

Parallax-Tolerant Image Stitching Based on Robust Elastic Warping

Jing Li¹, Zhengming Wang, Shiming Lai, Yongping Zhai, and Maojun Zhang

Abstract—Image stitching aims at generating high-quality panoramas with the lowest computational cost. In this paper, we propose a parallax-tolerant image stitching method based on robust elastic warping, which could achieve accurate alignment and efficient processing simultaneously. Given a group of point matches between images, an analytical warping function is constructed to eliminate the parallax errors. Then, the input images are warped according to the computed deformations over the meshed image plane. The seamless panorama is composed by directly reprojecting the warped images. As an important complement to the proposed method, a Bayesian model of feature refinement is proposed to adaptively remove the incorrect local matches. This ensures a more robust alignment than existing approaches. Moreover, our warp is highly compatible with different transformation types. A flexible strategy of combining it with the global similarity transformation is provided as an example. The performance of the proposed approach is demonstrated using several challenging cases.

Index Terms—Image stitching, image alignment, elastic warping, feature refinement, computational efficiency, computer vision.

I. INTRODUCTION

SITCHING images with parallax is still a challenging task in computer vision. It is crucial for the generation of high-resolution panoramic images and videos, which play an important role in applications of multimedia technology nowadays, such as surveillance [1], immersive communication [2] and virtual reality [3]. Traditional stitching approaches usually estimate an optimal global transformation for each input image. A representative example is AutoStitch [4], as proposed by Brown *et al.* Global methods work well for ideal cases when the camera translation is negligible [5] or the scene is near planar [6].

Manuscript received April 22, 2017; revised September 10, 2017 and November 2, 2017; accepted November 8, 2017. Date of publication November 24, 2017; date of current version June 15, 2018. This work was supported in part by the National Natural Science Foundation of China under Grants 61402491 and 61703415. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Sen-Ching Samson Cheung. (*Corresponding author: Jing Li.*)

J. Li, S. Lai, Y. Zhai, and M. Zhang are with the College of Information System and Management, National University of Defense Technology, Changsha 410073, China (e-mail: jingli@nudt.edu.cn; shiming413@nudt.edu.cn; zhaiyongping08@nudt.edu.cn; mjzhang@nudt.edu.cn).

Z. Wang is with the College of Science, National University of Defense Technology, Changsha 410073, China (e-mail: zmw@nudt.edu.cn).

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>. The material is a pdf that shows further results of experiments to demonstrate the performance of the proposed method in the paper. The file is 47 MB in size.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMM.2017.2777461

However, for general cases with parallax, such as the input images in Fig. 1, undesirable artifacts will appear in the resultant panorama.

To achieve better stitching quality, state-of-the-art approaches usually construct a local adaptive stitching field based on the point correspondences between images. Since the global transformation cannot meet these requirements, several space-varying transformation models [7]–[12] are introduced, of which the as-projective-as-possible (APAP) warp [7] is representative. APAP computes a local homography for every image patch to achieve high-precision local alignment. However, the homographic model is likely to cause projective distortions in the non-overlapped regions. To address this problem, the global similarity transformation is introduced to provide more natural-looking stitching results [12]–[14]. Inspired by the idea of content-preserving warps [15]–[17], researchers have proposed a series of direct warping approaches [13], [14], [18], in which the source image is divided into a uniform grid mesh, which is smoothly stretched to mitigate the inconsistency with the target image. Combined with seam cutting [13], [18], satisfactory results can be achieved for cases with large parallax.

Other than the mesh-based models, we have investigated non-rigid warping approaches that directly compute a deformation function across the image plane, such as optical flow [19]–[21], structure deformation [22], [23] and thin plate spline (TPS) [24]–[26]. Given the estimated global transformation, these methods regard the local alignment as a general matching problem with a smoothness constraint. The calculation of the optical flow field involves dense matching of image content, which is a time-consuming task. Moreover, the optical flow becomes less reliable as the disparity becomes larger. Therefore, it is seldom directly used to handle parallax in image stitching. For structure deformation and TPS, the alignment term is constructed based on a few anchor points. Therefore, the warps are very sensitive to matching errors. In other words, the existence of outliers in the matching data will lead to severe distortions. However, as far as we know, little research has focused on outlier removal for elastic registration of images. Due to these defects, the application of deformation functions to natural image stitching is seriously restricted.

In this paper, we concentrate on both accurate and efficient alignment of images. A parallax-tolerant image stitching method based on robust elastic warping is proposed. The fundamental mathematical framework of our method is built on the TPS model [25]. We first illustrate the basic theories of the TPS and the elastic warping model given a group of



Fig. 1. Comparison of global alignment and our local alignment. Left: Input images of the *railtracks* [7] database. Middle: Stitching result of global alignment, in which serious ghosting exists. Right: Stitching result of our method, in which the misalignments are eliminated.

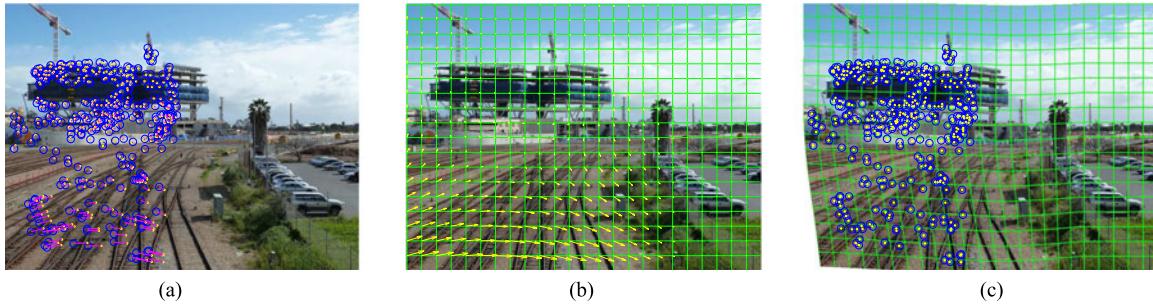


Fig. 2. Robust elastic warping. (a) Projection biases of feature points. (b) The meshed deformation field built using our method. (c) The warped source image in which the projection biases of feature points are eliminated.

point matches. Second, by analyzing the statistical regularities of the warping parameters, we propose a Bayesian model of feature refinement to adaptively remove the local outliers of the matching data. After a reliable and accurate alignment in the overlapping region is accomplished, we gradually change the local warps to the global transformation by linearly reducing the deformation function across the non-overlapping region, such that the global projectivity is preserved. Benefiting from the high compatibility of our warp, we also provide a flexible and efficient strategy to combine it with the global similarity transformation. The proposed method can be easily extended to multiple image stitching, and we conclude by presenting the entire extended work flow as well as some important principles.

The proposed warping can be regarded as a combination of the mesh-based model and the direct deformation strategy. As shown in Fig. 2(a), the projection biases of feature points are computed using the global transformation at first. Then the biases over the whole image can be interpolated using an analytical deformation function. To accelerate the computation, we set a grid mesh over image plane and only compute the function values on the mesh nodes, as shown in Fig. 2(b). The deformations of other pixels are directly obtained from bilinear interpolation. The computational efficiency is enhanced while preserving the alignment accuracy. Fig. 2(c) shows the warped image in which the projection biases between feature points are eliminated. Finally, the panorama is composed by re-projecting the warped images, as shown in Fig. 1.

The rest of the paper is organized as follows: Section II reviews important related work on image stitching. Section III

introduces the proposed method in detail. Section IV extends the method to stitch multiple images. Experimental results and comparisons with other methods are presented in Section V. Section VI concludes the paper.

II. RELATED WORKS

A thorough review of traditional stitching methods can be found in a previous work [27]. An excellent work named AutoStitch, which was proposed by Brown and Lowe [4], can automatically stitch multiple images. A global transformation was computed for each image under the assumption that all input images are captured rotationally. If this requirement is violated, visible artifacts will appear in the resultant panorama.

To resolve the parallax issues, state-of-the-art approaches have resulted in three kinds of strategies: 1) seam-driven stitching [13], [18], [22], [28], 2) content-preserving warps [11], [13], [29] and 3) local adaptive transformations [7]–[12]. Assuming the scene is composed by two predominating planes, Gao *et al.* proposed a dual homography warping (DHW) model [9]. It performs well for the specified scenes, but cannot handle more complex conditions. Noticing that seam-cutting is often the most crucial step for obtaining a perceptually seamless result, a seam-driven image stitching strategy [28] was proposed by which the transformation estimation is guided using the visual quality of the seam-cut. Lin *et al.* introduced a smoothly varying affine (SVA) field [8] to achieve local adaptive image stitching while maintaining global affinity. By explicitly segmenting and matching the planar regions, Lou and Gevers proposed the piecewise planar region matching approach [10]

to achieve more robust image registration. Zaragoza et al. proposed the as-projective-as-possible (APAP) warp [7] which applies local adaptive projective transformations to the meshed image plane using a simple moving direct linear transformation (moving-DLT). APAP achieves more accurate alignment than DHW in overlapping regions and better extrapolation quality than SVA in non-overlapping regions. Zhang et al. propose a multi-viewpoint panorama stitching method [30], in which a local DLT (direct linear transformation) approach is introduced to remove incorrect matching data between wide-baseline images. The constructed panoramas are locally perspective and globally orthogonal.

The seam-selection and content-preserving warping techniques are synthesized by Zhang and Liu to stitch images with large parallax [13]. A plausible seam is estimated by considering both geometric alignment and image content. The input images are then roughly aligned with the optimal homography and the alignment is further refined using content-preserving warping. Chang et al. proposed shape-preserving half-projective (SPHP) warping [11], by which the projective transformations are gradually changed to a global similarity transformation across the image. Combined with APAP, SPHP can provide good alignment accuracy with less projective distortion. Lin et al. proposed the adaptive as-natural-as-possible (ANAP) warp [12] for more perceptually natural stitching. Like SPHP, ANAP adopts a gradually changing transformation field composed by the linearized homography and the estimated global similarity transformation. To avoid the possible unnatural rotation problem, ANAP groups the matched points according to their dominated plane and chooses the point set with smallest rotation angle for computing the optimal similarity transformation. However, unnatural rotation and scaling still exist while stitching multiple images. A more robust stitching method with the global similarity prior is proposed by Chen and Chuang [14]. A matched point mesh between each adjacent image pair is generated using APAP. Then an objective function considering both local and global similarity are minimized to obtain natural-looking stitching results.

The above approaches are all based on point matches. However, as pointed out by Joo et al. [31], line matches can provide stronger correspondence than point matches and the straightness of lines can serve as additional supplemental cues. A line guided moving DLT (L-mDLT) framework is introduced. Combined with moving DLT [7], L-mDLT achieves more accurate alignment for some challenging cases. A complete workflow of line detection, matching, and refinement is provided by Li et al. [32]. Lin et al. propose a seam-guided local alignment (SEAGULL) approach [18], in which the final stabilized warp is accomplished through iteratively computing the seam location and the structure-preserving warping. SEAGULL detects line segments [33] in the target image to preserve line structures during the warping. Recently, Lin et al. propose a mesh-based photometric alignment (MPA) method [29] that combines the superior performance of dense photometric alignment with the efficiency of mesh-based image warping. Assuming the corresponding points in the two images obey the brightness constancy constraint, MPA minimizes pixel intensity difference instead of

Euclidean distance of known feature correspondences. Similar line structure preserving strategy is adopted in MPA as that in SEAGULL. The line-matches framework has evident advantages but suffers from high computational complexity that restricts its application range.

III. ROBUST ELASTIC LOCAL ALIGNMENT

A. Elastic Local Alignment

Given two overlapped images I_p and I_q and their matched points $\mathbf{p}_i = (x_i, y_i)^T$, $\mathbf{q}_i = (u_i, v_i)^T$, $i = 1, \dots, n$, the global transformation between I_p and I_q can be estimated using DLT [34] or the method presented by AutoStitch [4]

$$\hat{\mathbf{x}}' \sim \mathbf{H}\hat{\mathbf{x}} \quad (1)$$

where \sim denotes equality up to a scale, \mathbf{H} is the global homography, $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ are a pair of matching points (in homogeneous coordinates) of I_p and I_q . We use $\mathbf{p}'_i = (x'_i, y'_i)^T$ to denote the projection of \mathbf{p}_i in I_q computed by (1). The parallax error on \mathbf{p}'_i is represented by projection bias $\mathbf{g}_i = \mathbf{p}'_i - \mathbf{q}_i = (g_i, h_i)^T$, as shown in Fig. 2(a). For input images with parallax, $\mathbf{g}_i \neq 0$ and local adaptive alignment approaches are designed to eliminate these errors.

Originating with the elastic warp introduced to address the problem of registering medical images [26], [35], varieties of warping approaches with different characteristics are designed to suit specific application requirements. Radial basis functions (RBFs) have proven to be powerful tools in image warping given some anchor points [36]. We choose the thin plate spline (TPS) [25], [26] with a simple RBF type to formulate the image deformation due to its good performance in both alignment quality and efficiency.

Specifically, the deformation across image I_q can be represented as $\mathbf{g}(x, y) = (g(x, y), h(x, y))^T$, where $g(x, y)$ and $h(x, y)$ are deformations in the x and y directions, respectively. For simplicity, in the following analysis, we use $g(x, y)$ instead of $\mathbf{g}(x, y)$ since the computing of $g(x, y)$ and $h(x, y)$ are totally identical and mutually independent. The energy function for optimal warp contains two terms: the alignment term J_D and the smoothness term J_S , which are defined as

$$J_D = \sum_{i=1}^n (g(x'_i, y'_i) - g_i)^2 \quad (2)$$

$$J_S = \iint_{(x,y) \in \Omega} |\nabla^2 g|^2 dx dy \quad (3)$$

The overall energy function to be minimized is

$$J_\lambda = J_D + \lambda J_S \quad (4)$$

where λ is a weighting parameter used to balance the two terms. According to basic theories of TPS [26], the optimal solution to minimize (4) can be expressed as

$$g(x, y) = \sum_{i=1}^n \omega_i \phi_i(\mathbf{x}) + \alpha_1 x + \alpha_2 y + \alpha_3 \quad (5)$$

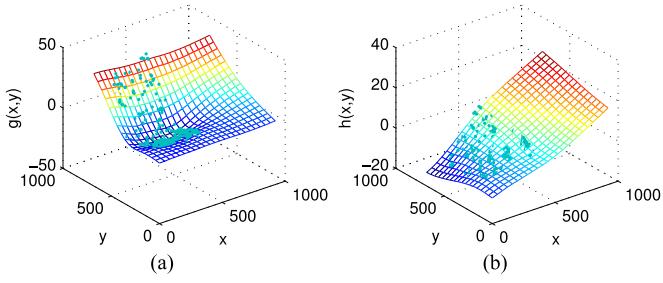


Fig. 3. Deformation function. Figures (a) and (b) are the two components of the computed deformation function $\mathbf{g}(x, y) = (g(x, y), h(x, y))^T$ given the scattered biases \mathbf{g}_i .

where $\phi_i(\mathbf{x}) = |\mathbf{x} - \mathbf{p}'_i|^2 \ln |\mathbf{x} - \mathbf{p}'_i|$ is the RBF. The coefficients $\mathbf{w} = (\omega_1, \dots, \omega_n)^T$ and $\mathbf{a} = (\alpha_1, \alpha_2, \alpha_3)^T$ can be computed by solving the following linear system

$$\begin{pmatrix} \mathbf{K} + \lambda C_{2,2} \mathbf{I} & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0}_{3 \times 3} \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \mathbf{a} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{0}_{3 \times 1} \end{pmatrix} \quad (6)$$

where $C_{2,2} = 8\pi$ is a constant, $\mathbf{K} = (\phi_i(\mathbf{p}'_j)) \in \mathbb{R}_{n \times n}$, $\mathbf{P} = (\hat{\mathbf{p}}'_1, \dots, \hat{\mathbf{p}}'_n)^T \in \mathbb{R}_{n \times 3}$, $\mathbf{f} = (g_1, \dots, g_n)^T$. (5) has a unique solution if the matches number $n \geq 4$ [26]. The deformation of an arbitrary position $(x, y)^T$ in the source image can be acquired by putting \mathbf{w} and \mathbf{a} into (5).

To accelerate the computation, a uniform grid mesh of $C_x \times C_y$ cells is set over the image plane. The deformations are first computed on the mesh nodes, and then linearly interpolated to all the other pixels. For input images presented in Fig. 1, the computed deformation function is shown in Fig. 3.

Based on the computed deformation $\mathbf{g}(x, y)$, I_q can be warped to eliminate the misalignments with I_p . An example of the meshed deformation field is depicted in Fig. 2(b). We denote the warped image as \tilde{I}_q . For an arbitrary point $\mathbf{q}_0 = (x_0, y_0)^T$ in I_q and its corresponding point $\mathbf{q} = (x, y)^T$ in \tilde{I}_q , we have

$$\begin{cases} x_0 = x - g(x, y) \\ y_0 = y - h(x, y) \end{cases} \quad (7)$$

The warping is accomplished by invoking *inverse* mapping with bilinear interpolation [37]. The warped image of Fig. 2(a) is shown in Fig. 2(c). In practice, the warping can be enclosed in the transformations of look-up-tables (LUTs) as an intermediate step. Thus \tilde{I}_q need not be explicitly generated. The seamless panorama is composed by re-projecting and blending the aligned images on the panoramic image plane.

B. Bayesian Refinement of Feature Matches

During feature detection and matching, position-error and mismatch problems are practically inevitable for natural images. It is necessary to eliminate the incorrect matches for accurate alignment. Most state-of-the-art approaches [7], [11], [12], [14] employ RANSAC [38] to remove outliers of the matching data. From the perspective of computational efficiency, global transformations usually serve as the minimal solver of RANSAC before estimating the local warps. Then a contradiction arises between the *global* RANSAC and the *local* adaptive model. This will result in two negative consequences: 1) inliers might

be removed if their projection biases are larger than the threshold; and 2) outliers might be preserved if their projection biases are less than the threshold. Fig. 4 shows a case (the *intersection* database [12]) with relatively large parallax. Fig. 4(a) is the outlier removal result using global RANSAC, in which some important matching points on the ground are mistakenly removed but a few outliers are preserved. It is very difficult to achieve accurate alignment in this situation. The stitching result of APAP is presented in Fig. 4(b), which contains evident artifacts.

As far as we know, little research attention has been paid on local outlier removal for image alignment. Zhang *et al.* propose a local DLT approach to remove outliers in their wide-baseline image stitching method [30]. A homography is computed for every feature point by employing DLT over its neighborhood of R pixels. A match is considered as an inlier if its residual error is less than the threshold γ . The local DLT approach is effective for textured areas, but would fail for areas with no sufficient features. Specifically, a reliable homography cannot be estimated for a position if the matched feature number within its neighborhood is less than 4 [34]. Fig. 4(c) shows the outlier removal result of local DLT setting $R = 50$ and $\gamma = 5$ as suggested by the authors of the orginal paper. Some inliers are mistakenly removed since there are no sufficient matches around them to support the judgment of whether they are correct or not. The stitching result of APAP combined with local DLT is presented in Fig. 4(b).

By analyzing the distribution of weights ω_i (for x-direction) and v_i (for y-direction) that are attached to the RBF $\phi_i(\mathbf{x})$, we propose a probabilistic model for robust feature refinement. As a subsequent step after a loose global RANSAC, it could remove local outliers adaptively. Since ω_i and v_i are computed globally, our model will not suffer from the issues caused by insufficient matches. The feature-refinement result of the proposed model is shown in Fig. 4(e), which is much better than those shown in Fig. 4(a) and (c). It is worth noting that the proposed feature refinement model involves manipulations of the weights computed in Section III-A. Therefore, it has to be employed along with the elastic local alignment model.

For convenience of quantitative analysis, two idealized assumptions about the matching data are made first.

Assumption 1: The variations of projection biases between two adjacent pixels are independent and identically distributed.

Assumption 2: Every anchor point is separated from all the others by an adequate number of pixels.

The term *pixel* in above assumptions is not the real measurement unit of digital images, but an abstract conception, which means the projections of the scene points on the image plane. Therefore, an idealized image could contain an infinite number of such *pixels*. An interesting corollary can be deduced from the central limit theorem under Assumptions 1 and 2.

Corollary 3: The difference between biases \mathbf{g}_i and \mathbf{g}_j , $j \neq i$ obeys the normal distribution.

According to the properties of TPS [24], the weights (ω_i, v_i) are roughly proportional to the difference between \mathbf{g}_i and the projection biases of other anchor points around \mathbf{p}'_i . In accordance with Corollary 3, the values of ω_i and v_i approximately obey the normal distribution, too. Specifically, $\omega_i \sim N(0, \sigma_\omega)$ and $v_i \sim N(0, \sigma_v)$, where σ_ω and σ_v are corresponding stan-

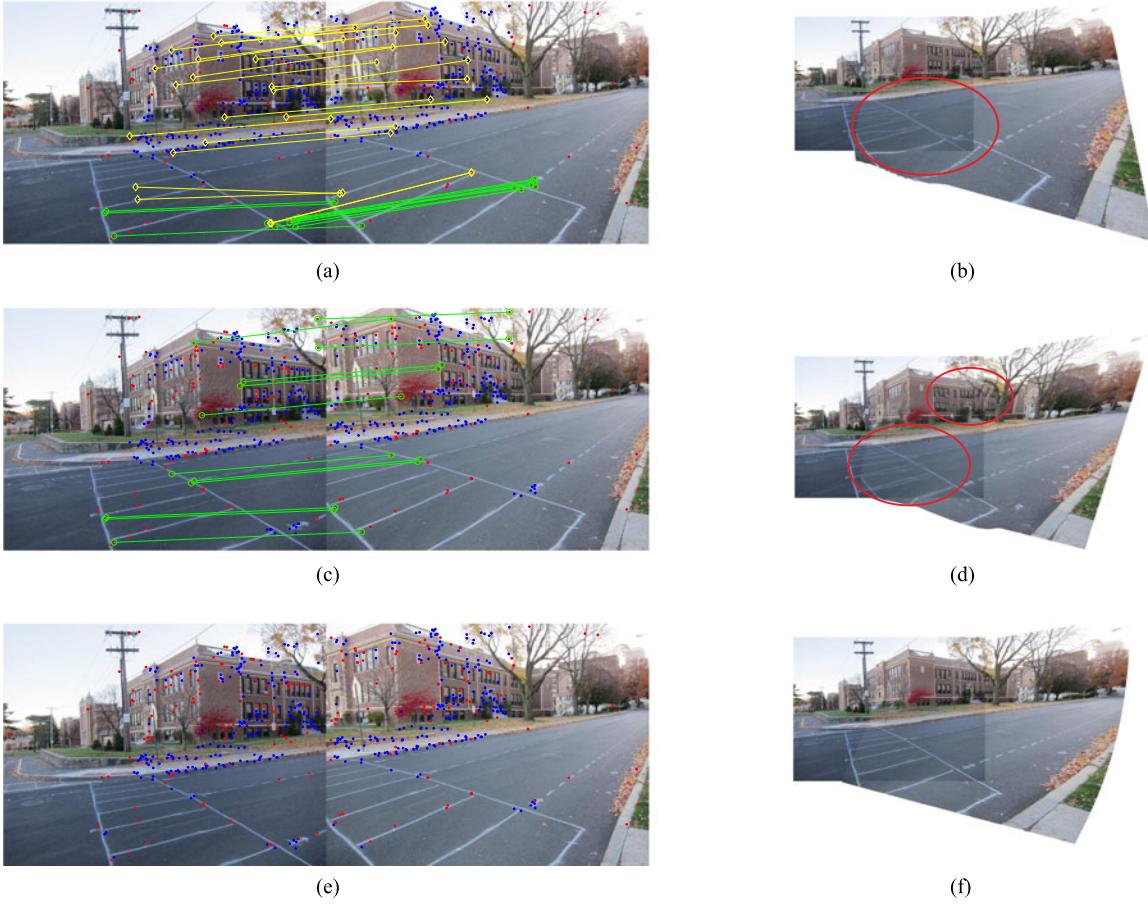


Fig. 4. Outlier removal. (a) The feature refinement result of global RANSAC. The blue points and the red ones respectively represent the preserved data and removed data. Some of the removed inliers are marked in green, while the preserved outliers are marked in yellow. (b) The stitching result of APAP, which uses the global RANSAC to remove outliers. (c) The feature refinement result of local DLT, in which some of the removed inliers are marked in green. (d) The stitching result of APAP combined with local DLT. (e) The feature refinement result of our Bayesian model by which the local outliers are removed adaptively. (f) The stitching result of our method.

dard deviations. The mean value of ω_i is equal to 0, since the bottom row of (6) ensures $\sum_{i=1}^n \omega_i = 0$, and the same is true of v_i .

Given a pair of matched points $\{\mathbf{p}_i, \mathbf{q}_i\}$, whether this match is correct or not is represented by a binary variable $\tau_i \in \{0, 1\}$. We denote the proportion of inliers in all matches as p_1 , such that $p(\tau_i = 1) = p_1$ and $p(\tau_i = 0) = 1 - p_1$. Considering the event $A_i = \{|\omega_i| > t\sigma_\omega\}$, we have $p(A_i | \tau_i = 1) = 2(1 - \Phi(t))$, where $\Phi(\cdot)$ is the standard normal distribution function. The probability of A_i conforms to

$$\begin{aligned} p(A_i) &= \sum_{k=\{1,2\}} p(A_i | \tau_i = k) p(\tau_i = k) \\ &\leq p(A_i | \tau_i = 1) p(\tau_i = 1) + p(\tau_i = 0) \\ &= 1 - (2\Phi(t) - 1)p_1 \end{aligned} \quad (8)$$

Under the condition of $|\omega_i| > t\sigma_\omega$, the posterior probability of $\tau_i = 0$ can be computed according to the Bayesian rule

$$p(\tau_i = 0 | A_i) = \frac{p(\tau_i = 0)}{p(A_i)} \geq \frac{1 - p_1}{1 - (2\Phi(t) - 1)p_1} \quad (9)$$

We set $t = 3$ in accordance with the three-sigma rule so that $\Phi(t) = 0.9987$. If $p_1 \in [0, 0.9973]$, then $p(\tau_i = 0 | A_i) \geq 0.5$.

It means that the match $\{\mathbf{p}_i, \mathbf{q}_i\}$ is more likely to be incorrect if the event $A_i = \{|\omega_i| > 3\sigma_\omega\}$ occurs.

Based on above analysis, the distributions of weights (ω_i, v_i) can serve as the criteria for removing outliers. On the premise that $p_1 \in [0, 0.9973]$, if $|\omega_i| > 3\sigma_\omega$ or $|v_i| > 3\sigma_v$, we regard $\{\mathbf{p}_i, \mathbf{q}_i\}$ as an outlier and remove it from the matching data. Geometrically, the criteria can be explained by observing that one match is very likely to be incorrect if its projection bias significantly differs from other matches nearby. Though a few inliers might be mistakenly removed due to the nature of the probabilistic model, the alignment quality will not be influenced. The stitching result based on the refined matching data is shown in Fig. 4(f).

Since the parameters σ_ω and σ_v are unknown in practice, we approximate them using the standard deviations $\tilde{\sigma}_\omega$ and $\tilde{\sigma}_v$ derived from all the current matches. The outliers are removed iteratively by invoking Algorithm 1. After feature refinement, the distribution of weights ω_i and v_i corresponding to Fig. 4(e) are depicted in Fig. 5. In our experiments, since the refinement finished within 10 loops in most cases, we set a maximum iteration number $L = 10$. This setting also prevents excessive removal of inliers in the exceptional conditions that the basic assumptions about the matching data are violated.

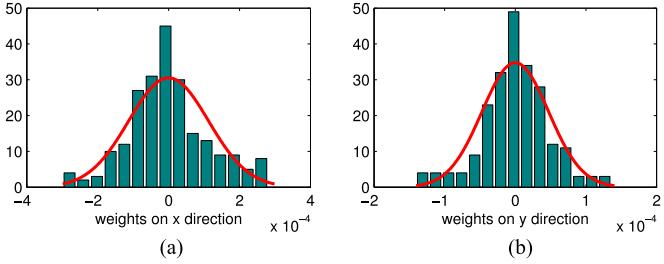


Fig. 5. Histograms of the weights. (a) and (b) are the weight distributions for $g(x, y)$ (x-direction) and $h(x, y)$ (y-direction) after feature refinement.

Algorithm 1: Bayesian refinement of feature matches

Require: Matched points $\{\mathbf{p}'_i, \mathbf{q}_i\}, i = 1, \dots, n$.

- 1: Formulate (6) based on $\{\mathbf{p}'_i, \mathbf{q}_i\}, i = 1, \dots, n$.
- 2: Set the current match number $n_l = n$.
- 3: Solve (6) to get the weights $(\omega_i, v_i), i = 1, \dots, n$.
- 4: **for** $l = 1, \dots, L$ **do**
- 5: Compute the standard deviations $\tilde{\sigma}_\omega$ and $\tilde{\sigma}_v$.
- 6: Mark all matches that make $|\omega_i| > 3\tilde{\sigma}_\omega$ or $|v_i| > 3\tilde{\sigma}_v$.
- 7: **if** the number of marked matches $< 0.0027n_l$ **then**
- 8: Break out of the loop.
- 9: **end if**
- 10: Remove all rows and columns regarding the marked matches from (6) and formulate the new equations.
- 11: Update the match number n_l .
- 12: Solve the new equations to get the updated weights $(\omega_i, v_i), i = 1, \dots, n_l$.
- 13: **end for**

C. Smooth Transition to Global Transformation

The robust elastic warping model to align the overlapping region has been introduced in the previous sections. However, directly extrapolating the warps to the non-overlapping region will suffer from the over-fitting problems. The undesirable trend can be observed in Fig. 3. We would like to mitigate this trend by smoothly changing the local warps to an estimated global transformation, particularly, the global homography \mathbf{H} . This can be easily done by gradually reduce the deformation function $\mathbf{g}(x, y)$ to zero across the non-overlapping region. The suppressed deformation $\mathbf{g}_s(x, y)$ is expressed as follows

$$\mathbf{g}_s(x, y) = \eta \mathbf{g}(x, y) \quad (10)$$

While the point (x, y) is moving away from the overlapping region, the scale parameter η is gradually changed from 1 to 0 as

$$\eta = 1 - \frac{\max(0, x - x_u, x_l - x, y - y_u, y_l - y)}{d_s} \quad (11)$$

where $[x_l, x_u]$ and $[y_l, y_u]$ are the boundaries of the overlapping region in the x and y directions, respectively, and d_s is the smooth transition width in pixels, which is set to be directly proportional to the maximum bias. Specifically,

$$d_s = K \times \max\{|g_1|, \dots, |g_n|, |h_1|, \dots, |h_n|\} \quad (12)$$

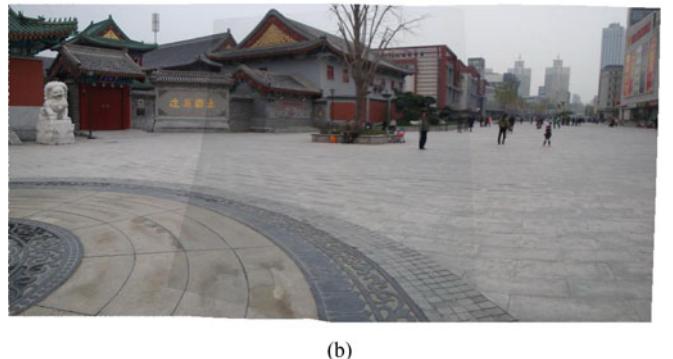
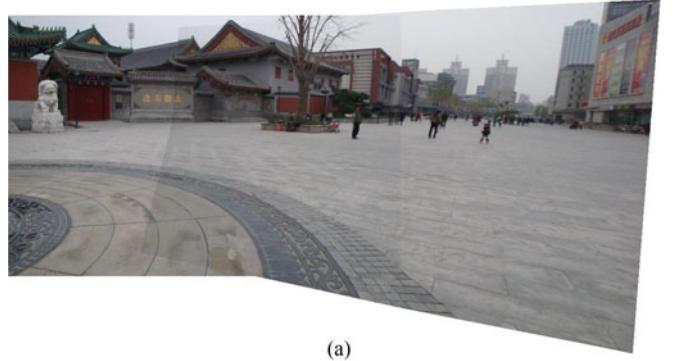


Fig. 6. Comparison between different global transformations. (a) The result using homography as the global transformation. (b) The result from combining similarity transformation with zero rotation.

where K is a scale parameter. The suppressed deformation $\mathbf{g}_s(x, y)$ becomes 0 outside the transition area so that the global projectivity in the non-overlapping region is well preserved.

D. Combination With Global Similarity Transformation

The proposed warp involves only operations on image content, which is not relevant to the transformation types. We use global homography because it is a good approximation of the local projectivity in the overlapping region. However, global homography may lead to projective distortions in the non-overlapping region while stitching wide-field-of-view images, as illustrated in Fig. 6(a) using the widely tested *temple* [9] dataset. Some state-of-the-art approaches [11]–[14], [39] employ the similarity transformation due to its good performance in mitigating projective distortions. An overall space-varying stitching field with both accurate alignments and natural looking projectivity can be constructed by combining the two transformations, as shown in Fig. 6(b), which looks more natural than Fig. 6(a). Different from existing local models, we warp the input images before re-projecting them so that the misalignments are eliminated before the transformations are applied. Then the subsequent stitching procedures can be interpreted as manipulations on the parallax-free images, which means arbitrary global transformation can be applied without any influence on alignment accuracy. Therefore, the proposed warping model is highly compatible with the transformation types.

After the refined point matches $\{\mathbf{p}_i, \mathbf{q}_i\}, i = 1, \dots, n_r$ are received from the method presented in Section III-B, we esti-

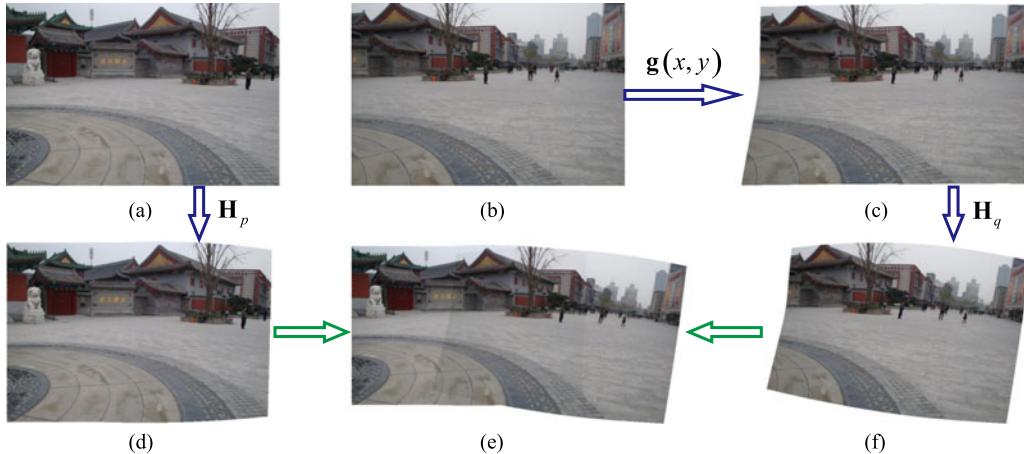


Fig. 7. Elastic warping combined with global similarity transformation. (a) and (b) are the input images [9]. (b) is warped using the deformation function $g(x, y)$ to produce the warped image (c). (c) is then transformed using H_q to get (f). Meanwhile, (a) is transformed using H_p to get (d). The panoramic image (e) is synthesized by (d) and (f) at last.

mate the global similarity transformation \mathbf{S} by minimizing

$$J_A = \sum_{i=1}^{n_r} \|\mathbf{S}\hat{\mathbf{p}}_i - \mathbf{q}_i\|^2 \quad (13)$$

where $\mathbf{S} = \begin{pmatrix} c & -s & t_x \\ s & c & t_y \end{pmatrix}$. The direct linear transform (DLT) [34] is invoked to compute the optimal \mathbf{S} . We mix the homography and the similarity transformation using the techniques proposed in ANAP [12].

$$\mathbf{H}_q = \mu_h \mathbf{H} + \mu_s \mathbf{H}_s \quad (14)$$

where \mathbf{H}_s is composed by adding row $(0, 0, 1)$ to the bottom of \mathbf{S} , and μ_h and μ_s are weighting coefficients with the constraint $\mu_h + \mu_s = 1$. μ_h linearly varies from 1 to 0 across the source image I_q (precisely, the warped source image \tilde{I}_q) following the setting of ANAP [12]. The target image I_p is also transformed accordingly as

$$\mathbf{H}_p = \mathbf{H}_q \mathbf{H}^{-1} \quad (15)$$

The overall procedures are illustrated in Fig. 7. If we set $s = 0$ while minimizing (14), then \mathbf{S} contains only scaling and offsetting (without rotation). The resulting panorama is shown in Fig. 6(b).

Different from ANAP and other approaches (such as SPHP [11]) that employ local homographies, our method only relies on the global homography \mathbf{H} as well as the global similarity transformation \mathbf{S} . This characteristic provides a relatively more efficient performance while combining the transformations.

E. Complexity and Efficiency Analysis

So far, the algorithm for pairwise image stitching has been presented. We summarize the methodology as follows: In elastic local alignment of Section III-A, a feature-based deformation function is constructed to eliminate the misalignments between images. The parameters of the function can be computed by solving a linear system according to the basic theories of TPS.

For a pair of overlapped images with n matches, the computational complexity of solving the n -dimensional equations is $O(n^3)$. To ensure the reliability of alignment, a probabilistic feature refinement strategy is introduced in Section III-B. The outliers are iteratively removed according to the probabilistic distribution of the computed weights attached to the RBFs. Therefore, the actual computational complexity of parameter estimation is $O(Ln^3)$, given the maximum number L . The deformation over the entire image is computed by summing the weighted RBFs. For input images of size $X \times Y$, the computational complexity of deformation computing will be $O(nXY)$. Without sacrificing the quality of alignment, we set a uniform grid mesh of $C_x \times C_y$ cells over the image plane to accelerate the computation, which reduces the computational complexity to $O(nC_xC_y + XY)$. In Section III-C, the local warps are smoothly changed to the global transformation by linearly decreasing the deformation function to 0 in the non-overlapping region. Moreover, benefiting from the high compatibility of our method, the global transformation is not restricted to the previously estimated homography. A flexible strategy of combining the proposed warp with the global similarity transformation is discussed and presented in Section III-D. Both the scaling operation on the deformation function and the transition to the global similarity transformation have $O(XY)$ complexity.

In a word, the overall computational complexity of our method is $O(Ln^3 + nC_xC_y + XY)$. Experimental results in Section V confirm that our method runs much faster than other local adaptive stitching methods.

IV. MULTIPLE IMAGE STITCHING

The proposed warping method can be easily extended to stitch multiple images. The elastic deformation is caused by directly moving every pair of biased feature points to be co-incident, which is irrelevant to the projection model. In this way, the global optimization technique bundle adjustment [40] does not fit this situation. In fact, the incremental stitching scheme is powerful enough to provide satisfying results with the help of the global projectivity preserving approach presented in

Section III-C. Specifically, the input images are incrementally aligned and composited onto the panoramic image plane. During the alignment, two additional matters should be considered:

(i) *The simultaneous processing of matches.* The feature points of an image are possibly matched with feature points extracted from more than one image. While aligning the current image, these matches must be handled as a whole. For image I_i to be aligned, a pair of sets P_i and P'_i are allocated to collect the matching points related to I_i . We use I_j to denote an aligned image that is overlapped with I_i . $\{\mathbf{p}, \mathbf{q}\}$ is a pair of matching points between I_j and I_i . The point \mathbf{p} in I_j is projected to I_i using the global transformation, and then the projection \mathbf{p}' is included in P'_i . Meanwhile, its matching point \mathbf{q} is included in P_i . After all alignment images are traversed, the point sets P_i and P'_i are built, and these contain all point matches between I_i and the aligned images. Then the matches are simultaneously refined to construct the elastic warping model of I_i .

(ii) *The update of feature locations.* The warping of an aligned image will cause a slight shift of feature points within it. Before further dealing with subsequent images overlapping with the current aligned one, the feature location must be updated. Unlike the *inverse* mapping employed for image warping in Section III-A, the update of feature locations is essentially a *forward* mapping process, hence the new location cannot be directly read from the LUT. In this circumstance, $\mathbf{q} = (x, y)^T$ becomes the unknown instead of the original location $\mathbf{q}_0 = (x_0, y_0)^T$ in (7), which is nonlinear. However, noticing that the deformation function $\mathbf{g}(x, y) = (g(x, y), h(x, y))^T$ is Lipschitz continuous [41] for the limited image size (proof of this proposition is beyond the scope of this paper), and its Lipschitz constant is much less than 1 in the overwhelming majority of cases, $\mathbf{p} = (x, y)^T$ can still be instantly computed by applying fixed-point iteration on (7).

Given N input images $I_i, i = 1, \dots, N$ as well as the matching points between images $\{\mathbf{p}_{i,j,k}, \mathbf{q}_{i,j,k}\}, i = 1, \dots, N - 1, j = i + 1, \dots, N, k = 1, \dots, n_{i,j}$, the workflow of stitching multiple images is presented in Algorithm 2. Without loss of generality, the first image is chosen as the reference image.

V. EXPERIMENTS

We implemented¹ the proposed approach and carried out a group of experiments to evaluate its performance. These included 1) comparison with existing methods in terms of stitching quality; 2) testing its flexibility while combining it with different types of global transformations; 3) comparison of computational efficiency with state-of-the-art approaches; and 4) stitching results of multiple images. The compared methods include global homography, AutoStitch [4], APAP [7], SPHP [11], SPHP+APAP and ANAP [12], in which the first two are traditional ones with global transformation models and the others are with local adaptive stitching fields. In our experiments, the corresponding points are detected and matched using SIFT [42]. The same matching data is applied in all the tested methods.

Algorithm 2: Multiple image stitching based on robust elastic local alignment

```

Require: Input images  $I_i, i = 1, \dots, N$  and matching
points between images  $\{\mathbf{p}_{i,j,k}, \mathbf{q}_{i,j,k}\}, i = 1, \dots,
N - 1, j = i + 1, \dots, N, k = 1, \dots, n_{i,j}$ .
1: for  $i = 1, \dots, N$  do
2:   Clear the point sets  $P_i$  and  $P'_i$ .
3:   for  $j = 1, \dots, i - 1$  do
4:     if  $I_i$  and  $I_j$  are overlapped then
5:       For  $k = 1, \dots, n_{j,i}$ , compute the projection of
       $\mathbf{p}_{j,i,k}$  in  $I_i$ , denote it as  $\mathbf{p}'_{j,i,k}$ .
6:        $P_i = P_i \cup \{\mathbf{q}_{j,i,k} | k = 1, \dots, n_{j,i}\}$ .
7:        $P'_i = P'_i \cup \{\mathbf{p}'_{j,i,k} | k = 1, \dots, n_{j,i}\}$ .
8:     end if
9:   end for
10:  Formulate (6) based on the matches  $\{\mathbf{p}'_k, \mathbf{q}_k\}$ ,
     $k = 1, \dots, \sum_{j=1}^{i-1} n_{j,i}$ , where  $\mathbf{p}'_k \in P_i$  and  $\mathbf{q}_k \in P'_i$ .
11:  Solve (6) to obtain the local alignment parameters.
12:  Align  $I_i$  and composite it to the panorama.
13:  for  $j = i + 1, \dots, N$  do
14:    Update the location of  $\mathbf{p}_{i,j,k}$  for every match
     $\{\mathbf{p}_{i,j,k}, \mathbf{q}_{i,j,k}\}, k = 1, \dots, n_{i,j}$ .
15:  end for
16: end for

```

Parameter settings. The parameter settings for the compared methods follow the recommendations of their corresponding papers and the parameters with the best results are chosen if the value ranges are provided. For our method, the weighting parameter λ is set to 0.1% of the image size $X \times Y$ and the scale parameter K is set to 5 times the maximum bias. In fairness to the performance evaluation, all the resulting panoramic images generated by different methods are constructed using a constant cell size of 10×10 pixels, and the overlapping regions are simply set to the average of the reprojected images.

A. Comparison of Stitching Quality

Qualitative evaluation. In the previous sections, we have presented the stitching results of the *railtracks* [7], *intersection* [12] and *temple* [9] databases. Accurate alignment could be achieved for all these databases using our method. Fig. 8 shows a challenging case that contains some areas with poor or repetitive textures. The results of different methods are presented by rows and three representative areas of each resulting image are highlighted. As the baseline of comparison, the first row shows the results of global homography. Significant artifacts appear in all three highlighted areas. Rows 2 to 4 respectively present the results of APAP, SPHP+APAP and ANAP which introduce similar local adaptive transformations in the overlapping regions to achieve better stitching quality. Row 5 shows the results of our elastic local alignment model introduced in Section III-A. All the four *local* adaptive approaches use *global* RANSAC to remove outliers of the matching data. Due to the limitation of global RANSAC, some outliers cannot be successfully removed.

¹https://github.com/gain2217/Robust_Elastic_Warping

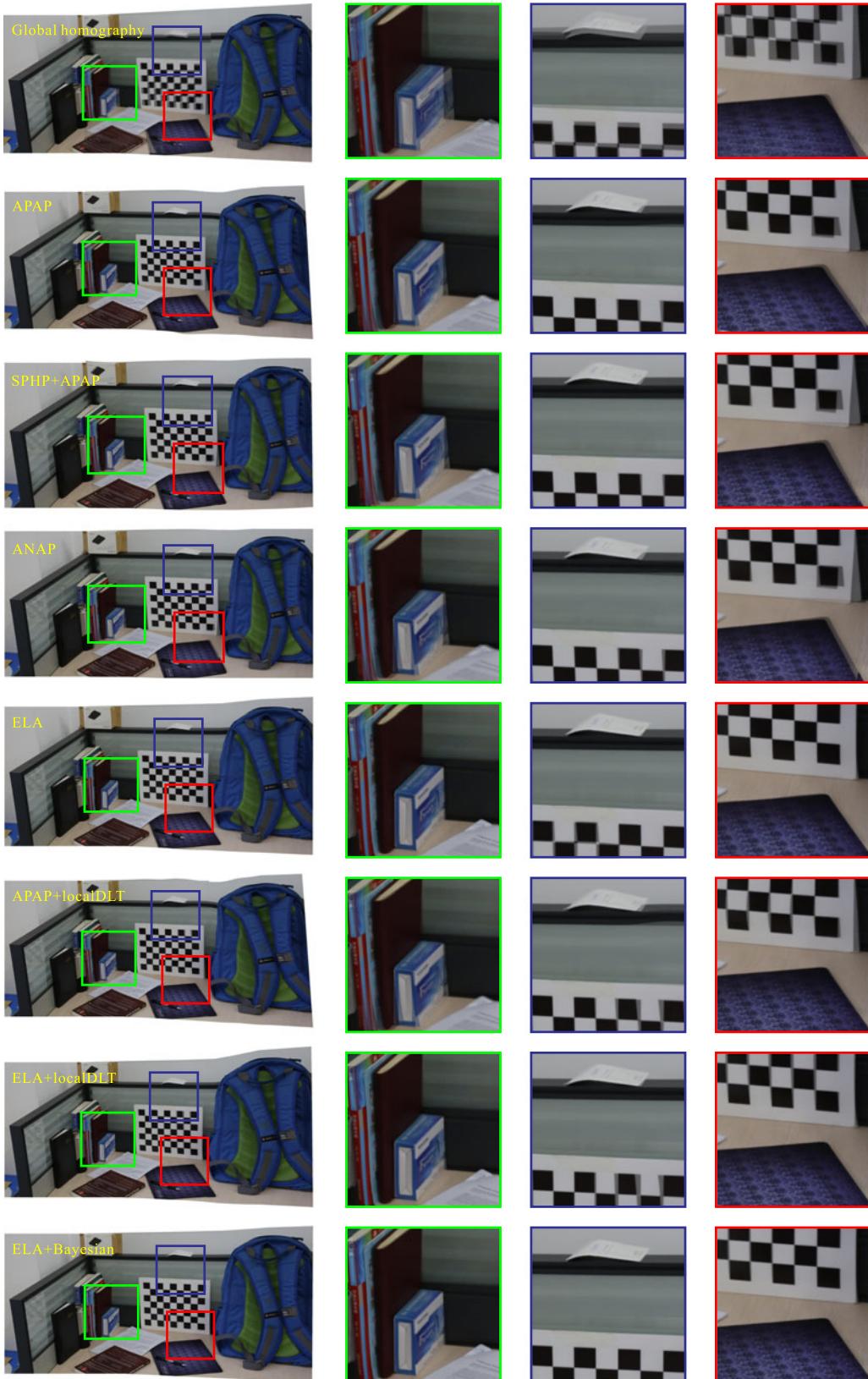


Fig. 8. Comparison of stitching quality on a *work table* scene. Row 1: Results using global homography. Row 2: Results using APAP [7]. Row 3: Results using SPHP [11] combined with APAP. Row 4: Results of ANAP [12]. Row 5: Results of our elastic local alignment (ELA) model. Row 6: Results of APAP combined with local DLT [30]. Row 7: Results of our ELA model combined with local DLT. Row 8: Results of our robust ELA model which involves Bayesian feature refinement.

TABLE I
COMPARISON OF SSIM ON DIFFERENT DATABASES

database	approaches using global outlier removal					approaches using local outlier removal		
	Baseline	APAP	SPHP+APAP	ANAP	ELA	APAP+IDLT	ELA+IDLT	robust ELA
<i>temple</i> [9]	0.4625	0.8945	0.9077	0.9145	0.8995	0.8897	0.8874	0.9072
<i>railtracks</i> [7]	0.3241	0.8938	0.8693	0.8732	0.9099	0.8688	0.9085	0.9176
<i>chess girl</i> [8]	0.7139	0.7924	0.7952	0.7908	0.7926	0.7874	0.7906	0.7933
<i>building</i> [12]	0.5911	0.7665	0.7817	0.8125	0.7709	0.7708	0.7714	0.7926
<i>fence</i> [12]	0.7472	0.8814	0.8825	0.8912	0.8927	0.8811	0.8959	0.8982
<i>park</i> [11]	0.7168	0.8657	0.8692	0.8699	0.8761	0.8636	0.8779	0.8766
<i>campus</i> [11]	0.5110	0.6551	0.6587	0.6594	0.6659	0.6529	0.6680	0.6634
<i>theater</i>	0.7019	0.9250	0.9299	0.9257	0.9393	0.8853	0.9223	0.9471
<i>cabin</i>	0.5237	0.9001	0.9042	0.9420	0.9603	0.8990	0.9621	0.9657
<i>footpath</i>	0.6453	0.9020	0.9039	0.9186	0.9176	0.9027	0.9181	0.9178
<i>wall</i>	0.7325	0.8978	0.9134	0.9243	0.9600	0.8984	0.9519	0.9586
<i>cabinet</i> [32]	0.9020	0.9295	0.9433	0.9278	0.9197	0.9263	0.9328	0.9486
<i>corner</i> [32]	0.7766	0.8432	0.8064	0.8440	0.8384	0.7939	0.7650	0.9042
<i>desk</i> [32]	0.7872	0.8717	0.9097	0.9070	0.8927	0.8935	0.8969	0.9139
<i>window</i> [32]	0.8171	0.8210	0.8085	0.8500	0.8454	0.7981	0.8279	0.8576
<i>tower</i>	0.8064	0.9429	0.9499	0.9474	0.9671	0.9488	0.9761	0.9826
<i>work table</i>	0.5289	0.8663	0.8803	0.8647	0.9296	0.8836	0.9177	0.9487
<i>racetracks</i>	0.4390	0.6726	0.6930	0.6881	0.7140	0.6727	0.7412	0.8403
<i>campus square</i>	0.8984	0.8952	0.9080	0.9374	0.9472	0.9014	0.9209	0.9734
<i>gym</i>	0.5928	0.9654	0.9553	0.9539	0.9459	0.9019	0.9012	0.9767

The spatially biased matches in the monochromatic areas might cause inaccurate alignment of the nearby image content, and the mismatches of repetitive structures would lead to undesirable distortions. Therefore, the results of APAP, SPHP+APAP, ANAP and ELA are better than global homography, but still contain evident artifacts. Since the smoothness term of our ELA model performs well in suppressing drastic local warping, the results of ELA are more stable than the other three.

Rows 6 to 8 of Fig. 8 provide the results of approaches using local outlier removal techniques. Rows 6 shows the results of APAP+local DLT [30], in which the misalignments are better eliminated than the results using global RANSAC. Rows 7 shows the results of our ELA model combined with local DLT, in which there are only a few slight artifacts. As analyzed in Section III-B, the proposed Bayesian feature refinement could ensure more robust feature refinement than local DLT. The results of our integrated robust ELA model involving Bayesian feature refinement are presented in Row 8. All three highlighted areas are accurately aligned and no visible artifacts exist.

Quantitative evaluation. We use the SSIM (structural similarity) [43] index to quantitatively evaluate the alignment quality of different approaches. The test images are collected from public available databases [7]–[9], [11], [12], [32] or captured by ourselves. The overlapping regions of every pair of input images are extracted to compute the corresponding SSIM after alignment. Table I shows the computed scores of the eight tested approaches, five of which are with global outlier removal and the other three are with local outlier removal. For the upper eleven cases of Table I with small or large parallax, all the seven local adaptive approaches could achieve perceptually satisfactory stitching results, thus yield very similar scores. Our two approaches (ELA and robust ELA) obtain relatively higher total scores than the others. The input images of the next nine cases are with weakly-textured or repetitively-textured areas, which may lead to matching errors and then influence the accuracy of

alignment. Compared with global RANSAC, the local outlier removal techniques can eliminate these matching errors more effectively, hence attain significantly higher scores. Among all the eight tested approaches, the proposed robust ELA constantly achieves the highest scores for all the nine difficult cases. The detailed stitching results of all the listed databases can be found in the supplementary materials at <http://ieeexplore.ieee.org>.

B. Flexibility Evaluation

As we have analyzed in Section III-D, our warping model is very flexible in that it can be integrated with different types of transformations. A more comprehensive evaluation is presented in Fig. 9, in which our warp combined with homography, equirectangular projection, and two types of similarity transformations are respectively tested. Fig. 9(a), (b), and (c) show the alignment results of global homography, SPHP, and AutoStitch, respectively, among which AutoStitch employs the equirectangular projection to form its global transformation and SPHP mixes global homography and the global similarity transformation. Visible artifacts (circled in red) appear in all three results using only global transformations. The result of our method in equirectangular projection is shown in Fig. 9(d), which contains no visible artifacts, such that it performs better than Fig. 9(c). As shown in Fig. 9(e) and (f), accurate alignment in perspective projection could be achieved by both APAP and our method for the current database. By applying the combination strategy presented in Section III-D, the warps in the overlapping region could be gradually changed into a previously estimated similarity transformation across the image. Fig. 9(h) shows the resulting panorama, which has the same shape as the result of SPHP+APAP presented in Fig. 9(g). If we restrict the rotation angle to zero while estimating the similarity transformation which means only scaling and offsetting remain, the result of our method would be comparable to ANAP, as shown in Fig. 9(i).



Fig. 9. Flexibility evaluation. (a) (b) (c) (e) (g) and (i) are respectively the results of global homography, SPHP [11], AutoStitch [4], APAP [7], SPHP+APAP and ANAP [12]. (d) (f) (h) and (j) are results of our warp combined with different global transformations.

and (j). The flexibility of our method ensures a wider scope of application than existing approaches.

C. Comparison of Computational Efficiency

The proposed warping is directly derived from the matching data and can be efficiently computed on the meshed stitching

field. Its computational complexity has been analyzed in detail in Section III-E. Here, we further evaluate its computational efficiency on different databases, and compare it with state-of-the-art local adaptive stitching methods, namely APAP [7], SPHP [11]+APAP, and ANAP [12]. All the tested methods are run in the same experimental settings with a 2.0G-Hz CPU and 16 GB RAM. For more comprehensive comparison, we

TABLE II
COMPARISON OF ELAPSED TIME USING CONSTANT CELL SIZE

database	image Size	matches	elapsed time (s)				
			APAP	accelerated APAP	SPHP+APAP	ANAP	robust ELA
<i>desk</i> [32]	500 × 375	245	1.129	2.182	1.234	1.334	0.683
<i>temple</i> [9]	730 × 487	482	5.569	4.012	5.254	4.443	0.421
<i>window</i> [32]	800 × 600	216	5.475	2.265	3.851	4.114	0.317
<i>cabinet</i> [32]	800 × 600	260	4.275	3.629	5.486	4.948	0.285
<i>corner</i> [32]	800 × 600	263	2.204	3.191	2.354	2.424	0.660
<i>chess girl</i> [8]	1008 × 755	742	6.807	5.410	8.991	6.987	0.458
<i>building</i> [12]	816 × 1088	839	16.470	6.623	13.910	13.254	0.661
<i>fence</i> [12]	816 × 1088	762	21.374	5.658	17.135	16.171	2.579
<i>tower</i>	960 × 1280	919	30.871	6.854	24.337	23.116	2.441
<i>wall</i>	1280 × 960	903	52.919	6.678	22.650	27.089	1.168
<i>footpath</i>	1280 × 960	1652	60.967	9.002	41.514	46.291	2.059
<i>cabin</i>	1280 × 960	1907	53.938	9.948	39.401	37.977	2.375
<i>campus square</i>	1280 × 960	1276	77.878	9.178	29.697	29.910	1.618
<i>gym</i>	1280 × 960	1936	51.300	11.262	33.321	32.733	1.810
<i>park</i> [11]	1920 × 1440	910	167.155	11.891	54.519	52.701	2.775
<i>campus</i> [11]	1920 × 1440	2050	392.298	19.173	112.875	109.968	5.093
<i>railtracks</i> [7]	2000 × 1500	3036	142.013	22.522	143.181	127.412	5.227
<i>work table</i>	2160 × 1440	2341	76.604	17.860	74.980	72.618	2.214
<i>racetracks</i>	2160 × 1440	3749	227.616	28.045	155.968	153.674	7.541
<i>theater</i>	2160 × 1440	3700	273.985	27.558	167.915	152.838	7.277

TABLE III
COMPARISON OF ELAPSED TIME USING CONSTANT CELL NUMBER

database	image Size	matches	elapsed time (s)				
			APAP	accelerated APAP	SPHP+APAP	ANAP	robust ELA
<i>desk</i> [32]	500 × 375	245	3.728	2.865	5.176	4.778	1.192
<i>temple</i> [9]	730 × 487	482	6.516	4.434	8.312	8.738	0.980
<i>window</i> [32]	800 × 600	216	4.363	2.779	5.697	6.521	0.934
<i>cabinet</i> [32]	800 × 600	260	4.813	2.850	6.615	6.170	0.893
<i>corner</i> [32]	800 × 600	263	3.832	2.745	5.595	5.109	1.434
<i>chess girl</i> [8]	1008 × 755	742	7.219	5.734	9.939	9.228	1.046
<i>building</i> [12]	816 × 1088	839	8.418	5.586	11.371	11.082	1.194
<i>fence</i> [12]	816 × 1088	762	8.851	6.033	11.739	12.121	3.256
<i>tower</i>	960 × 1280	919	9.958	6.478	10.693	12.731	3.648
<i>wall</i>	1280 × 960	903	11.391	5.976	12.948	12.401	1.408
<i>footpath</i>	1280 × 960	1652	15.176	8.901	16.430	17.562	2.092
<i>cabin</i>	1280 × 960	1907	15.059	9.700	18.699	19.050	2.399
<i>campus square</i>	1280 × 960	1276	16.133	9.466	17.088	17.260	1.663
<i>gym</i>	1280 × 960	1936	17.655	10.229	21.203	20.609	2.029
<i>park</i> [11]	1920 × 1440	910	19.689	10.502	18.152	17.681	2.607
<i>campus</i> [11]	1920 × 1440	2050	35.293	15.590	29.037	27.998	4.186
<i>railtracks</i> [7]	2000 × 1500	3036	27.695	16.695	35.523	35.214	4.484
<i>work table</i>	2160 × 1440	2341	19.742	14.005	24.599	25.423	2.451
<i>racetracks</i>	2160 × 1440	3749	30.951	22.857	40.233	41.610	7.010
<i>theater</i>	2160 × 1440	3700	32.942	22.429	42.585	40.469	6.763

use two different mesh resolution settings for the evaluation: 1) the constant cell size of 10×10 pixels, and 2) the constant cell number of 100×100 . The computational costs in terms of elapsed time under the two mesh settings are respectively presented in Table II and Table III. The timing starts right after the feature matching and ends before the image reprojection. For both settings, the time consumptions of our method are constantly much lower than other methods in all cases with different resolutions and matched point numbers. Noticing that the implementation provided by the authors of APAP employs several pre-compiled executable components to accelerate the computation, the elapsed time of the accelerated APAP are also

presented in Table II and Table III. It still costs more time than our method, which did not employ such accelerations. The experimental results are depicted by curves in Fig. 10, in which the databases are sorted by their resolutions in ascending order. It can be easily seen that our method outperforms all the compared approaches in computational efficiency.

D. Stitching Multiple Images

To demonstrate the performance of the proposed method with multiple images, Fig. 11 shows the stitching results of the *round-about* [12] database with four input images. Fig. 11(a) and (b)

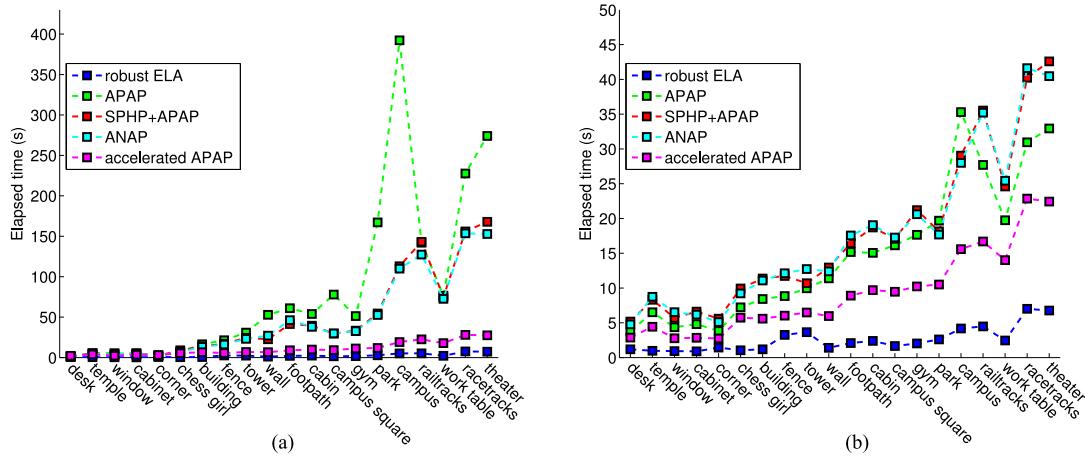
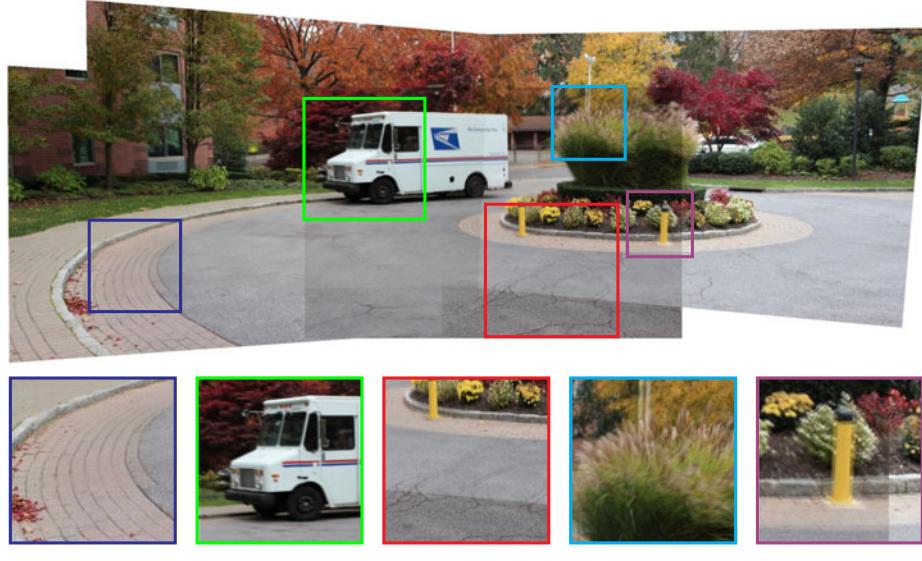
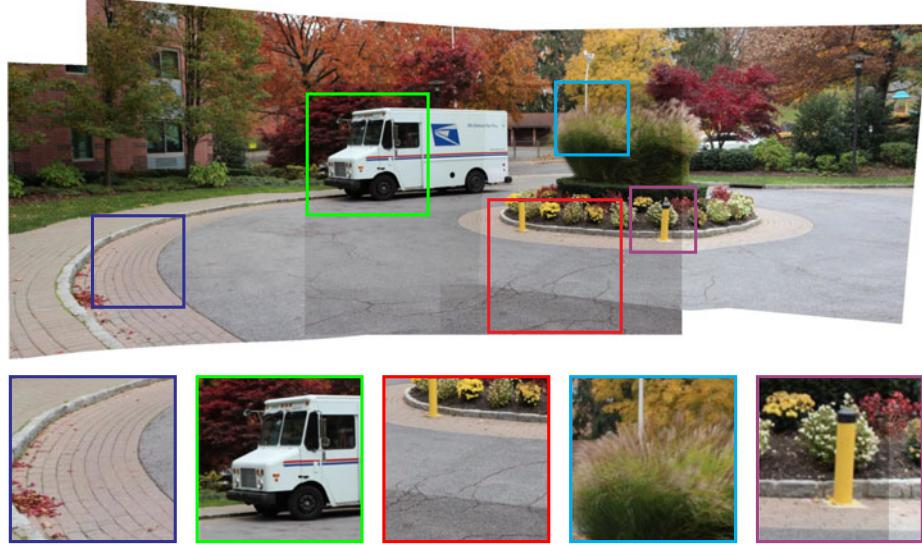


Fig. 10. Comparison of computational efficiency. (a) Results of the constant cell size 10×10 pixels. (b) Results of the constant cell number 100×100 .



(a)



(b)

Fig. 11. Stitching results of multiple images in perspective projection. The *roundabout* [12] database containing four input images is used to test our method. (a) Results of global alignment. (b) Results of our method.

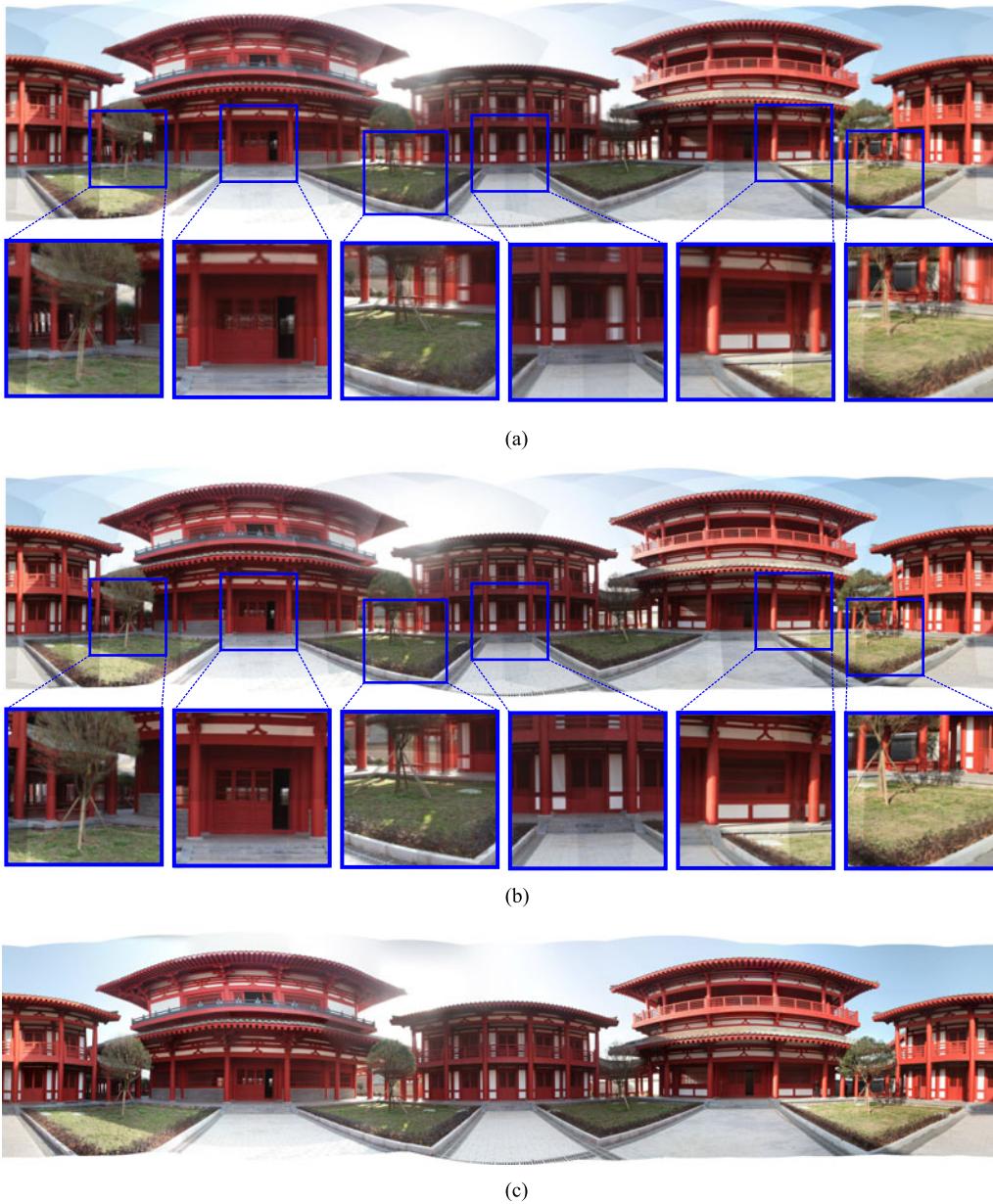


Fig. 12. Stitching results of multiple images in equirectangular projection. The panorama is constructed using a database captured by ourselves, which contains 24 input images of a 360° view palace scene. (a) Results of global alignment. (b) Results of our method. (c) Final blended panorama.



Fig. 13. Failure case. The input images of both cases are with severe occlusions. The artifacts in the resulting panoramas are circled in red.

are respectively generated using the global alignment and our method. Five representative areas are highlighted to show the alignment quality in detail. The structural misalignments of Fig. 11(a) are eliminated in Fig. 11(b) while preserving the global projectivity. Fig. 12 shows the results of another case containing 24 input images of a 360° view palace scene. There are no visible parallax errors in the resulting panorama using our method. Moreover, the panoramas are displayed in equirectangular projection in which the left and right borders are connected in space. Our method could ensure the consistency across the two borders which is seldom considered by other local adaptive methods. As the final step of a complete stitching process, the pyramid blending [44] is employed to smooth the transition between images. Fig. 12(c) presents the blended panorama.

E. Limitations and Failure Cases

The proposed elastic warping regards the scene as a continuous surface. Therefore, our method is incapable of aligning images with severe occlusions. Fig. 13 presents two failure cases in which ghosting effects appear on the foreground objects. Another limitation of our method is that the inliers should be the majority of the matching data to support the probabilistic feature refinement.

VI. CONCLUSION

An effective and efficient image stitching method based on robust elastic warping is proposed in this paper. First, an elastic warping model is constructed to eliminate the parallax errors. Without sacrificing the accuracy of alignment, a uniform grid mesh is set over the image plane to accelerate the computation of the deformation functions. The warping model is directly derived from the matching data and then applied to the original input images. Therefore, it can be interpreted as a correction of the global model. The panorama can be composed by directly re-projecting the warped images. Second, a Bayesian model of feature refinement is proposed to adaptively remove the local outliers from the matching data. This is an important complement of the proposed warping model that ensures a more robust alignment than existing approaches. Third, we provide the strategy of combining our warp with the global similarity transformation to benefit from its high flexibility. Finally, we provide the entire work flow of stitching multiple images using our method, as well as some important principles. The stitching quality, flexibility and computational efficiency of our method are carefully evaluated by several comparative experiments on a series of challenging cases. Our future work will concentrate on more robust occlusion handling and video stitching.

REFERENCES

- [1] V. R. Gaddam, M. Riegler, R. Eg, C. Griwodz, and P. Halvorsen, “Tiling in interactive panoramic video: Approaches and evaluation,” *IEEE Trans. Multimedia*, vol. 18, no. 9, pp. 1819–1831, Sep. 2016.
- [2] S. Kasahara, S. Nagai, and J. Rekimoto, “Jackin head: Immersive visual telepresence system with omnidirectional wearable camera,” *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 3, pp. 1222–1234, Mar. 2017.
- [3] R. Anderson *et al.*, “Jump: Virtual reality video,” *ACM Trans. Graph.*, vol. 35, no. 6, 2016, Art. no. 198.
- [4] M. Brown and D. G. Lowe, “Automatic panoramic image stitching using invariant features,” *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 59–73, 2007.
- [5] J. Kopf, M. Uyttendaele, O. Deussen, and M. F. Cohen, “Capturing and viewing gigapixel images,” *ACM Trans. Graph.*, vol. 26, no. 3, 2007, Art. no. 93.
- [6] Z. Zhu, E. M. Riseman, and A. R. Hanson, “Parallel-perspective stereo mosaics,” in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, 2001, vol. 1, pp. 345–352.
- [7] J. Zaragoza, T.-J. Chin, Q.-H. Tran, M. S. Brown, and D. Suter, “As-projective-as-possible image stitching with moving dlt,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1285–1298, Jul. 2014.
- [8] W.-Y. Lin, S. Liu, Y. Matsushita, T.-T. Ng, and L.-F. Cheong, “Smoothly varying affine stitching,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2011, pp. 345–352.
- [9] J. Gao, S. J. Kim, and M. S. Brown, “Constructing image panoramas using dual-homography warping,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2011, pp. 49–56.
- [10] Z. Lou and T. Gevers, “Image alignment by piecewise planar region matching,” *IEEE Trans. Multimedia*, vol. 16, no. 7, pp. 2052–2061, Nov. 2014.
- [11] C.-H. Chang, Y. Sato, and Y.-Y. Chuang, “Shape-preserving half-projective warps for image stitching,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2014, pp. 3254–3261.
- [12] C.-C. Lin, S. U. Pankanti, K. N. Ramamurthy, and A. Y. Aravkin, “Adaptive as-natural-as-possible image stitching,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2015, pp. 1155–1163.
- [13] F. Zhang and F. Liu, “Parallax-tolerant image stitching,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2014, pp. 3262–3269.
- [14] Y.-S. Chen and Y.-Y. Chuang, “Natural image stitching with the global similarity prior,” in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 186–201.
- [15] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, “Optimized scale-and-stretch for image resizing,” *ACM Trans. Graph.*, vol. 27, no. 5, 2008, Art. no. 118.
- [16] C.-H. Chang and Y.-Y. Chuang, “A line-structure-preserving approach to image resizing,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2012, pp. 1075–1082.
- [17] K. He, H. Chang, and J. Sun, “Rectangling panoramic images via warping,” *ACM Trans. Graph.*, vol. 32, no. 4, 2013, Art. no. 79.
- [18] K. Lin, N. Jiang, L.-F. Cheong, M. Do, and J. Lu, “Seagull: Seam-guided local alignment for parallax-tolerant image stitching,” in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 370–385.
- [19] S. Baker and I. Matthews, “Lucas-kanade 20 years on: A unifying framework,” *Int. J. Comput. Vis.*, vol. 56, no. 3, pp. 221–255, 2004.
- [20] C. Liu, J. Yuen, and A. Torralba, “Sift flow: Dense correspondence across scenes and its applications,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 978–994, May 2011.
- [21] F. Perazzi *et al.*, “Panoramic video from unstructured camera arrays,” *Comput. Graph. Forum*, vol. 34, no. 2, pp. 57–68, 2015.
- [22] J. Jia and C.-K. Tang, “Image stitching using structure deformation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 617–631, Apr. 2008.
- [23] J. Li *et al.*, “Efficient video stitching based on fast structure deformation,” *IEEE Trans. Cybern.*, vol. 45, no. 12, pp. 2707–2719, Dec. 2014.
- [24] F. L. Bookstein, “Principal warps: Thin-plate splines and the decomposition of deformations,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 567–585, Jun. 1989.
- [25] R. Sprengel, K. Rohr, and H. S. Stiehl, “Thin-plate spline approximation for image registration,” in *Proc. IEEE 18th Annu. Int. Conf. Eng. Med. Biol. Soc. Bridging Disciplines Biomed.*, 1996, vol. 3, pp. 1190–1191.
- [26] K. Rohr *et al.*, “Point-based elastic registration of medical image data using approximating thin-plate splines,” in *Visualization in Biomedical Computing*. Berlin, Germany: Springer-Verlag, 1996, pp. 297–306.
- [27] R. Szeliski, “Image alignment and stitching: A tutorial,” *Found. Trends Comput. Graph. Vis.*, vol. 2, no. 1, pp. 1–104, 2006.
- [28] J. Gao, Y. Li, T.-J. Chin, and M. S. Brown, “Seam-driven image stitching,” in *Proc. Eurographics*, 2013, pp. 45–48.
- [29] K. Lin, N. Jiang, S. Liu, L.-F. Cheong, and M. D. J. Lu, “Direct photometric alignment by mesh deformation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2017, pp. 2405–2413.
- [30] G. Zhang, Y. He, W. Chen, J. Jia, and H. Bao, “Multi-viewpoint panorama construction with wide-baseline images,” *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3099–3111, Jul. 2016.
- [31] K. Joo, N. Kim, T.-H. Oh, and I. S. Kweon, “Line meets as-projective-as-possible image stitching with moving dlt,” in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 1175–1179.

- [32] S. Li, L. Yuan, J. Sun, and L. Quan, "Dual-feature warping-based motion model estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 4283–4291.
- [33] 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.
- [34] Z. Zhang, "Parameter estimation techniques: A tutorial with application to conic fitting," *Image Vis. Comput.*, vol. 15, no. 1, pp. 59–76, 1997.
- [35] R. Bajcsy and S. Kovačić, "Multiresolution elastic matching," *Comput. Vis., Graph., Image Process.*, vol. 46, no. 1, pp. 1–21, 1989.
- [36] N. Arad and D. Reisfeld, "Image warping using few anchor points and radial functions," *Comput. Graph. Forum*, vol. 14, no. 1, pp. 35–46, 1995.
- [37] P. S. Heckbert, "Fundamentals of texture mapping and image warping," Master's thesis, Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, CA, USA, 1989.
- [38] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [39] F. Diego, J. Serrat, and A. M. López, "Joint spatio-temporal alignment of sequences," *IEEE Trans. Multimedia*, vol. 15, no. 6, pp. 1377–1387, Oct. 2013.
- [40] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—A modern synthesis," in *Proc. Int. Workshop Vis. Algorithms*, 1999, pp. 298–372.
- [41] M. O'Searcoid, *Metric Spaces*. Berlin, Germany: Springer-Verlag, 2006.
- [42] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [44] P. J. Burt and E. H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Trans. Graph.*, vol. 2, no. 4, pp. 217–236, 1983.



Jing Li received the B.S. degree from National University of Defense Technology, Changsha, China, in 2011. He is currently working toward the Ph.D. degree in the College of Information System and Management, National University of Defense Technology. His research interests include computer vision, image processing, and image analysis.



Zhengming Wang received the M.Sc. degree from Huazhong University of Science and Technology, Wuhan, China, in 1986 and the Ph.D. degree from National University of Defense Technology, Changsha, China, in 1998. He is currently a distinguished Professor with the College of Science, National University of Defense Technology. He authored three academic monographs and more than 50 papers in the field of data mining, system analysis and time series analysis.



Shiming Lai received the B.S. and Ph.D. degrees in system engineering from National University of Defense Technology, Changsha, China, in 2008 and 2014, respectively. He is currently a Lecturer with the Department of System Engineering, National University of Defense Technology. His research interests include computer vision, computational photography, imaging systems, and virtual reality.

Yongping Zhai received the M.Sc. degree in instrumentation science and technology, and the Ph.D. degrees in communication engineering, both from National University of Defense Technology, Changsha, China, in 2007 and 2013, respectively. He is currently a Postdoctoral Researcher with the Department of System Engineering, National University of Defense Technology. His current research interests include multimedia technology, virtual reality, and machine learning.

Maojun Zhang received the Ph.D. degree in systems engineering from the National University of Defense Technology, Changsha, China, in 1997. He is currently a Professor with the Department of System Engineering, National University of Defense Technology. His research interests include computer vision, information system engineering, system simulation, and virtual reality.