2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Locally Warping-based Image Stitching by Imposing Line Constraints

Tianzhu Xiang, Gui-Song Xia, Liangpei Zhang
State Key Lab. LIESMARS, Wuhan University
Wuhan 430079, China
{*tzxiang, guisong.xia, zlp62*}@whu.edu.cn

Ningning Huang
Intelligence Detachment
Nanning Bureau of Public Security
Nanning 530028, China

***Abstract*—Warping-based image stitching methods often suffer from perspective variations among multiple images and lead to shape and perspective distortions in stitching results. Moreover, they also quickly lose their efficiency in low-textured images, due to the lack of reliable point correspondences. To solve these problems, this paper presents a locally warping-based image stitching by imposing line constraints. First, a two-stage alignment scheme with line constraints is introduced to achieve accurate alignment. More precisely, line features are adopted as alignment constraints to jointly estimate local homographies with point correspondences, which provides strong correspondences especially in low-textured cases. Then line constraints are also imposed to the content-preserving warping framework to further reduce alignment errors and preserve image structures. Second, in order to preserve shape and perspective information, a global similarity transform is introduced to mitigate projective distortions. Experimental results demonstrate the efficiency of our method, which yields more encouraging image stitching results in contrast with state-of-the-art methods.**

I. Introduction

Image stitching, a process of warping a set of images into a single while larger mosaic with a wider field of view (FoV) [1], has been widely used in many tasks, such as panorama [2], scene understanding [3], photogrammetry [4], remote sensing [5] and many others.

A. Motivation and Objective

Recently warping-based methods are proposed to improve image stitching performance. These methods can be roughly classified into two categories: multiple homographies [6]–[9] and mesh-based warping [10]–[13]. In contrast with global models [1], [14], these methods are more flexible to handle parallax due to higher degree of freedom (DoFs). However, there are still some problems, as follows.

- Most warping-based methods suffer from projective distortions, especially shape distortion and perspective distortion. This is because these methods only focus on alignment accuracy without considering distortions resulted by projective transformation. As shown in Fig. 1, the as-projective-as-possible (APAP) warping method [7] and content-preserving warping (CPW) method [10] suffer from some projective distortions, for example, the chair in red box is obviously stretched, the perspective of the stitched image faces unsatisfactory distortions.

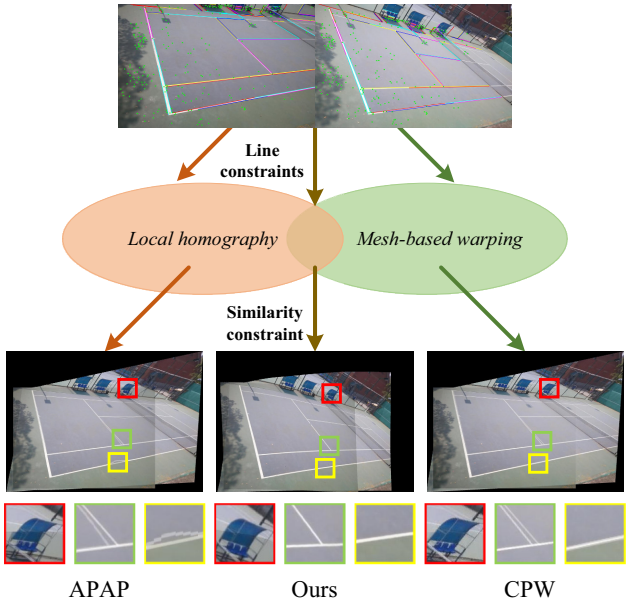


Fig. 1. Illustration of the overview of our method. The point and line features are detected and matched. The line feature is first adopted as another alignment constraint to jointly estimate local homographies with points. Furthermore, line constraints are integrated into mesh-based warping to improve alignment and preserve image structures. Similarity transformation is introduced to alleviate projective distortions. Observe that APAP [7] (one of the local homography-based methods) and CPW [10] (one of the mesh-based warping methods) produce misalignments and suffer from projective distortions. The distortions of APAP and CPW can be seen in red boxes. The misalignments of APAP and CPW can be seen in green and yellow boxes.

- These methods may fail to provide accurate alignment in low-texture images. In low-texture or homogeneous regions, such as white walls in indoor images, sky regions in outdoor images, low-textured regions of man-made structures, the key points are difficult to be extracted. It is unlikely to provide sufficient and reliable correspondences of local geometries [15] to estimate an accuracy model for image alignment. This may also damage the structures where local transformations are erroneously estimated. In Fig. 1, owing to the insufficient corresponding information, the white lines on ground in green box in APAP and CPW are obviously misaligned, the straight line in yellow box in APAP is bended.

978-1-5090-4846-5/16/$31.00 ©2016 IEEE

From literature, some studies have been devoted to dealing with the first problem, e.g. [8] and [9], among which similarity transformation shows many advantages and promising results in the reduction of projective distortions. But they are incapable of handling low-texture images. Meantime, as far as we know, few methods can handle the latter problems which mainly resulted by the poor correspondences of low-texture region. Recently, [16] and [17] introduced the line feature to improve the alignment accuracy for image stitching. However, they also suffer from projective distortions and may be weak to handle parallax.

To deal with these problems discussed above, the locally warping-based image stitching by imposing line constraints is proposed in this paper. The line feature can not only be regarded as another alignment constraint to provide rich correspondences for accurate warping model estimation, but also be employed as line constraints, such as line colinearity or correspondence, to further improve alignment and preserve image structures. In addition, similarity transformation is adopted to reduce projective distortions in non-overlapping regions by weight integration with local warping model. The outline of our method is illustrated in Fig. 1.

B. Contributions

Our work is distinguished by the follow contributions.

- We propose a two-stage alignment scheme with line constraints for image stitching. We first adopt the line features as the alignment constraint to jointly estimate the local homographies.Then with line correspondence constraint, the mesh-based warping is employed to further refine the alignment. This method combines the strength of local homographies and mesh-based warping.
- We introduce line features as structure constraints to preserve the image structures. The line correspondence and line colinearity are integrated into mesh-based warping framework to prevent the structural distortions.

The remainder of paper is organized as follows. Section II describes the details of our approach. In section III the experimental results are shown and discussed. Finally, we conclude our work in Section IV.

II. Our Method

A. Local Homography Estimation with Line Feature

Compared with other image stitching methods, the line feature is adopted as another alignment constraint to jointly estimate local homographies with point correspondences in our method, which provides rich correspondences in low-texture images. LSD [18] is employed for line detection. Line-point invariants (LPI) matching [19] is adopted for lines matching [20] [21].

Given a pair of matching points $p_i = [x_i, y_i, 1]^T$ and $p_i^{'} = [x_i^{'}, y_i^{'}, 1]^T$ between image I and $I^{'}$ respectively, the homography H for p_i to $p_i^{'}$ can be expressed as: $p_i^{'} = Hp_i$, and H can be estimated by DLT from a group of matching points. The equation can be rewrote as $A_{p_i}h = p_i \times Hp_i^{'} = 0_{3\times1}$, h is a vector representation of H. So:

$$A_{p_i} = \begin{bmatrix} x_i, y_i, 1, 0, 0, 0, -x_i^{'}x_i, -x_i^{'}y_i, -x_i^{'} \\ 0, 0, 0, x_i, y_i, 1, -y_i^{'}x_i, -y_i^{'}y_i, -y_i^{'} \end{bmatrix} \tag{1}$$

Given a pair of corresponding lines $(l_j, l_j^{'})$, they can be parameterized as $l_j = [a_j, b_j, c_j]^T$ and $l_j^{'} = [a_j^{'}, b_j^{'}, c_j^{'}]^T$. Besides, $p_j^{0,1} = [x_j^{0,1}, y_j^{0,1}, 1]^T$ denotes its two endpoints of l_j. The transformed endpoints of line l_j should lie on the target line $l_j^{'}$ for line-to-line mapping, that is: $l_j^{'T} Hp_j^{0,1} = 0$. From $A_{l_j}h = l_j^{'T} Hp_j^{0,1} = 0$, we can get

$$A_{l_j} = \kappa_j \begin{bmatrix} a_j^{'}p_j^0, & b_j^{'}p_j^0, & c_j^{'}p_j^0 \\ a_j^{'}p_j^1, & b_j^{'}p_j^1, & c_j^{'}p_j^1 \end{bmatrix} \tag{2}$$

where κ_j is a scalar factor for the balance with A_{p_i}, $\kappa_j = 1/\sqrt{a_j^{'2} + b_j^{'2}}$.

Simply stacking the coefficient matrices of points A_{p_i} and lines A_{l_j}, i.e. $A = [A_{p(:)}; A_{l(:)}]$, $A \in \mathbb{R}^{2(N+M)*9}$, N is the number of point correspondences and M is the number of line correspondences. For numerical stability, the all entries of the stacked matrix $[A_{p(:)}; A_{l(:)}]$ should be normalized.The point-centric normalization approach proposed in [22] is adopted.

Inspired by APAP, the local homographies can be jointly estimated by points and lines. The input image is first partitioned into meshes, and the local homography of each mesh is estimated by:

$$h_* = \arg\min_h \|WAh\|^2, \text{ subject to } \|h\| = 1 \tag{3}$$

where $W = diag\left(\left[w^p, w^l\right]\right)$ denotes weight factor of point and line correspondences, respectively.

The weight factor of point w^p is computed with Gaussian kernel defined in [7]. The line weight factor w^l is computed as:

$$w^{l_j} = \max\left(\exp\left(-d_l(X_*, l_j)^2/\sigma^2\right), \eta\right) \tag{4}$$

where $d_l(X_*, l_j)$ is the shortest distance between X_* and l_j.

$$d_l(X_*, l_j) = \begin{cases} \min(\|X_* - X_1\|, \|X_* - X_2\|) & (a) \\ |a_j x_* + b_j y_* + c| / \sqrt{a_j^2 + b_j^2} & (b) \end{cases} \tag{5}$$

where X_1, X_2 are the endpoints of l_j. As shown in Fig. 2, if X_* is in the R_1 or R_2 region, the d_l is calculated by (a), and if X_* is in the R_3 region, d_l is calculated by (b).

From (4) and (5), the weight is high when grid point is closer to point or line correspondences, and nearly equal when grid point is far from the matching points and lines.

B. Global Similarity Transformation Constraint

To mitigate the projective distortions, the global similarity transformation [8] is employed as constraint to tweak the local homographies to reduce distortions and preserve the image shape and perspective.

The global similarity transformation is estimated by RANSAC iteratively on matching points to find a group of

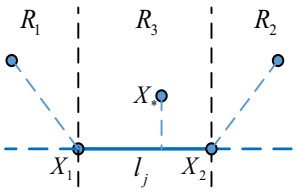


Fig. 2. The shortest distance between the cell point X_* and line segment l_j.

points with the smallest rotation angle, see [9] for details. Then it is combined with local homographies as below:

$$H_i^{'} = \alpha H_i + \beta S \tag{6}$$

where H_i is the local homography in the i^{th} grid, $H_i^{'}$ is the final local warping, S is the similarity transformation, α and β are weight coefficients and $\alpha + \beta = 1$.

Accordingly, the reference image should take the corresponding warping so that the two input images can be well aligned. The warping is:

$$T_i^{'} = H_i^{'} H_i^{-1} \tag{7}$$

where $T_i^{'}$ is the local warping for reference image in i^{th} grid.

Due to space limitations, the estimation of weight coefficients can refer to [9], [23].

C. Line Feature Constraints

Line feature constraints are adopted to extend the content-preserving warping [10] framework so that it can further align the pre-warping result as well as preserve image structures. Line feature constraints mainly contain line alignment constraint and line collinearity constraint detailed below.

The input image I is first divided into $m \times n$ regular grid, and the grid can be transformed to the pre-warping image by local warping. The vertices of the mesh in pre-warping image for solving denote V. For arbitrary point P_i in pre-warping image, it can be represented by a linear combination of four grid vertices $V_i = [V_i^1, V_i^2, V_i^3, V_i^4]^T$ in its locating quad: $P_i = w_i^T V_i$, and $w_i = [w_i^1, w_i^2, w_i^3, w_i^4]^T$ is calculated by the inverse bilinear interpolation and sum to 1. The image warping problem is then changed to a mesh warping problem. The energy terms in this paper is detailed below.

1) Line alignment term: The line correspondences should be aligned. Assuming that l_j, $l_j^{'}$ are a pair of line correspondence. Line l_j is cut into several short line segments by the edges of mesh if l_j goes across this mesh. $\{P_k\}$ denotes all endpoints of short line segments from l_j. $\{P_k^{'}\}$ denotes the endpoints in the pre-warping image transformed from $\{P_k\}$ by local warping. The distance from all transformed endpoints $\{P_k^{'}\}$ to the corresponding line $l_j^{'}$ should be minimized. So

$$E_l(V) = \sum\nolimits_{j,k} \left\| (l_j^{'\,T} \cdot w_{j,k}^T V_{j,k}) / \sqrt{a_j^{'\,2} + b_j^{'\,2}} \right\|^2 \tag{8}$$

Lines are used to not only enhance the alignment for lines but also preserve the straightness property of lines together with the line collinearity term below.

2) Line collinearity term: It is used to preserve the line structures as much as possible. Line $\hat{l_j}$ is computed by the head and tail endpoints of $\{P_k^{'}\}$. The term is expressed by the distance from the transformed endpoints $\{P_k^{'}\}$ to the line $\hat{l_j}$, that is

$$E_c = \sum\nolimits_{j,k} \left\| (\hat{l_j}^T \cdot w_{j,k}^T V_{j,k}) / \sqrt{\hat{a}_j^2 + \hat{b}_j^2} \right\|^2 \tag{9}$$

3) Other terms: Point alignment term, global alignment term and smoothness term are directly from [10]. See [10] for detailed information.

The above five energy terms are combined as energy minimization problem, the objective function is:

$$E = \alpha E_p + \beta E_l + \gamma E_c + \delta E_s + \rho E_g \tag{10}$$

where $\alpha, \beta, \gamma, \delta, \rho$ are weight factors of each term. In our paper, $\alpha = 1, \beta = 1, \gamma = 0.001, \delta = 0.01, \rho = 0.001$. Since the above function is quadratic, the function can be solved by sparse linear solver. The final result is obtained through texture mapping.

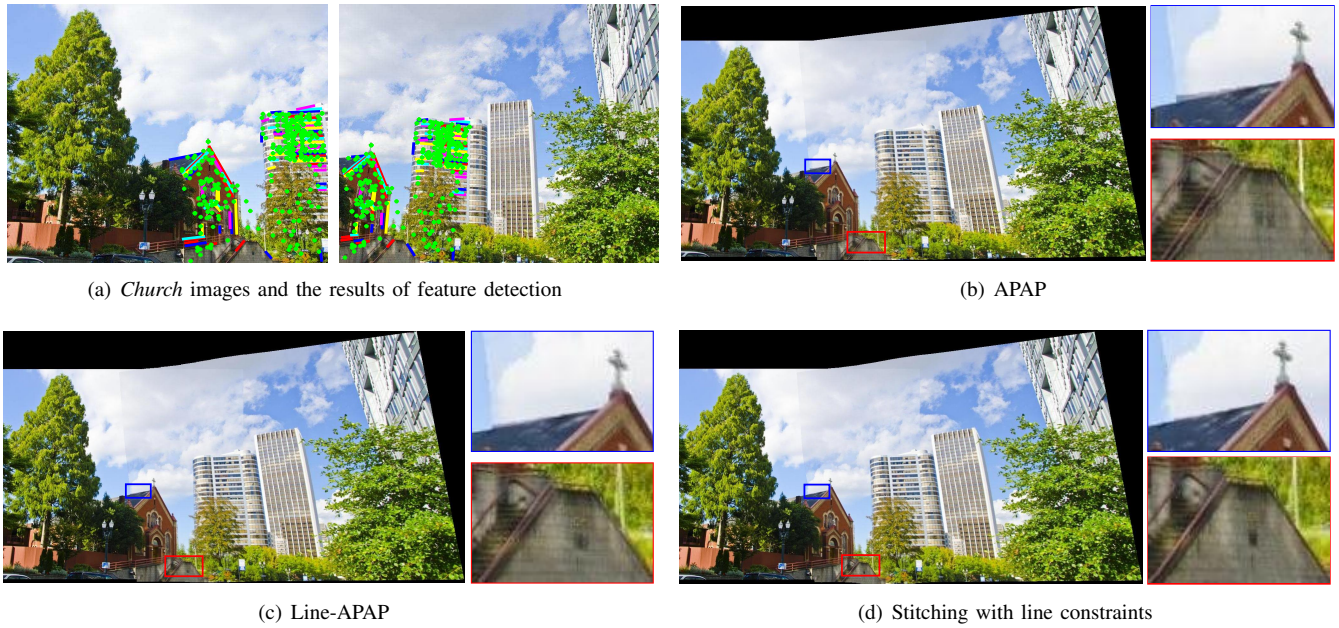
III. Results and Analysis

A set of experiments are conducted to verify the effectiveness of the proposed method on a range of challenging images. Our method is compared with other algorithms, including global method [14], APAP [7], SPHP with global homography [8], and SPHP with APAP (SPHP+APAP) [8].

For better comparison, the post-processing methods like blending or seam cutting detailed in [1] are avoided in case of interference. The stitching results are simply blended by intensity average so that misalignments remain obvious. The parameters for other methods are set as the same suggested in each paper. The parameters in our paper are: $\sigma = 8.5, \eta = 0.01$. To quantify the alignment accuracy of our method, the root mean squared error (RMSE) suggested in [7] on point and line matches is adopted.

Fig. 3 illustrates the performance of line constraints on the image stitching. As can be seen in Fig. 3 (b), misalignments are obvious at the roof of church and the handrail, owing to the lack of point correspondences, which can be seen in the results of feature detection in Fig. 3 (a). With the line correspondences that provide strong alignment constraints, there obtains a better result. But the handrail still suffers from the slight misaligments. Thanks to the integration of line feature constraints into content-preserving warping, there is a clear improvement in alignment. Our method achieves an encouraging image stitching performance. Table I depicts the quantitative evaluation on the *Church* experiment. It is clear that our approach provides the smallest errors.

Fig. 4 illustrates the original images for compared experiments: *Building, Temple, Desk* and *Wall* images. Experimental images include two categories: (a) outdoor images from public available dataset [7], [12] and (b) indoor images captured by ourselves. Each image pair is taken at different view points, so they may be not fitted by global models. Furthermore, *Building, Desk* and *Wall* images are lack of textures, which bring



(a) *Church* images and the results of feature detection

(b) APAP

(c) Line-APAP

(d) Stitching with line constraints

Fig. 3. The performance of line constraints on image stitching. (a) Original images of *church* and feature detection. (b) Stitching by APAP (one of the keypoint-based approach). (c) Stitching by Line-APAP. The local homography is estimated jointly by line and point correspondences, i.e. the first-stage alignment. (d) stitching by our method, i.e. the two-stage alignment. Other than alignment constraint, content-preserving warping framework with line feature constraints is adopted. For better comparison, some local details are highlighted.

TABLE I
RMSE ON *Church* IMAGES

Method	APAP	Line-APAP	Our method
RMSE (Point)	1.69	1.48	**0.87**
RMSE (Line)	2.69	2.04	**1.56**
RMSE	2.18	1.76	**1.21**

about some challenges for image stitching. Fig. 5 shows the stitching results on the four pairs of images. Some details are highlighted for better comparison. Red boxes show alignment errors or distortions, and green boxes show the satisfactory stitching. The enlarged views are shown below each stitching result.

In Fig. 5, it is obvious that the global method is hard to tackle parallax well for the defective model. It produces obvious misalignments, which can be seen in four pairs of the stitched results. In addition, because of the projective transformation adopted by the global method, it inevitably suffers from the projective distortions. This can be seen in *Building* and *Temple* images, the tall buildings in non-overlapping regions in *Building* are enlarged and become tilt, the buildings in *Temple* are severely inhomogeneous stretched and become parallelograms. The perspective of *Temple* images is undesirely changed with serious incline.

Owing to higher DoF, APAP is more flexible to handle parallax than the global method, it can reduce much alignment errors in the overlapping regions. For *Building, Desk* and *Wall* images, the misalignments of APAP are less than the global method. For *Temple* images, we can see the striking performance of alignment. It achieves accurate alignment owing to the rich and exact point correspondences. However, because of the insufficient and unreliable correspondences in low-texture regions, it cannot completely handle alignment errors. Besides, due to the simple extrapolation of projective transformation to non-overlapping regions, APAP also produces projective distortions in the non-overlapping regions. This can be obviously discovered in *Temple* result. APAP may break the image structures when there are rare correspondences for the estimation of local homographies, especially line structures. For instance, some straight lines in *Building* and *Wall* results are bended, because of the lack of key point correspondences around these lines.

Combined with similarity transformation, SPHP and SPHP+APAP [8] are proposed to handle the shape distortion in non-overlapping regions. They adopt three continuous warping (homography, transition transformation, similarity transformation) from overlapping to non-overlapping regions to reduce the shape distortion in non-overlapping regions. This can be easily seen in the *Temple* results, the shape of distant buildings experience less distortion and is preserved well. However, due to the using of the global transformation, SPHP is not flexible to handle parallax and thus produces misalignments. The four pairs of stitched results suffer from alignment errors at different extent, which are shown in red boxes. Owing to the combination of SPHP and APAP, SPHP+APAP can improve the alignment performance. For instance, the images are aligned accurately in *Temple* result. But it is also unable to deal with misalignments in low-texture images. Because of the local estimation of APAP, it may also break the image structures. For instance, some lines in *Building* and *Wall* results

(a) Outdoor images (left: *Building*; right: *Temple*)

(b) Indoor images (left: *Desk*; right: *Wall*)

Fig. 4. The images for experiments.

TABLE II
QUANTITATIVE EVALUATION (RMSE) ON FOUR IMAGE PAIRS

Method	Global Method	APAP	SPHP	SPHP+APAP	Ours
Building	4.87	3.35	4.58	3.27	**1.81**
Temple	8.16	2.25	7.30	2.05	**0.74**
Desk	9.29	7.51	9.42	7.68	**1.52**
Wall	7.53	5.08	7.47	5.19	**1.95**

become curving. Besides, these two methods don't considering perspective distortion, so it remians in stitched images, which can be seen in *Temple* results.

Our method provides accurate alignment and preserves the image content and perspective. The method adopts multiple local homographies so that it is flexible enough to handle parallax. Due to the use of line features, our method can estimate an accurate warping model, so it provides the accuracy alignment in the overlapping regions on four pairs of images. For the line constraints, the image structures can be preserved, lines maintain straightness in the four esperiments. Owing to the integration of similarity transformation, it can significantly mitigate the projective distortions. The proposed approach preserves the image shape and perspective well. Table II shows RMSE of compared methods on four image pairs. As we can see, our methods consistently yields better accuracy than other methods. From the comparison, our method achieves an impressive image stitching results.

IV. CONCLUSION

This paper proposes a local warping-based image stitching method by imposing line constraints. Our method introduces the global similarity transformation to reduce the projective distortions, including shape and perspective distortions, by the way of combination with the locally warping model. Our method adopts the line features to handle the low-textured images. Line features are first used to provide rich and reliable correspondences for the accurate alignment, especially in low-texture cases that points are extracted with much difficulty. Then line features compose two line constraints (line collinearity and alignment constraints) integrated into content-preserving warping framework to improve alignment as well as preserve image structures. From experiments, an accurate distortion-free image stitching is achieved with our method.

REFERENCES

[1] R. Szeliski, "Image Alignment and Stitching: A Tutorial," in Handbook of Mathematical Models in Computer Vision, Springer, 2005, pp. 273-292.

[2] M. Brown and D. G. Lowe, "Recognising panoramas," in ICCV, Nice, France, 2003, pp. 1218-1225. F. Zhang and F. Liu, "Casual Stereoscopic Panorama Stitching," in CVPR, Boston, MA, USA, 2015, vol. 23, pp. 2002-2010.

[3] Y. Zhang, S. Song, P. Tan, and J. Xiao, "PanoContext: A Whole-Room 3D Context Model for Panoramic Scene Understanding," in ECCV, Zurich, Switzerland, 2014, vol. 8694, pp. 668-686.

[4] S. Pang, M. Sun, X. Hu, and Z. Zhang, "SGM-based seamline determination for urban orthophoto mosaicking," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 112, pp. 1-12, 2016.

[5] X. Li, N. Hui, H. Shen, Y. Fu, and L. Zhang, "A robust mosaicking procedure for high spatial resolution remote sensing images," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 109, pp. 108-125, 2015.

[6] J. Gao, S. J. Kim and M. S. Brown, "Constructing Image Panoramas using Dual-Homography Warping," in CVPR, Colorado Springs, CO, USA, 2011, pp. 49-56.

[7] J. Zaragoza, T. Chin, M. S. Brown, and D. Suter, "As-Projective-As-Possible Image Stitching with Moving DLT," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, pp. 1285-1298, 2014.

[8] C. Chang, Y. Sato and Y. Chuang, "Shape-Preserving Half-Projective Warps for Image Stitching," in CVPR, Columbus, OH, USA: IEEE, 2014, pp. 3254–3261.

[9] C. Lin, S. U. Pankanti, K. N. Ramamurthy, and A. Y. Aravkin, "Adaptive As-Natural-As-Possible Image Stitching," in CVPR, Boston, MA, USA, 2015, pp. 1155–1163.

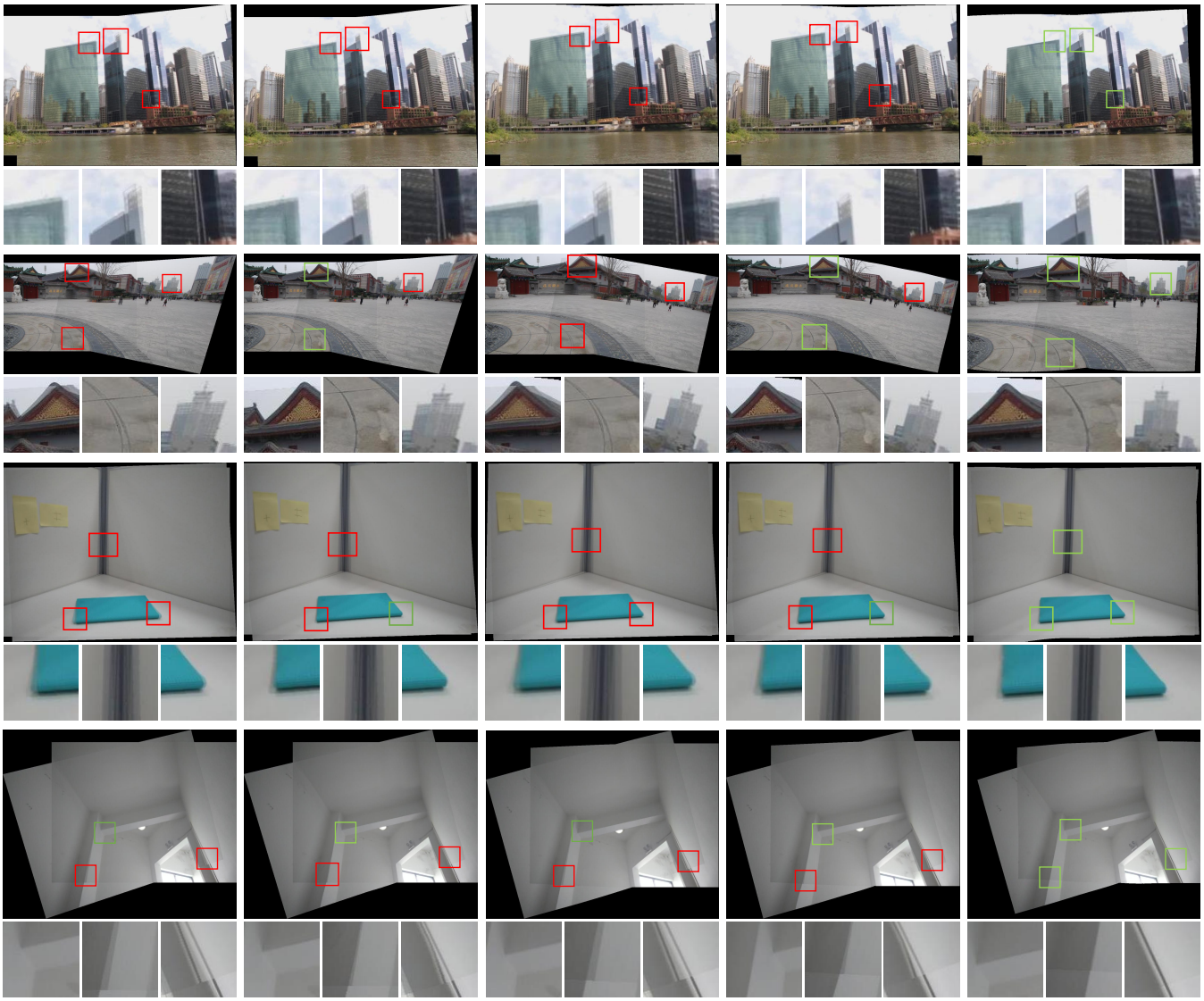


Fig. 5. Comparison of image stitching on four image pairs. From left to right, the results are: global method [14], APAP [7], SPHP [8], SPHP+APAP [8], and our method. For better comparison, some details are highlighted. Red boxes denote alignment errors or distortions, green boxes denote the satisfactory stitching. The enlarged views are displayed below each image stitching result.

[10] F. Liu, M. Gleicher, H. Jin, and A. Agarwala, "Content-preserving warps for 3D video stabilization-CPW," ACM Transactions on Graphics, vol. 28, p. 1, 2009-07-27 2009.

[11] W. Lin, S. Liu, Y. Matsushita, T. Ng, and L. Cheong, "Smoothly Varying Affine Stitching," in CVPR, Colorado Springs, CO, USA, 2011, pp. 345-352.

[12] F. Zhang and F. Liu, "Parallax-tolerant image stitching," in CVPR, Columbus, OH, USA: IEEE, 2014, pp. 3262–3269.

[13] C. Chang, C. Chen and Y. Chuang, "Spatially-Varying Image Warps for Scene Alignment," in ICPR, Stockholm, Sweden, 2014, pp. 64-69.

[14] M. Brown and D. G. Lowe, "Automatic Panoramic Image Stitching using Invariant Features," International Journal of Computer Vision, vol. 74, pp. 59-73, 2007.

[15] G.-S. Xia, J. Delon, Y. Gousseau. "Accurate junction detection and characterization in natural images," International Journal of Computer Vision, vol.106, No.1, pp: 31-56, 2014.

[16] S. Li, L. Yuan, J. Sun, and L. Quan, "Dual-Feature Warping-based Motion Model Estimation," in ICCV, Santiago, 2015, pp. 4283 - 4291.

[17] K. Joo, N. Kim, T. Oh, and I. S. Kweon, "Line meets as-projective-as-possible image stitching with moving DLT," in ICIP, Quebec City, QC, 2015, pp. 1175 - 1179.

[18] R. G. von Gioi, J. Jakubowicz, J. M. Morel, and G. Randall, "LSD: A Fast Line Segment Detector with a False Detection Control," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, pp. 722-732, 2010.

[19] B. Fan, F. Wu and Z. Hu, "Robust line matching through line-point invariants," Pattern Recognition, vol. 45, pp. 794-805, 2012.

[20] Z. Fu, Z. Sun, "An algorithm of straight line features matching on aerial imagery," in XXI ISPRS Congress, Beijing, China, 2008.

[21] J. Zhu, M. Ren, "Image Mosaic Method Based on SIFT Features of Line Segment," Computational and Mathematical Methods in Medicine, vol 2014, pp. 1-11, 2014.

[22] E. Dubrofsky and R. J. Woodham, "Combining Line and Point Correspondences for Homography Estimation," in International Symposium on Visual Computing, Las Vegas, NV, USA, 2008, pp. 202-213.

[23] T. Xiang, G.-S. Xia, L. Zhang, "Image stitching with perspective-preserving warping," in XXIII ISPRS Congress, Prague, Czech Republic, 2016.