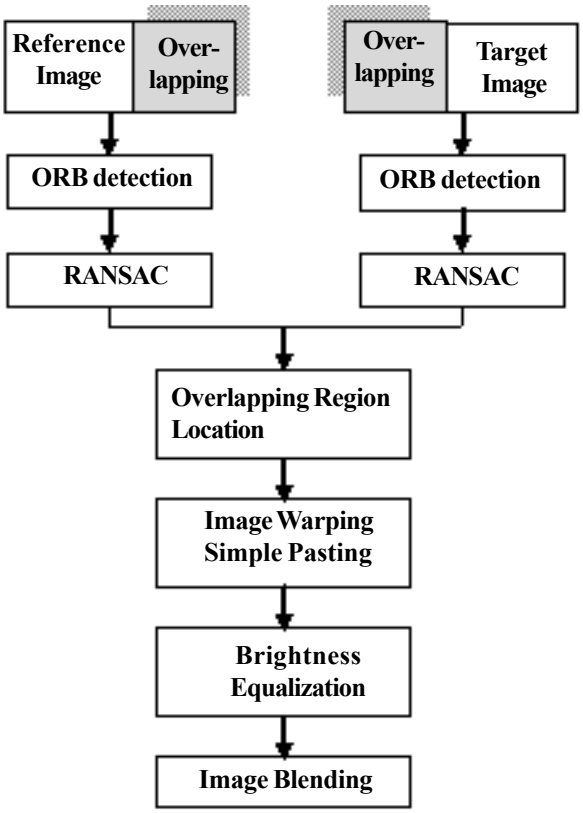


Figure 1. System Design

inherits the speed and low cost of BRIEF descriptor [9], and improves its rotational invariance. ORB outperforms SIFT and SURF in speed and could apply to real time applications.

In this paper, we present an approach of image mosaic using ORB descriptor and an improved blending algorithm. In Section II, the general procedure of image mosaic and design of the proposed approach are described. Section III details ORB and RANSAC algorithms applied in our approach. Principles of the proposed blending method and the comparison with its counterparts are presented in Section IV. Experimental results are shown in Section V to validate our approach, while conclusion and discussion are given in Section VI.

2. Procedure of image mosaic

The general procedure of image mosaic, regardless of concrete methods, consists of the following steps [10]:

• **Image Acquisition:** Original images for mosaic could be captured either by panning or tilting a camera which is mounted on a tripod. A more flexible way is to handhold a camera and capture images with scale, range and viewpoints differences. Images we use to test the validity of our algorithm are acquired in this way.

• **Image Registration:** Images with overlapping regions are transformed into the same coordinate system for pasting and blending in this step. Feature-based registration establishes a point-to-point correspondence of key points within the overlapping area. With such correspondence, a geometrical transformation could be computed to map the target image to the reference one. Note that the size of the overlapping region should be cautiously chosen: too small regions may result to inaccuracy of registration for the shortage of key point pairs, while too large regions are redundant. Our algorithm sets a threshold of the number of key point pairs to determine whether the size of the overlapping region between the reference image and the target image is suitable for mosaic.

• **Image Warping:** Once the projective transformation matrix, which represents the mapping relation between every corresponding point pair within two images, is obtained, the target image would be warped to the reference coordinate by this

matrix for further processing.

• **Image Blending:** Differences of brightness and viewpoints between reference images and warped target images could lead to obvious artifacts if simply pasted together. Accordingly, aims of this step are to eliminate artificial edges and to blend images seamlessly.

In reference of the procedure explained above, we designed a system using the proposed algorithm.

3. ORB Matching and RANSAC Mismatch Rejection

3.1 ORB Feature Descriptor

ORB is a rotation-invariant, noise-resistant and very fast binary descriptor built on BRIEF (Binary Robust Independent Elementary Features). Generated by simple intensity difference tests, BRIEF evaluates similarities between descriptors with Hamming distance, thus is very fast to build and to match. Despite that it is highly discriminative and strikingly efficient, it lacks sufficient scale- and rotation-invariance [9]. ORB enhances the performance of BRIEF by computing an orientation component to FAST (Features from Accelerated Segment Test) and adding it to BRIEF features.

FAST is an ideal choice for finding key points that match visual features, nevertheless it does not produce a measure of conerness and lacks of multi-scale features [11]. To fill these gaps, ORB first employs a Harris corner measure to order FAST key points, then utilizes a scale pyramid of the image with each level producing certain FAST features. Orientation of FAST features is produced by intensity centroid, which assumes that there is an offset between the intensity of a certain corner and its center. The orientation of a patch is:

$$\theta = a\tan 2\,(m_{01}, m_{10}), \tag{1}$$

where m_{pq} represents the moments of a patch and is defined as:

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y) \tag{2}$$

ORB then uses the orientation of a patch to steer the original BRIEF descriptor. For a smoothed patch P, a binary test on the patch could be defined as

$$\tau(p; x, y) = \begin{cases} 1 & p(x) < p(y) \\ 0 & p(x) \geq p(y) \end{cases} \tag{3}$$

where $p(x)$ is the intensity of patch P at point (x, y). With n such binary tests on this patch, the feature could be defined as a vector:

$$f_n(p) = \sum_{1 \leq i \leq n} 2^{i-1} \tau(p; x_i, y_i) \tag{4}$$

Performance of BRIEF depends on distributions of tests. Experimental results show Gaussian distribution around the center of a patch to be the best one. Moreover, $n = 256$ is demonstrated as a proper number of tests.

To steer BRIEF in accordance with the orientation of key points, ORB defines any feature set of n binary tests at (x_i, y_i) as

$$S = \begin{pmatrix} x_1 x_2 \cdots x_n \\ y_1 y_2 \cdots y_n \end{pmatrix} \tag{5}$$

then a steered BRIEF becomes

$$g_n(p, \theta) = f_n(p) \mid (x_i, y_i) \in S_\theta, \tag{6}$$

where $S_\theta = R_\theta S$ with S_θ being the corresponding rotation matrix.

ORB also develops a learning method for choosing binary test with low correlation and improving variance. This method could be regarded as a greedy search for a set of binary test, which meets the requirements of having the highest variance and the lowest correlation. The variance of tests is evaluated by their distance from a mean of 0.5, with the best one being closest to it. The improved version of BRIEF descriptor prevails over the steer BRIEF in diversity.

When detecting roughly the same number of features on same images, ORB could be an order of magnitude faster than SURF

and more than two orders faster than SIFT [8]. We have compared the performance of ORB and SURF with 640 × 480 images captured by handheld cameras. TABLE I shows the time consumed for feature detecting using ORB and SURF of 10 image pairs.

Note that the contrast tests were conducted using the same personal computer whose CPU, operating system and installed RAM remains unchanged during all the tests. Obvious from TABLE I is that the detecting speed of ORB is outstanding, which verifies the conclusion drawn by [8].

3.2 Projective Transformation

Different static 3D scenes whose objects are located approximately in the same plane could be denoted by 2D points in the image plane and described by a 2D planar projective transformation using the following matrix multiplication [12]:

$$\begin{bmatrix} X_2 \\ Y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \\ 1 \end{bmatrix} \tag{7}$$

(a)

(b)

Figure 2. Key point pairs before and after RANSAC

where (x_1, y_1) and (x_2, y_2) denote the 2D points in reference image and target image respectively. The projective transformation has *8* degrees of freedom, say h_1, h_2, …, h_8. h_1, h_2, h_4 and h_5 represent rotation parameters and scale parameters; h_3 and h_6 denote horizontal displacement and vertical displacement separately; h_7 and h_8 are the horizontal and vertical deformations.

The system we designed regards the first input image as the reference with the overlapping region on the right, and the second input as the target with a left-located overlapping region. The projective transformation H is computed to map the

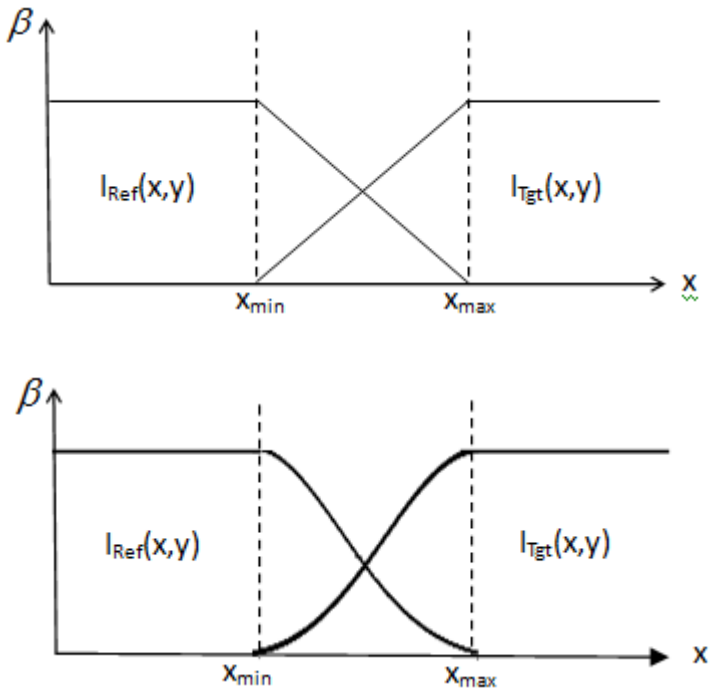


Figure 3. Distribution of $\beta(x, y)$ in linear cross-fade method

second image to the first one. When images inputted are inverted, the inverse matrix of H should be utilized for mapping. We control the sequence of input images by comparing numbers of key points on the left half and right half of an image.

3.3 RANSAC Mismatch Rejection

The key point pair set detected by ORB contains mismatches among those good matches inevitably. For this reason, we apply RANSAC to refine the raw set of matching pairs and reject mismatches [13].

The basic assumption of RANSAC is that one set contains data which fit a certain model (inliers) and those whose distribution doesn't (outliers). RANSAC refines the data set by estimating parameters of the model which encompasses the largest number of good matches. In the case of image matching, this model becomes a projective transformation matrix H with *8* degrees of freedom (say, $h_1, h_2, \ldots, h_8$ in (7)). When conducting RANSAC,

• Four point pairs within the overlapping region are chosen randomly to compute the initial matrix H_0;

• When the error between the computational (x_2', y_2') and (x_2, y_2) falls in a certain range, the point pair (x_1, y_1) and (x_2, y_2) is supposed to be inliers, and the total number of point pairs under this model H_i *is counted and denoted by* M_i;

• Repeat the above-mentioned two steps until certain H_k produces the largest M_k, this H_k is thus considered to be the best transformation matrix and those pairs subjected to this model are good matches.

A comparison of key point pairs before and after RANSAC refinement is given in Figure 2 (a) and Figure 2 (b) respectively. The matching points detected by ORB are lined with color lines for clarity. Figure 2 (a) has enough key point pairs for mosaicking, yet contains a few ones which have been mismatched. Surrounding areas of mismatching pairs share similar shapes or corners, making it difficult for ORB to distinguish and resulting to mismatch. Nonetheless, with its greedy search algorithm, RANSAC is able to tell the difference between the true matching pairs and the false ones, eliminating incorrect points and retain adequate accurate ones, as shown in Figure 2(b). RANSAC could cut down the time consumed for mosaicking, for the time saved by rejecting mismatches exceeds the time consumed for conducting RANSAC.

Table 1 shows the number of pairs before and after RANSAC refinement. Statistical data also demonstrate its effectiveness of

rejecting mismatches and reducing number of pairs waiting for further processing.

4. Improved image blending algorithm

Current image blending techniques include weighed mean method [14] [15], multi-resolution spline method [16] and wavelet transform method [17] [18] [19]. The main idea of weighed mean methods is that pixels at different part of the overlapping area contribute differently to the final composite. They are easy to realize and perform well. Consider a pixel $I(x, y)$ in the overlapping area of the composite, it could be defined as

$$I(x,y) = \beta(x,y)\, I_{Ref}(x,y) + (1-\beta(x,y))\, I_{Tgt}(x,y) \tag{8}$$

with $I_{Ref}(x, y)$ and $I_{Tgt}(x, y)$ being pixels in the reference image and the target image at the same location of the overlapping region. Weighting function $\beta(x, y)$ is subjected to the position of the pixel and would change from 1 to 0 when (x, y) remotes from the left edge of the overlapping area and gets closer to the right edge.

Among all the variances of weighed mean methods, cross-fade function is a commonly used approach [20]. The weighting factor $\beta(x, y)$ of a point (x, y) is defined as:

$$\beta(x,y) = 1 - \frac{x - x_{min}}{x_{max} - x_{min}} \tag{9}$$

with x_{min} and x_{max} representing the left and right edge of a certain row in the overlapping region respectively. Figure 3(a) shows the distribution of $\beta(x, y)$ within a composite image.

Despite of its comparatively satisfying performance, linear cross-fade function generates sudden change at the edge of the overlapping area and causes blurring. As a result, artificial seams still exist when scrutinizing the mosaic (See Figure 4 (b)). In this paper we employed a variant of Gaussian function as the weighting function and proposed an improved algorithm to eliminate artifacts.

4.1 Brightness Equalization

Inconformity of brightness between images and color distortion at the edge of each image accounts for an unnatural color transition of the mosaic image (See Figure 4 (a)). To address this problem, brightness equalization is involved in the proposed algorithm with the following steps:

• Converting both reference and target images from RGB color space to HSV color space.

• Computing V_{iRef} / V_{iTgt} of each pixel pair p_i within a certain part of the overlapping region. In this paper we choose 20% of the

Image Pair No.	Parameters		Good Matches		Detecting Time	
	Over-lappin g (%)	Rotation (°)	ORB	RANSAC	ORB (ms)	SURF (s)
1	84.5	4	294	219	91.6	3.884
2	77.2	11	177	156	109.5	5.819
3	71.1	15	182	103	139.7	5.023
4	62.2	7	228	206	122.9	3.759
5	55.0	3	207	173	112.9	4.586
6	50.6	4	136	115	106.7	4.025
7	49.4	0	171	153	96.5	5.273
8	34.5	0	49	28	113	3.853
9	34.2	5	98	64	118.6	4.415
10	25.6	8	29	9	107	2.449

Table 1. Performance of the Algorithm

overlapping area close to the seam as the part of interest;

• Adjusting V of every pixel in the target with the mean of ratios calculated in the previous step.

After such processing steps, the discrepancy of brightness between the reference image and the target could be significantly weakened, bringing about a better mosaic result compared with the image pairs without equalization.

4.2 Weighting function based on Gaussian model

Gaussian function is nonlinear and ranges smoothly between 0 and 1. In addition, its convergence rate could be adjusted for various purposes. The proposed algorithm in this paper employs a variant of Gaussian model, with the value reaching 1 at x_{min} and 0 at x_{max}. The Gaussian weighting function is defined as follows:

$$\beta(x,y) = 2e^{-\frac{(x-\mu)^2}{2\sigma^2}} - \frac{1}{2}, \tag{9}$$

where $\mu = x_{max}$ and $2\sigma^2 = (x_{max} - x_{min})^2$. The distribution of β (x, y) is given in Figure 3 (b).

The variant of Gaussian model we used takes only a half part of the standard Gaussian forms; in addition to such modification, we revised its coefficients to make it converge at x_{max} and reach 1 at x_{min}. This weighting function keeps the smoothness of the original Gaussian function and could avoid blurring.

Figure 4 (b) and Figure 4 (c) show the blending results using linear weighting function and the composite image after conducting brightness equalization and Gaussian cross-fade method respectively. When simply attaching the target image to the reference, the result image contains an unbearable stitching seam, as emphasized with a yellow circle in Figure 4 (a). With linear weighting functions, the seam becomes a little vaguer than the original one but still perceptible, as shown in Figure 4 (b). After brightness equalization, the luminance difference between two sides of the seam is largely suppressed, and the artifact is also weakened. With the blending method proposed in this paper, the stitching seam has been removed on the whole, and colors of the result image transit naturally from the reference to the target, as circled in Figure 4 (c).

It is evident in Figure 4 that our method could effectively reduce artifacts in comparison with the linear cross-fade approach.

5. Results

The performance of this proposed system is evaluated by image pairs of 640 × 480 which are highly discriminative in contents and captured by handheld cameras. Also, the images are of various formats like bmp, jpeg and png. The computer used for testing was equipped with an Intel (R) Core (TM) i7-2600 CPU and had a RAM of 4GB. The system was programmed in C++ language using VS2010. Figure 5 (a) shows a sample pair of reference and target images inputted in the system. The mosaic result is presented in Figure 5 (b).

Images in Figure 5 (a) share 53% of its original size and the target image has rotated for about 13.8°. Processed by the above-mentioned mosaic steps, the artifacts in Figure 5(b) are largely eliminated and the satisfying result demonstrates the effectiveness of the proposed blending method.

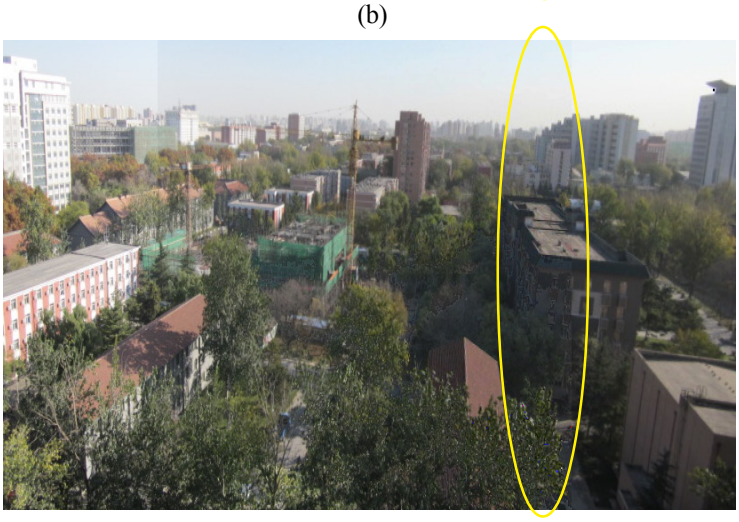
Characteristics of 10 image pairs from the testing set are given in Table 1. Main concerns of the experiment include the availability of this method for different overlapping size, the robustness against rotation, and the detecting speed in compare with SURF. All of the contrast tests were conducted using the same personal computer whose CPU, operating system and installed RAM remains unchanged during the tests.

As shown in Table 1, the proposed system could process images with a limited overlapping area (Image pair No.10 only shares 25.6% of the original size) and a large rotation angle (since larger angles may result in visual discomfort, rotation angles of image pairs in our tests are confined within 15 degrees).

Numbers of good-matching pairs is considerably reduced by RANSAC refinement, yet retains enough pairs to describe the

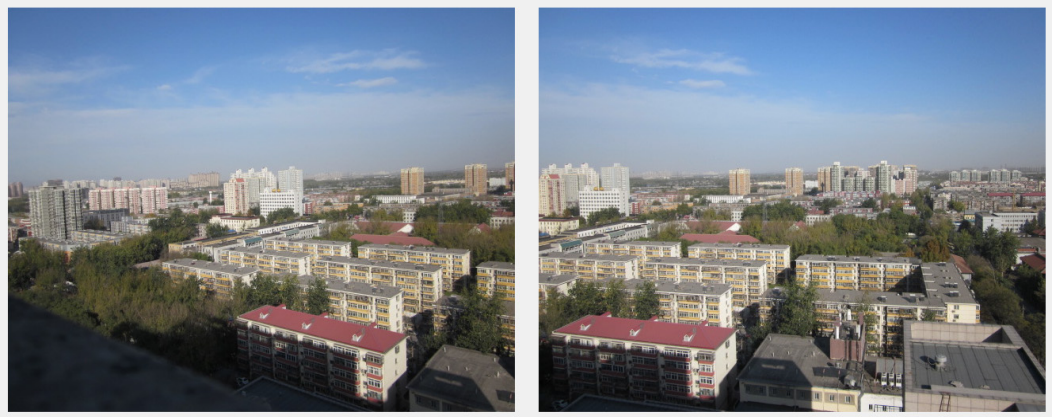
(a)

(b)

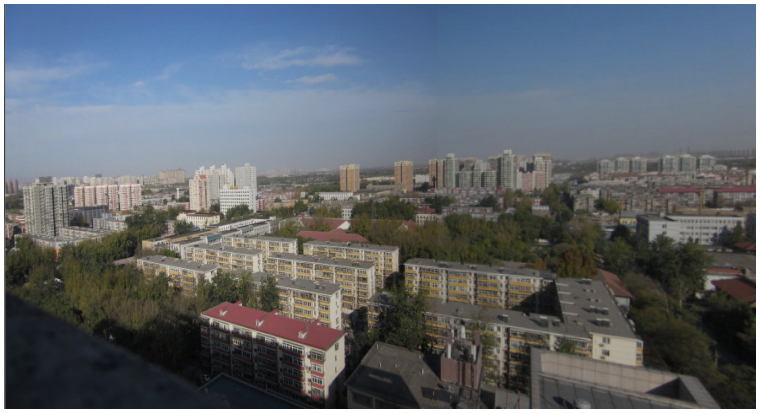


(c)

Figure 4. Composite before brightness equalization (a), blending results of linear cross-fade (b) and Gaussian cross-fade method (c)



(a) (b)



(c)

Figure 5. Images before (a) and after (b) mosaic

correlation between images.

"*Detecting Time*" refers to the time consumed by ORB and SURF to detect roughly the same number of key points in same image pairs. Clearly from the data is that ORB could be at one order of magnitude faster than SURF, which verifies the conclusion drawn by [8]. Note that the overlapping size, rotation angle and contents of images could influence the number of good matches and processing speed.

Owing to its efficiency, this algorithm is suitable for real time applications such as surveillance, which requires real time processing speed and decent robustness.

6. Conclusion and Discussion

The main contribution of this paper is the design and implementation of an image mosaic system. The approach proposed by this paper-namely, introducing ORB descriptor to detect and to describe feature points quickly, and refining the key point set with RANSAC method to reject mismatching and to speed up processing, then employing brightness equalization method to pre-remove stitching seam and to smooth color transition, and finally blending the target with the reference with Gaussian weighting functions—has a compelling advantage over other stitching methods with SIFT and SURF descriptor, or those with linear weighting functions. Experiments also demonstrate this approach's robustness against confined size of overlapping area and rotations within a certain limit. In short, the system is proven to be fast, accurate and robust.

Image mosaic provides us with a brand-new vision to scrutinize the world. It creates detailed environmental scenes needed for surveillance, virtual reality and other applications, without investing in luxurious equipments like fish-eye cameras. A promising extension of image mosaic is video mosaic, which stitches every corresponding frame of two video clips and generates a

dynamic presentation with a broader view. The approach proposed seems especially fit for video-based mosaic application, owing to its simplicity and speediness. However, according to experimental data listed above, the average time consumed to obtain matching point pairs in a relatively small image (640 × 480) is over 100*ms*, let alone taking the time used for following processing steps or larger images into consideration. If applying the approach to video, it is unlikely to meet the requirements of real-time processing, and the result video would be slower than the original versions. Consequently, further study should be focused on the nature of videos and improvement of procedures to accelerate the process.

Image mosaic and subsequent techniques will definitely open up new study and application domains in computer graphics, virtual reality and other practical areas.

References

[1] ZHANG Ya-juan. (2013). Research on image and video mosaic based on SURF. Xidian University Master's Thesis.

[2] LIU Ying., CAO Jian-zhong., XU Zhao-hui., et al. (2007). Improvement of image matiching algorithm based on gray correlation. *Jounal of Applied Optics*, 28 (9) 536-540

[3] LI Huan-huan., HUANG Shan., ZHANG Hong-bin. (2012). The automatic image mosaic based on Harris and SIFT algorithms. Computer Engineering & Science, 34, p.104-108

[4] LI Yun-xia., ZENG Yi., ZHONG Rui-yan., GUO Tao. (2012). Algorithm of image stitching based on SIFT feature matching.Computer Technology and Development, 19, p.43-52

[5] GAO Sheng-feng., WANG Zheng-yong., ZANG Yuan-yuan., HE Xiao-hai., (2013). Color image stitching based on color invariant and SURF. Computer Engineering and Applications, p.113-128

[6] Lowe., David, G. (2004). Distinctive image features from scale-invariant key points, *International Journal of Computer Vision*, 60, p. 91-110

[7] Herbert Bay., Andreas Ess., Tinne Tuytelaars., et al. (2008). SURF: Speeded up robust features. *Computer Vision and Image Understanding* (CVIU), 110 (3) 346-359.

[8] Ethan, R., Vincent, R., Kurt, K., et al. (2011). ORB:an efficient alternative to SIFT or SURF. *In*: Proc of International Conference on Computer Vision. p. 2564-2571

[9] Calonder, M., Lepetit, V., Konolige, K., et al. (2010). Brief: Binary robust independent elementary features. *In:* European Conference on Computer Vision.

[10] WANG Jun-Jie, LIU Jia-mao, HU Yun-fa, YU-Yu. (2013). Research and development of image mosaics. Computer Science, 30, p.141-150.

[11] Rosten, E., Drummond, T. (2006). Machine learning for highspeedcorner detection. *In:* European Conference on Computer Vision, V. 1.

[12] Ma Songde, Zhang Zhengyou. (1998). Computer Vision-Compute Theory and Arithmetic Foundation(in Chinese). Beijing: Science Press.

[13] Fischler, M. A., Bolles, R. C. (1981). Random Sample Consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* , 24 (6) 381–395.

[14] Jia,Y. H. (1998). Fusion of Landsat TM and SAR images based on principal component analysis. *Remote Sensing Technology and Application*, 13 (1) 46-49

[15] Shun, H. Y., Sezeliski, R. (2000). Construction of panoramic mosaics with global and local alignment. *International Journal of Computer Vision*, 36, p.101-130.

[16] Peter, J.B. & E.H. Adelson. (1983). A multiresolution spline with application to image mosaics. ACM Transactions on Graphics, 2 (4), p.217-236

[17] Ranchin, Wald, T. & L. (1993). The wavelet transform for the analysis of remotely sensed images. International Journal of Remote Sensing, 14 (3), p .615-619

[18] Chipman, L. J., Orr, T. M., Graham, L.N. (1995). Wavelets and image fusion. Orlando: In Proc. of Int. Conf. on Image Processing, p.248-251

[19] Li, H., Manjunath, B.S., Mitra, S. (1995). Multisensor image fusion using the wavelet transform. Graphical Models and Image Process, 57 (3), p. 235-245

[20] LIN, Lu. (2010). Research on image registration and stiching techniques. Changchun University of Science and Technology Master's Thesis.