Wide Baseline Image Stitching with Structure-preserving

Mingjun Cao, Wei Lyu, Zhong Zhou[*], Wei Wu
State Key Laboratory of Virtual Reality Technology and Systems
School of Computer Science and Engineering, Beihang University, Beijing, China
zz@buaa.edu.cn

Abstract—**This paper presents a novel stitching approach for wide-baseline images with low texture. Firstly, a three-phase feature matching model is applied to extract rich and reliable feature matching, in the case of low texture, our line matching and contour matching will compensate for the poor quality of point matching. Then, a structure-preserving warping is performed, by defining several constraints and minimizing the objective function to solve the optimal mesh, with which we obtain multiple affine matrices to warp images. Furthermore, we synthetically consider alignment error, color difference and saliency difference to find the optimal seam for image blending. Experiments both on common data sets and challenging surveillance scenes illustrate the effectiveness of the proposed method, and our approach has outstanding performance when compared with other state-of-the-art methods.**

Keywords-stitching; wide-baseline; low texture; three-phase feature matching; structure-preserving warping.

I. INTRODUCTION

Traditional image stitching assumes that the camera is at a fixed viewpoint or the scene is roughly planar, these two assumptions both require that there is no great depth change in the image. Violation of above assumptions, it is obviously inadequate to use only one global homography matrix for image alignment. Since the content of the image varies in depth, there will be noticeable artifacts in the final panorama. The misalignment between the target image and the reference image is usually called parallax.

However, for surveillance images, the situation is usually more challenging. In daily life, camera position, orientation, and other attributes vary much, thus the quality is rather poor, even the camera is affected by the surrounding environment, causing problems such as pollution, sheltering, blurring, etc. For this kind of images with wide-baseline, large parallax, and low texture, existing stitching algorithms cannot achieve satisfactory result, even some commercial softwares directly say that these videos cannot be stitched together.

In this paper, to overcome the problems mentioned above, we propose a novel framework which combines the advantages of several advanced approaches: firstly, a local homography model based on super pixel segmentation is applied to obtain rich and reliable feature matching, in the case of low texture, we add line matching and contour matching to compensate for the poor quality of point matching; then we perform a structure-preserving warping by combining several constraints and minimizing the objective function to solve the optimal mesh, with which we obtain multiple affine matrices to warp images; Finally, we synthetically consider alignment error, color difference and saliency difference to find the optimal seam for image blending.

II. RELATED WORK

In this section, we will briefly review the related work and the latest progress, the specific content can be referred to the literature review of image stitching [1].

A. Image alignment

Traditional method aligns images with one global homography matrix, which work well when the scene is planar and camera undergoes only rotation. But for images with large parallax, it will cause noticeable artifacts. Gao et al. [2] divided the image into distant plane and close plane, each was aligned with one homography matrix, which works well when the scene contains mainly two planes. Lin et al. [3] proposed to employ a smoothly varying affine model to align images, which works well with moderate parallax. Zaragoza et al. [4] divided images into hundreds of grids, each grid using smoothly varying homography to align images. In this paper, a local homography model is used to filter feature points on different planes, which works even better.

B. Image blending

Now mainstream approaches are based on the graph-cut, while only taking color difference as the quantitative standard will cause obvious structural fracture and seriously affect the quality of panoramic image. Zhang et al. [5] combined alignment error and color difference to make the seam across areas with good alignment. Lin et al. [6] combined color difference and saliency information to define the difference map so that the seam will not break significant structure. Our method is the combination of them, also has the advantages of both.

C. Image naturalness

Traditional global homography model usually sets an image as the reference plane, so images away from the reference view will suffer obvious distortion. Change et al. [7] proposed a SPHP warp, which smoothly transforms the homography of overlapping region into the similarity of non-overlapping region. Lin et al. [8] proposed an AANAP warp, which combines linearized homography and global similarity to generate nature panorama. Chen et al. [9] proposed a GSP warp, which optimizes naturalness of panorama by adding global similarity constraint. Unlike previous methods, reference plane is not set in our method, but solved with scale-preserving constraint, so the final panorama suffers less distortion.

D. Mesh optimization

Mesh optimization is widely used in image retargeting [10], [11], image resizing [12], [13], image rectangling [14], video stabilization [15], [16] and so on. Recently, mesh optimization has been tried for local warping in image stitching and has a good effect.

III. THREE-PHASE FEATURE MATCHING

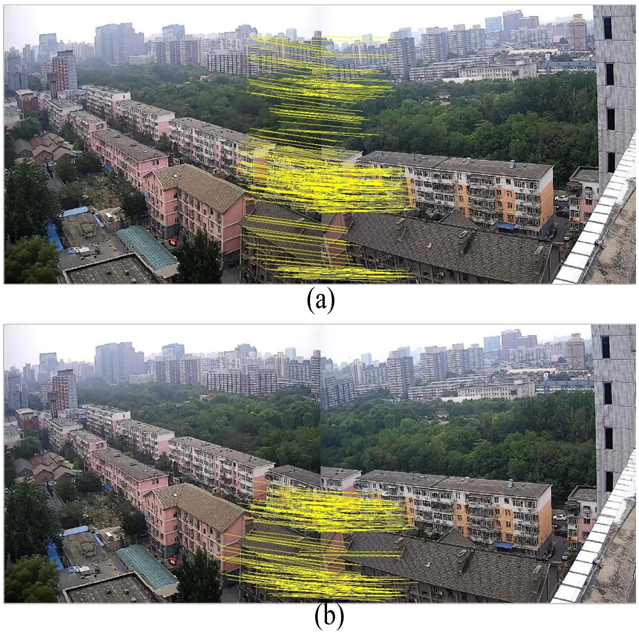
In the feature matching part, a three-phase matching strategy is adopted, namely, point matching based on local homography, line matching based on neighboring feature points and contour matching based on neighboring pixels. Point matching is the first and most important matching method. Considering the clarity and texture characteristics of surveillance scene, line matching and contour matching can be introduced as a supplement to a certain extent.

A. Point matching based on local homography

Like many previous method, we use SIFT [17] to detect feature points, for challenging data, ASIFT [18] can be adopted, and then we use K-D tree to perform initial matching of feature points. As for outlier rejection, we combine Zhang's [5] plane hypothesis and Lin's [6] method of super pixel segmentation. First, we use SLIC [19] to perform super pixel segmentation on image pair (I_i, I_j), and the number of super pixels is set to 50. Next, we take I_i as reference, for each super pixel of I_j, we use DLT [20] to estimate a homography H for all feature points within this super pixel, if residual error γ is less than 5 pixels, we take it as an inlier, after enumerating all feature points in I_j, we obtain the inlier set S_1. Then, we swap I_i and I_j to get the inlier set S_2. Finally, the inlier set for (I_i, I_j) is $S_1 \cap S_2$.

B. Line matching based on neighboring feature points

Just like [21], our line matching is based on the result of point matching. First, we use the LSD [22] algorithm to detect lines; then, we divide neighboring feature points set for each line according to the distance, if the distance between the feature point and the line is less than 30 pixels, the feature point will be included in neighboring set of the line; finally, we match each line and its neighboring set together as a line-points variants to generate the voting matrix, if there are some feature points not used, we can generate some candidates from the already matched pairs, and the final results are selected from the voting matrix.



(a)

(b)

Figure 1. Point matching. (a) The result of our local homography; (b) The result of traditional global homography.

C. Contour matching based on neighboring pixels

Contour matching is similar to region-based matching. First, we use the SOBEL operator to obtain the contour map of the image; then, we extract the branch nodes of contour as feature points, for each feature point, we take its 10*10 neighboring pixels as feature descriptors; finally, the normalized cross correlation (NCC) is used to measure similarity, the pairs with the highest similarity are selected as the final result.

IV. STRUCTURE-PRESERVING WARPING

In this section, we create initial mesh on all images and define objective function with the mesh as the independent variable, by minimizing the objective function to obtain the optimal mesh. In accordance with the corresponding relationship between the initial mesh and the optimal mesh, the image deformation can be performed.

A. Definition of objective function

We propose a structure-preserving warping, which is formulated as an optimization problem of energy function including several constraint terms.

1) Alignment term: The alignment term constraints each feature points pair to be transformed to the same location to ensure the alignment of overlapping regions between images. Each feature point is expressed as form of a bilinear interpolation of the vertex coordinates, because the mesh is finally affine transformed, it can be derived that the form of bilinear interpolation is unchanged. The alignment term can be calculated as follows:

$$E_A = \sum \frac{1}{N}\left\|p_i^* - p_j^*\right\|^2 = \sum \frac{1}{N}\left\|\sum w_{i,k}V_{i,k} - \sum w_{j,k}V_{j,k}\right\|^2 \quad (1)$$

where N is the number of feature points, P_i*, P_j* is the transformed positions of P_i and P_j.

2) Regular term: The regular term constraints adjacent mesh vertices to do similar transformation, according to the characteristics of similar transformation, it can be derived that the position relationship between adjacent mesh vertices is unchanged.

As shown in Fig.2, we use the coordinate system constructed by V_b and V_c to calculate the position of V_a, since the location relationship remains unchanged after transformation, the error value of a single triangle can be expressed as:

$$C_{tri} = \left\|V_a^* - (V_b^* + u(V_c^* - V_b^*) + vR_{90}(V_c^* - V_b^*))\right\|^2 \quad (2)$$

The final regular term can be expressed as follows:

$$E_R = \sum_{i=1}^{NI}\sum_{j=1}^{NT} Ctri^{\,i,j} \quad (3)$$

3) Scale-preserving term: The scale-preserving term constraints no large scale changes after transform, however, simply requires that all images are close to the original size is illogical, we hope that the image with rich feature information is relatively large, vice versa.

First, we calculate the relative scale factor between matched images. Specifically, for a image pair (I_i,I_j), we build a convex polygon Pi on the feature points from I_i and find its corresponding polygon P_j on image I_j. The relative scaling factor is defined as the perimeter ratio of P_i and P_j.

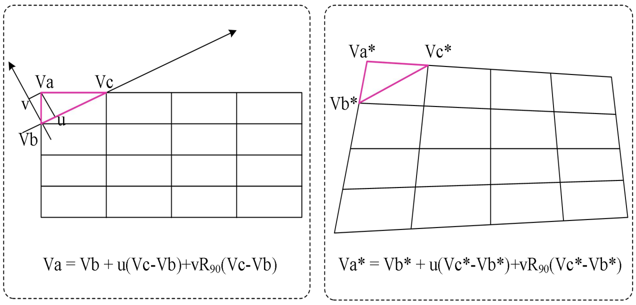

Va
Vc
v
u
Vb
Va = Vb + u(Vc-Vb)+vR90(Vc-Vb)
Va*
Vc*
Vb*
Va* = Vb* + u(Vc*-Vb*)+vR90(Vc*-Vb*)

Figure 2. Regular term.

Then, we calculate the absolute scale factor for each image, and solve the problem by minimizing the global scale change and maintaining the relative scaling relationship between the matched images:

$$\arg\min_{s} \sum \left|\gamma_{ij} s_j - s_i\right|^2 \quad s.t. \sum s_i = N_I \qquad (4)$$

Finally, the error value can be expressed as the sum of the deviation between the ideal scale and the true scale of each image:

$$E_S = \sum \left\| S(I_i^*) - s_i S(I_i) \right\|^2 \qquad (5)$$

where $S(I_i)$ is the scale vector of the original image I_i, and $S(I_i^*)$ is the scale vector of the warped image I_i^*.

4) Line-preserving term: The line-preserving term constraints that the lines in the image remain linear after transformation. In section 3.2, we have detected all the lines in the image, first we sample some discrete points on the line as $\{P_1, P_2, ..., P_n\}$, the coordinates of each point can be calculated using bilinear interpolation, for the convenience of later optimization, it is required here that sampling points are in different grid. Then, the orthogonal vector $[a_l, b_l]\perp$ is calculated according to P_1^* and P_n^*. Finally, the line-preserving term can be expressed as sum of the deviation of all sub line segment and line :

$$E_L = \sum \sum_{i=1}^{n-1} \left([a_l, b_l]_\perp \bullet \left(\sum w_{i,k} V_{i,k} - \sum w_{i+1,k} V_{i+1,k} \right) \right) \quad (6)$$

5) Contour-preserving term: The contour-preserving term constraints that significant structures in the image do not distort much after transformation, similar to of regular term, we construct a series of triangles on image contour, contour-preserving term can be expressed as the sum error of all triangles. As shown in Fig.3, the contour-preserving term can be calculated as follows:

$$E_C = \sum \left\| V_{key}^* - (V_b^* + u(V_c^* - V_b^*) + vR_{90}(V_c^* - V_b^*)) \right\|^2 \qquad (7)$$

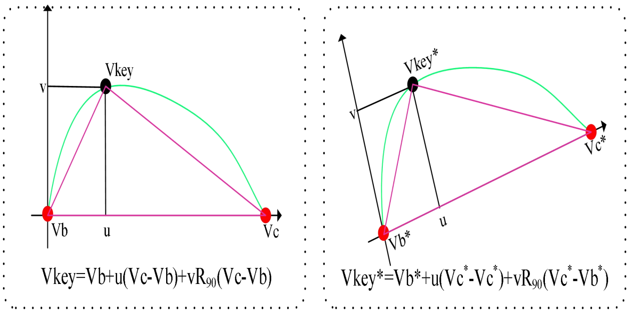

Vkey
v
Vb
u
Vc
Vkey=Vb+u(Vc-Vb)+vR90(Vc-Vb)
Vkey*
v
Vb*
u
Vc*
Vkey*=Vb*+u(Vc*-Vc*)+vR90(Vc*-Vb*)

Figure 3. Contour-preserving term.

B. *Optimization of objective function*

We combine the above five constraint terms into the following energy minimization problem:

$$E = E_A + \lambda_R E_R + \lambda_S E_S + \lambda_L E_L + \lambda_C E_C \quad (8)$$

where λ_R, λ_S, λ_L, λ_C are the weights of each term, in our experiments, they are all set to 1. The above minimization problem is mostly quadratic and is solved by linear approximation and sparse linear solver. Once we obtain the optimal mesh, in accordance with the corresponding relationship between the initial mesh and the optimal mesh, the image deformation can be performed.

V. Seamless Blending

After warping images, we need to blend warped images. The blending strategy we employ is multi-band blending [23] based on the optimal seam. To ensure the quality of the optimal seam under different cases, we synthetically consider alignment error, color difference and saliency difference to define the difference map.

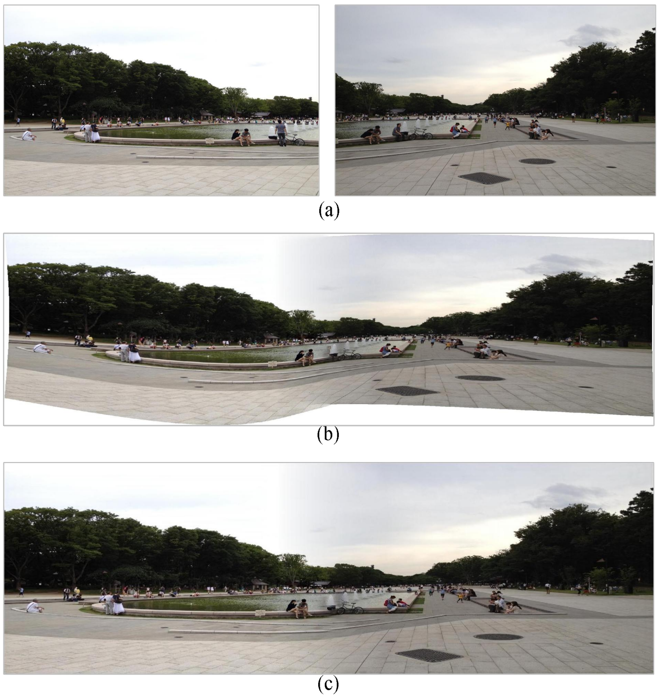
1) Alignment error : Just like [5], given the image pair (I_i, I_j), we compute the alignment errors for each image:

$$S_{I_i}(x) = \frac{\sum_p \omega_{p,x} s_{p,q}}{\sum_p \omega_{p,x}} \qquad (9)$$

where (p,q) is a pair of matched feature points, respectively in I_i and I_j. $s_{p,q}$ is the alignment error of (p,q), $\omega_{p,x}$ is the weight coefficient. ψ_i and ψ_j are transform functions of I_i and I_j. The final alignment error value is:

$$S_{align} = \frac{1}{2}\left(\psi_i(S_{Ii}) + \psi_j(S_{Ij})\right) \qquad (10)$$

2) Color difference : For the image pair (I_i, I_j), the color difference is expressed as the Euclidean distance of of pixel's RGB in the overlapping region , and the value is



(a)

(b)

(c)

Figure 4. Panoramas of different object functions. (a) Two input images. (b) The panorama of object function $E=E_A+E_R+E_S$; (b) The panorama of object function $E=E_A+E_R+E_S+E_L+E_C$.

mapped to the [0,1] interval via the Gauss function:

$$S_{color} = exp(-\frac{\left|\psi_i(I_i) - \psi_j(I_j) - \mu\right|^2}{\sigma^2}) \qquad (11)$$

where ψ_i and ψ_j are transform functions of I_i and I_j. μ is the mean value of pixel distance in overlapping region, while σ is the standard deviation.

3) Saliency difference : We take the contour in the image as the saliency features, and ensure that ideal seam does not cross the contour, avoiding obvious structural fracture in the final panorama.

4) Specific combination : Firstly, we combine the alignment error with the color difference and constrain the numerical to [0,1]:

$$S_{total} = \frac{(S_{align} + S_{color}) - min}{max - min} \qquad (12)$$

Secondly, we extract the contour and extend the contour to both sides to generate image mask.

Then, we apply the mask to the difference map, that is, only the difference value of the contour is preserved, and the difference value of the other region is set to 0.

Finally, we use the graph-cut algorithm to solve the optimal seam on the basis of the difference map.

VI. Experiment evaluation

We experiment the proposed approach both on common data sets and a variety of challenging images captured from surveillance cameras in urban scenes.

A. *Experiments on common data sets*

First we verify whether our feature matching method can obtain rich and reliable matching information. Line matching and contour matching are introduced to make up for the insufficient of point matching. As shown in Fig.1, our local homography model can provide more matching than traditional global homography model.

Then, we verify whether the structure-preserving model can achieve enough ideal reference plane and image deformation effect. We do not set one of the images as reference plane, but add mesh for all images, and solve the optimal reference plane by calculating the optimal mesh.

Finally, we verify whether the blending algorithm can obtain the ideal seam under different alignment conditions. As shown in Fig.5, our seam crosses areas with good alignment and strictly converges to the contour of the image. We propose a ZNCC score for the final seam, and evaluate the applicability of the blending algorithm by monitoring the score of the seam under various data sets. After a large number of experimental results statistics, our blending algorithm can achieve satisfactory results in most cases.

An example of specific comparison is shown in Fig.7.

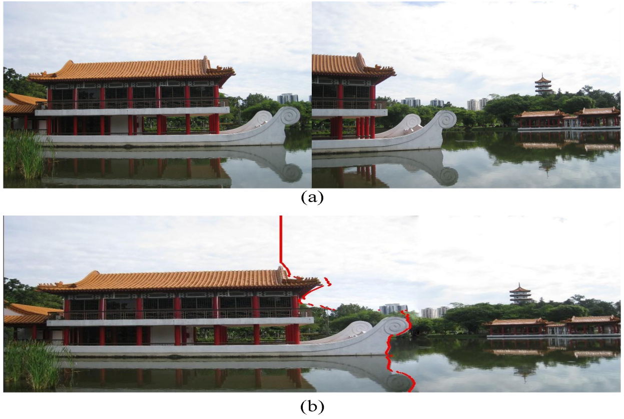
B. *Experiments on surveillance scenes*

After experiments above, we can argue that our stitching algorithm is effective on common data sets. Next, we need to compare with other advanced algorithms to verify the performance of our method. We compare with AutoStitch [25], ICE, SPHP [7], WB [5] and SEAGULL [6] in challenging surveillance scenes.

As shown in Fig.6, for AutoStitch, the first image cannot be stitched together and there is obvious distortion in the red box as the high building is missing directly. For ICE, the first image cannot be stitched either, and obvious repetition appears in two red boxes because of the difficulty of alignment with large parallax. For SPHP , the whole panorama is blurring and distorted, because the SPHP is composed of interpolation from homography transform in overlapping regions to similarity transformation in non-overlapping regions, although there is a certain degree of shape correction, but there is no optimization of alignment in overlapping regions. For WB, structure fracture appears in the red box, this is because the WB only considers alignment error and color difference to find the seam, the solved seam cross the misaligned structure, causing significant structural fracture. For SEAGULL, obvious dislocation appears in the red box, this is because the SEAGULL only considers the pixel information of the image contour and its surrounding to seek the seam, but no pixel information far away from the contour, the solved seam is not the global optimal. Finally for our method, It can be seen that there is no obvious distortion or artifacts in comparison with the above algorithms.

VII. Conclusion

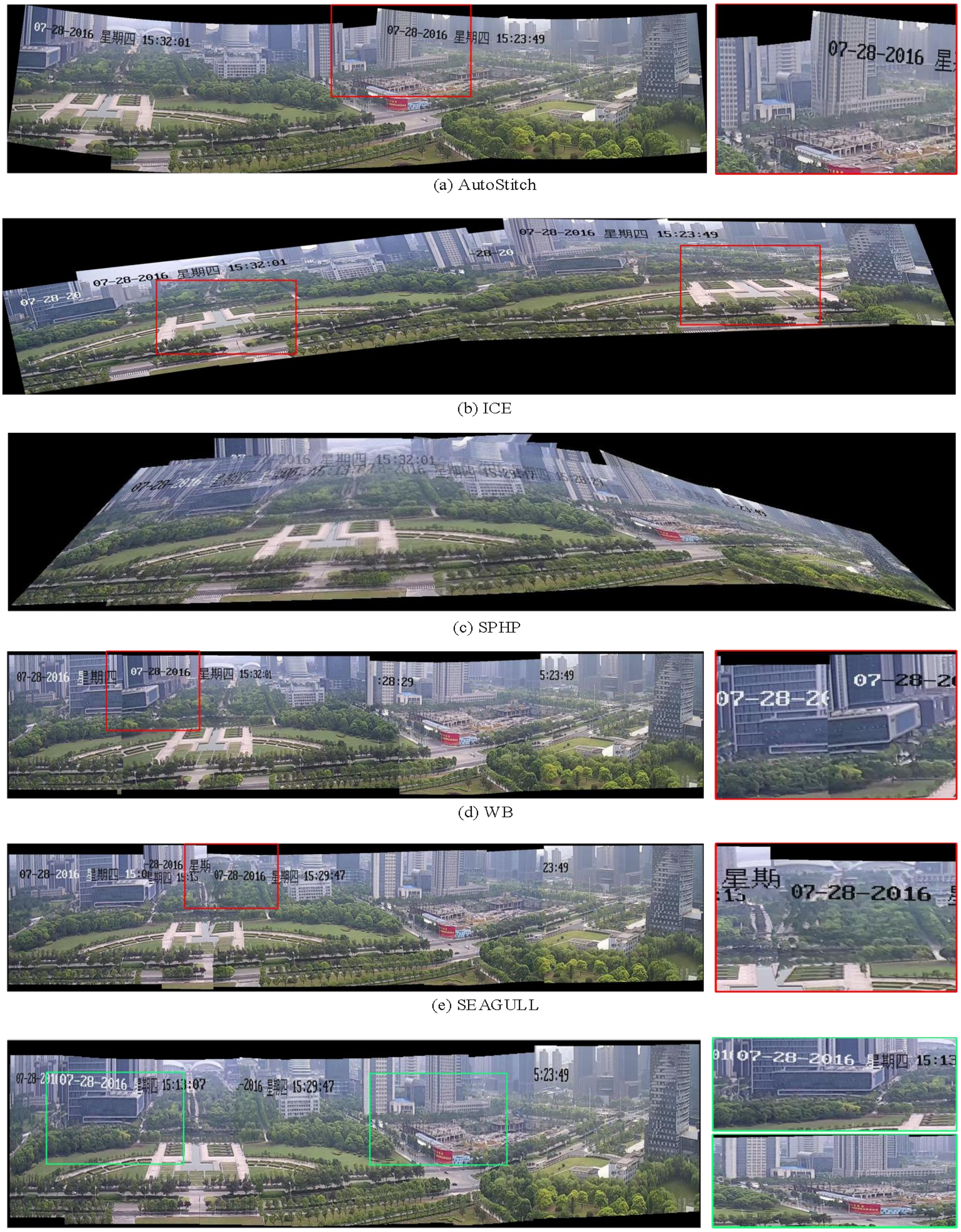
This paper presented a novel stitching method for wide-baseline images with low texture on the basis of combining several state-of-the-art stitching algorithms. The three-phase feature matching method effectively compensates for the deficiencies of the traditional matching based on global homography, even in the case of low texture, we can still obtain rich and reliable feature matching. The structure-preserving warping model effectively balances the projective distortion and the perspective distortion, and the optimal reference plane can keeps the scale relationship between images and the visual habit of people. In image blending, we synthetically consider alignment error, color difference and saliency information to find optimal seam, the seam try pass regions with good alignment and similar color, and not break the internal structure. Compared with several existing top stitching algorithms, it can be proved that our approach has outstanding performance.



(a)

(b)

Figure 5. The optimal seam. (a) Two input images. (b) The panorama with the red optimal seam.


07-28-2016 星期四 15:32:01
07-28-2016 星期四 15:23:49
07-28-2016 星期四 15:29:47
5:23:49

(a) AutoStitch

(b) ICE

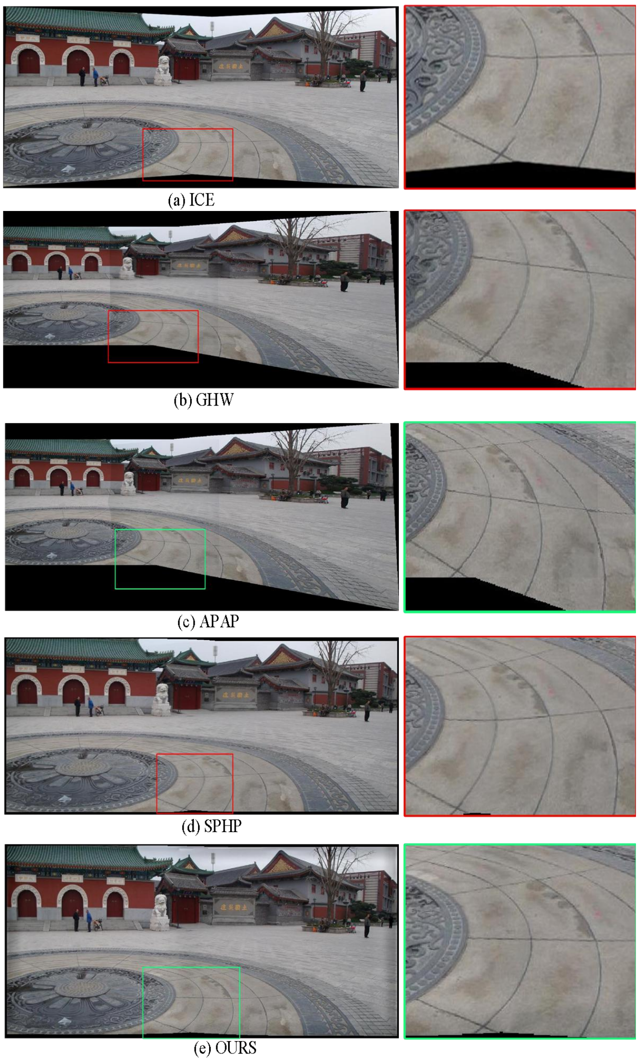
(c) SPHP

(d) WB

(e) SEAGULL

(f) OURS

Figure 6. Comparisons among various methods on surveillance scenes with 6 input images.(a) The result of AutoStitch with seamless blending, the image 1 cannot be stitched together; (b) The result of Microsoft ICE with seamless blending, the image 1 cannot be stitched together; (c) The result of SPHP without seamless blending; (d) The result of WB without seamless blending; (e) The result of SEAGULL without seamless blending; (f) The result of our method without seamless blending.



(a) ICE

(b) GHW

(c) APAP

(d) SPHP

(e) OURS

Figure 7. Comparisons among various methods on common data sets. (a) The result of ICE, the line in the red box is obviously broken. (b) The result of GHW, there is obvious artifact in the red box. (c) The result of APAP, the content of the green box is fairly well maintained. (d) The result of SPHP, the line in the red box is curved. (e) The result of our method, the content of the green box maintains fairly well.

ACKNOWLEDGE

This work is supported by the National 863 Program of China under Grant No.2015AA016403 and the Natural Science Foundation of China under Grant No.61572061, 61472020.

REFERENCES

[1] R. Szeliski. Image alignment and stitching:a tutorial. Found.Trends. Comput.Graph. Vis., 2(1):1–104, 2006.

[2] J. Gao, S. J. Kim,and M. S. Brown, Constructing image panoramas using dual-homography warping. In IEEE CVPR, pages 49–56, 2011.

[3] W.-Y. Lin, S. Liu, Y.Matsushita, T.-T. Ng, and L.-F. Cheong, Smoothly varying affine stitching. In IEEE CVPR, pages 345–352, 2011.

[4] J. Zaragoza, T.-J. Chin, M. S. Brown, and D. Suter. As-projective-as-possible Image stitching with moving DLT. In IEEE CVPR, 2013.

[5] G.-F. Zhang and Y. He, "Multi-Viewpoint Panorama Construction With Wide-BaseLine Images," IEEE.Trans,Image Processing, vol. 25, no. 7, jul. 2016, pp. 3099-3111

[6] K.-M. Lin,N.-J Jiang, "SEAGULL:Seam-guided Local Alignment for Parallax-tolerant Image Stitching," in ECCV,2016

[7] C.-H. Chang, Y. Sato, and Y.-Y. Chuang, "Shape-preserving-half-projective warps For image stitching," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014,pp.3254–3261.

[8] C. C. Lin, S. U. Pankanti, K. N. Ramamurthy, and A. Y. Aravkin, "Adaptive-as-natural-as-possible image stitching," in CVPR, 2015, pp.1155-1163.

[9] Y.-S. Chen and Y.-Y. Chuang, "Natural image stitching with the global similarity prior," in ECCV, 2016, pp. 186-201.

[10] Y. Guo, F. Liu, J. Shi, Z.-H. Zhou, and M. Gleicher, "Image retargeting using Mesh parametrization," IEEE Trans. Multimedia, vol. 11, no. 5, pp. 856–867, 2009.

[11] W. Hu, Z. Luo, and X. Fan, "Image retargeting via adaptive scaling with Geometry preservation," IEEE J. Emerg. Sel. Topics Circuits Syst.,vol. 4, no. 1, pp.70–81, Mar. 2014.

[12] G.-X. Zhang, M.-M. Cheng, S.-M. Hu, and R. R. Martin, "A shape-preserving approach to image resizing," Comput. Graph. Forum, vol. 28, no. 7, pp. 1897–1906,2009.

[13] C.-H. Chang and Y.-Y. Chuang, "A line-structure-preserving approach to image resizing," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,Jun. 2012, pp.1075–1082.

[14] K. He, H. Chang, and J. Sun, "Rectangling panoramic images via warping,"ACM Trans. Graph., vol. 32, no. 4, pp. 79:1–79:10, Jul. 2013.

[15] F. Liu, M. Gleicher, H. Jin, and A. Agarwala. Content preserving warps for 3d Video stabilization. ACM Transactions on Graphics, 28(3):44:1–44:9, 2009.

[16] S. Liu, L. Yuan, P. Tan, and J. Sun, "Bundled camera paths for video stabilization," ACM Trans. Graph., vol. 32, no. 4, p. 78, 2013.

[17] D. G. Lowe. Distinctive image features from scale-invariant keypoints. Int. J.Comput. Vision, 60(2):91–110, 2004.

[18] G. Yu and J.-M. Morel, "ASIFT: An algorithm for fully affine invariant comparison," Image Process. On Line, vol. 1, 2011.

[19] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: Slic Superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach.Intell. 34(11), 2274-2282 (Nov 2012)

[20] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: Slic Superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach.Intell. 34(11), 2274-2282 (Nov 2012)

[21] Q. Jia and X.-K. Gao, "Novel Coplanar Line-points Invariants for Robust Line Matching Across Views," in ECCV, 2016, pp. 599-611

[22] R. G. von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "LSD:A fast line segment detector with a false detection control," IEEE Trans.Pattern Anal. Mach.Intell., vol. 32, no. 4, pp. 722–732, Apr. 2010.

[23] P. J. Burt and E. H. Adelson. A multi-resolution spline with application to image mosaics. ACM Transactions on Graphics, 2(4):217–236, 1983.

[24] V. Kwatra, A. Schödl, I. Essa, G. Turk, and A. Bobick, "Graphcut textures: Image and video synthesis using graph cuts," ACM Trans.Graph., vol. 22, no. 3, pp. 277–286, 2003.

[25] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. Int. J. Comput. Vision, 74(1):59–73, 2007.