IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 30, 2019, accepted September 12, 2019, date of publication October 1, 2019, date of current version November 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2944852

As-Aligned-As-Possible Image Stitching Based on Deviation-Corrected Warping With Global Similarity Constraints

JIALIANG LI, PINQUN JIANG, SHUXIANG SONG, HAIYING XIA, AND MIN JIANG
College of Electronic Engineering, Guangxi Normal University, Guilin 541004, China

Corresponding author: Pinqun Jiang (pqjiang@mailbox.gxnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 11465004 and Grant 61762014, in part by the Science and Technology Project of Guangxi under Grant 2018JJA170089, in part by the Guilin Scientific Research and Technology Development Plan Project under Grant 20170113-4, and in part by the Innovation Project of Guangxi Graduate Education under Grant XYCSZ2019075.

ABSTRACT Handling local misalignment caused by the local warp remains a common and challenging task for image stitching. Moreover, the stitched image is prone to appearing ghosting due to the variations of the image viewpoint between images. To solve the problem of local misalignment, we propose a projection deviation-corrected local warping method with a global similarity constraint for image stitching. Recent warps prove that the warp of the local mesh guide image effectively improves the accuracy of image alignment. Geometric projection deviation is well used to accurately correct pixel offsets in image warping. To correct pixel offsets, we first remove the outliers from matching points by using the normal distribution model. The retained matches are more precise and can improve the accuracy of image alignment. Next, we use the local warping model combining local homography and global similarity for image warping. To further address the misalignment problem caused by local warping, we describe the local projection deviation of the local warping model by adopting a three-dimensional mesh interpolation model. Finally, the warped images are blended by a linear smoothing model. Experimental results show that our method outperforms the state-of-the-arts in alignment accuracy, and also provides better visual effects on challenging images.

INDEX TERMS Image stitching, global similarity, local warping, image alignment, parallax.

I. INTRODUCTION

Image stitching, a method of combining multiple images into a wide-angle panorama containing information of original images [1], is the most widely used algorithms in computer vision, such as panorama, video surveillance [2], and virtual reality [3]. The natural stitching of parallax images remains a challenging task [4]. Conventionally, the first stage in image stitching is to determine the warp relationship of each image and transform it into a common coordinate system [5]-[6]. Then, the warped images are blended into a natural panorama by linear weight.

Early methods focus on global alignment of images. AutoStitch is one of the representative methods [7]. A global homography transformation is only correct under special conditions. For example, it will result in artifacts like ghosting or broken image structures in a roughly coplanar scene. To address the inadequacy of global warps, several local warp models have been proposed, such as the smoothly varying affine (SVA) [8] and the as-projective-as-possible (APAP) [9]. These methods calculate multiple local warps to achieve better alignment accuracy, but they only work for images with moderate parallax.

Recently, several approaches attempt to define spatially-varying warping parameters to alleviate the distortion and the restricted view when aligning the input images. The shape-preserving half-projective (SPHP) warp [10] and the adaptive as-natural-as-possible (ANAP) warp [11] improve the perspective in the non-overlapping region by using a combination of local homography and global similarity transformations. Moreover, the content-preserving warping (CPW) method is also used to overcome the shortcomings of methods

The associate editor coordinating the review of this manuscript and approving it for publication was Guitao Cao.

based on global homography [12]. A series of direct warping approaches are proposed to obtain a natural stitching image [13]–[16]. These methods directly apply geometric constraints to guide mesh deformation in the localized image and can be combined with seam cutting to cope with large parallax images. However, it is difficult to optimize multiple constraint terms simultaneously.

Besides the models described above, the non-rigid warping method is also applicable to image warping by directly calculating the deformation function on the image plane. Bookstein [17] introduced the thin-plate spline (TPS) for elastic warping of medical images. TPS is the 2D generalization of the cubic spline that decomposes the deformation between images into affine and non-affine components to align images. The warp has been widely used in medical image registration [18]-[20]. These methods adopt the TPS to warp the image so that it falls on the fixed anchor point with a smooth surface to achieve image alignment. The warp is extremely sensitive to matching errors, i.e. severe distortion caused by the presence of outliers. Moreover, little research has focused on the removal of outliers in matching points after image alignment [4].

To deal with these problems discussed above, an image stitching method combining several techniques is proposed in this paper. Our work is inspired by Zaragoza's local projection [11] and Li's elastic warping [4] ideas that the local homography transformation and projection correction can effectively improve the ability of image alignment. Our method is built upon an observation that the misalignment caused by projection deviation will still occur after local warping. The 3D smooth surface, a deformation field interpolated by the deviation between the feature point and the projected point, is the key to eliminate projection correction and align the image. A feature point refinement model with local outliers removal, is proposed to calculate the smooth surfaces. In addition, we present a warping method based on local homography and global similarity, to eliminate warping deviation and further optimize the image alignment. The model reduces visual artifacts caused by misalignment that are difficult to be handled by traditional warp, such as APAP [9], ANAP [11], GSP [15], and REW [4].

The remainder of this paper is organized as follows. Section II briefly discusses recent related works. Section III describes our approach of projection deviation correction based on local warping in detail. Experimental results are shown and discussed in Section IV. Finally, a conclusion of our work is drawn in Section V.

II. RELATED WORK

Image stitching has been well studied, and more concepts about it can be found in [1]. The global homography worked well [7] under the assumption that the input image was roughly coplanar. Gao *et al.* [21] proposed the dual-homography warp (DHW) to address scenes with two dominant planes by a linear combination of two homography matrices. Since a few homography transformations cannot account for parallax, these methods were difficult to handle more complicated situations.

Lin *et al.* [8] proposed the smoothly varying affine (SVA) field while allowing local deformations to handle parallax flexibly. The as-projective-as-possible (APAP) warp, which used a moving direct linear transformation (MDLT) to assign global homography to each mesh for better local alignment, was proposed by Zaragoza *et al.* [9]. Multiple local warping models were adopted by [8] and [9] for better alignment accuracy.

After obtaining better alignment, several methods were more concerned with addressing the distortion problem in the stitched image. Chang *et al.* [10] proposed the shape-preserving half-projective (SPHP) warp to smoothly turn homography into global similarity, which maintained good alignment in overlapping region and protected viewing angles in non-overlapping region without distortion. The method used global homography to derive a global similarity, which caused unnatural rotation to occur. Lin *et al.* [11] proposed the adaptive as-natural-as-possible (ANAP) warp based on linearity homography and combined with global similar transformation to solve the unnatural rotation of SPHP by minimizing the rotation angle. However, the problem of image misalignment still exists.

Other methods model the warp as mesh deformation by energy minimization, which achieves natural visual effect by applying geometric constraints to guide the mesh to deforming. Lin *et al.* [14] proposed a seam-guided local alignment (SEAGULL) method [18] based on structure-preserving warping, which effectively preserves the line structure during warping. The global-similarity-prior (GSP) warp based on APAP pre-matching is proposed by Chen and Chuang [15]. The warp of the image is constrained to a global similarity transform by GSP. Zhou *et al.* [20] proposed a vector-field interpolation method for non-rigid image deformation, which calculates the spatial transformation of each pixel by the TPS function. Li *et al.* [4] proposed a parallax-tolerant image stitching method based on global homography, which uses the TPS function to calculate the global deformation vector to alleviate the misalignment of global homography in the overlapping region.

III. THE PROPOSED APPROACH

In this section, we detail our proposed approach. Fig. 1 illustrates all stages of the proposed approach. The general process of image stitching involves feature extraction, outliers culling, image warping and image blending. We mainly perform error elimination processing at the stage of outliers culling and image warping to improve the accuracy of the image alignment.

Firstly, the outliers are effectively identified by our feature point refinement model. This process translates geometric errors into statistically probabilistic events that effectively avoids the limitations of a fixed threshold that is not flexible enough. After obtaining the inliers, the local homography and global similarity are calculated on the meshed image.

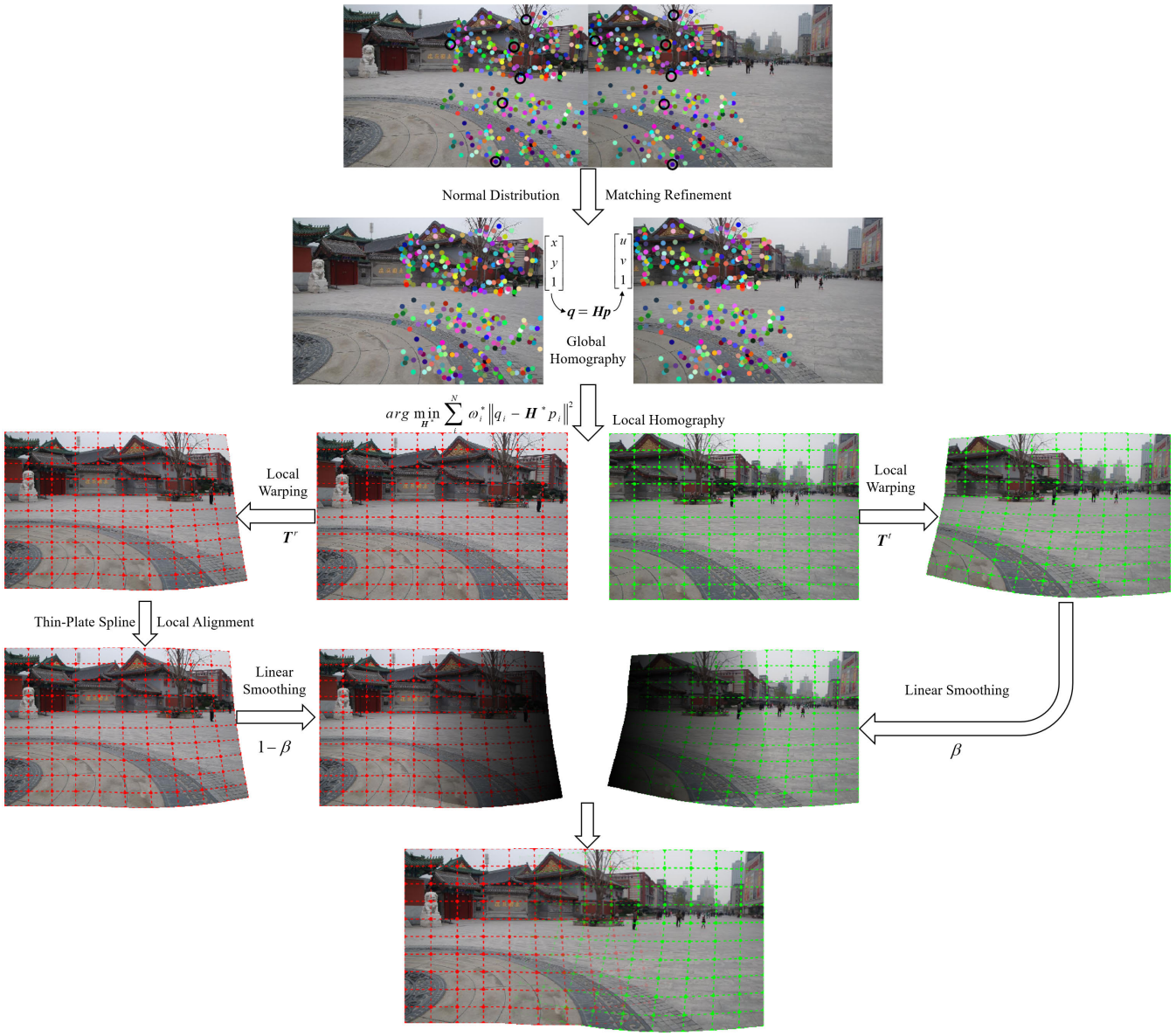


FIGURE 1. Flowchart of the proposed image stitching method.

secondly, the method based on TPS local deviation correction is applied to the warp of the image to align the image as much as possible. Finally, a linear-based pixel smoothing model is used for the blending of warped images to eliminate ghosting and color difference seams.

A. REFINEMENT OF FEATURE MATCHES

The image stitching method based on feature points, uses the scale invariant feature transform (SIFT) [22] combined with a random sample consensus (RANSAC) [23] to filter feature points. Although, the matching strategy performs well, some imperceptible wrong matches are inevitable. Reference [16] uses a fixed threshold to remove outliers. And reference [4] removes some mismatches by using the coefficient distribution of TPS as a criterion for rejecting outliers. As can be seen from the mark of Fig. 2, the methods remove the wrong point and also deletes some correct points that should be retained. In particular, these points are precious for low-texture images. As assumed in reference [4], the projection deviation of the matched point obeys the normal distribution. We directly predict the correctness of the matches by judging whether the projection deviation obeys the normal distribution or not.

Given the reference image I1 and the target image I2 with overlapping region. The points $\boldsymbol{p} = [x, y, 1]^{\mathrm{T}}$ and $\boldsymbol{q} = [u, v, 1]^{\mathrm{T}}$ are a pair of matching points of the region. The homography $\boldsymbol{H}$ is estimated by DLT from the matching points. $\boldsymbol{p}' = [x', y', 1]^{\mathrm{T}}$ is the projective point of $\boldsymbol{p}$ in I2

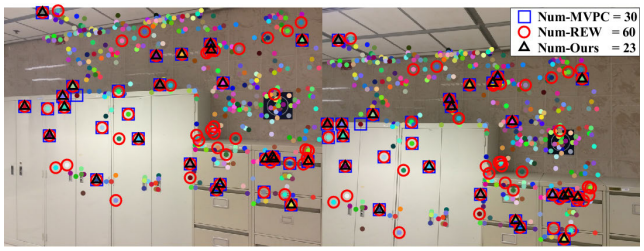


FIGURE 2. Comparison of refinement results of feature matching. Feature points are represented as filled points. The blue box is the 30 error points marked by MVPC [16], and the red circle is the 60 error points of the REW [4] mark, which contain some correct points. Twenty-three error matches are marked by black triangles in our method, which retains more correct matches.

estimated by MDLT [9]. The projection deviations of the matching point is projected into the x and y directions. The projection deviation $\boldsymbol{PD} = \boldsymbol{p}' - \boldsymbol{q} = [Dx,\ Dy]^{\mathrm{T}}$, the mean (μ_x, μ_y) and the standard deviation (σ_x, σ_y) of projection deviation Dx and Dy are calculated, respectively.

Given a pair of matching points $\{\boldsymbol{p}_i,\ \boldsymbol{q}_i\}$, the projection deviation Dx_i and Dy_i can be calculated subsequently.

Let the event $Ax_i = \{|Dx_i - \mu_x| < n\sigma_x\}$ and $Ay_i = \{|Dy_i - \mu_y| < n\sigma_y\}$ represent the thresholds of the projection deviations in the x and y directions, respectively. The probability of Ax_i and Ay_i conforms to

$$\begin{aligned} p(Ax_i) &= p\{|Dx_i - \mu_x| < n\sigma_x\} \\ &= p\{-n < \frac{Dx_i - \mu_x}{\sigma_x} < n\} = 2\Phi(n) - 1 \end{aligned} \tag{1}$$

$$\begin{aligned} p(Ay_i) &= p\{\left|Dy_i - \mu_y\right| < n\sigma_y\} \\ &= p\{-n < \frac{Dy_i - \mu_y}{\sigma_y} < n\} = 2\Phi(n) - 1 \end{aligned} \tag{2}$$

where $\Phi(\cdot)$ is the standard normal distribution function.

According to the three-sigma rule, if the projection deviation Dx_i or Dy_i of feature point is outside the interval $(\mu\text{-}n\sigma, \mu + n\sigma)$ and $n > 2$, then it is called a small probability event with a probability less than 5%. Generally, set $n = 3$, the probability of this event is only 0.27%. We regard $\{\boldsymbol{p}_i, \boldsymbol{q}_i\}$ as an outlier and remove it from the matching set. The matching result based on the normal distribution is shown in the black triangle of Fig. 2.

B. LOCAL WARPING

A homography $\boldsymbol{H}$ is estimated by solving the linear equation of $\boldsymbol{p} = [x, y, 1]^{\mathrm{T}}$ to $\boldsymbol{q} = [u, v, 1]^{\mathrm{T}}$, $\boldsymbol{q} = \boldsymbol{Hp}$, as shown in (3).

$$\begin{bmatrix} \mathrm{u} \\ v \\ 1 \end{bmatrix} \sim \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3}$$

where $\sim$ denotes equality up to a scale factor, $\boldsymbol{H} \in \mathbb{R}^{3\times3}$.

Taking the cross product on both sides of formula (3), the homology transformation can be rewritten as $\mathbf{0}_{3\times1} = \boldsymbol{q} \times \boldsymbol{Hp}$. The homogeneous equation is as follows

$$\mathbf{0}_{3\times1} = \begin{bmatrix} \mathbf{0}_{1\times3} & -\boldsymbol{p}^{\mathrm{T}} & v\boldsymbol{p}^{\mathrm{T}} \\ \boldsymbol{p}^{\mathrm{T}} & \mathbf{0}_{1\times3} & -u\boldsymbol{p}^{\mathrm{T}} \\ -v\boldsymbol{p}^{\mathrm{T}} & u\boldsymbol{p}^{\mathrm{T}} & \mathbf{0}_{1\times3} \end{bmatrix} \boldsymbol{h} = \boldsymbol{Ah} \tag{4}$$

where $\boldsymbol{h} = [\,h_{11}\ h_{12}\ \ldots\ h_{33}\,]^{\mathrm{T}}$. $\boldsymbol{A} = [\,\boldsymbol{a}_1\ \boldsymbol{a}_2\ \ldots\ \boldsymbol{a}_{\mathrm{N}}\,]^{\mathrm{T}}$, Let $\boldsymbol{a}_i \in \mathbb{R}^{2\times9}$ be the first-two rows of (4), $i = 1, 2, \ldots, \mathrm{N}$.

The matrix $\boldsymbol{h}$ in the formula (4) can be estimated by the linear equation called DLT as shown in the following.

$$\boldsymbol{h} = \arg\min_{\boldsymbol{h}} \sum_{i=1}^{\mathrm{N}} \|\boldsymbol{a}_i\boldsymbol{h}\|^2 = \arg\min_{\boldsymbol{h}} \|\boldsymbol{Ah}\|^2,\ ||\boldsymbol{h}||^2 = 1 \tag{5}$$

The global homography $\boldsymbol{H}$ reshapes from $\boldsymbol{h}$. Inspired by APAP, images are aligned by weighted global homography.

Image I1 is divided into a grid of C1×C2 cells, and the coordinates of the grid vertices are $p_* = [x_*\ y_*\ 1]^{\mathrm{T}}$ in homogeneous coordinate system. The local homography at arbitrary position p_* is estimated by the MDLT framework [9] under the following rules:

$$\boldsymbol{h}_* = \arg\min_{\boldsymbol{h}} \sum_{i=1}^{\mathrm{N}} \left\| \omega_*^i \boldsymbol{a}_i \boldsymbol{h} \right\|^2 = \arg\min_{\boldsymbol{h}} \|W_*\boldsymbol{Ah}\|^2 \tag{6}$$

where $\omega_*^i = \max(\exp(-\|p_* - p_i\|^2/\delta^2), \gamma)$, δ is a scale parameter and $\gamma \in [0,\ 1]$, $\boldsymbol{W}_* = \mathrm{diag}([\omega_*^1\ \omega_*^1\ \omega_*^2\ \omega_*^2\ \ldots\ \omega_*^{\mathrm{N}}\ \omega_*^{\mathrm{N}}])$.

Equation (6) can be solved by the smallest right singular vector of $\boldsymbol{W}_*\boldsymbol{A}$.

After local homography is obtained, the image I1 can be projected onto the image I2 to achieve alignment of the image. However, this will result in the distortion of the image in the non-overlapping region. As the measures taken by the ANAP, a global similar transformation is introduced for addressing the distortion to preserve the viewing angle.

The optimal global similarity $\boldsymbol{S}_{\mathrm{g}}$ is also calculated by DLT. The homography and similarity are linearly combined into a projection matrix as shown in (7).

$$\boldsymbol{T}_{i,j}^t = \alpha \boldsymbol{H}_{i,j}^{mdlt} + (1 - \alpha)\boldsymbol{S}_{\mathrm{g}} \tag{7}$$

$$\boldsymbol{T}_{i,j}^r = \boldsymbol{T}_{i,j}^t (\boldsymbol{H}_{i,j}^{mdlt})^{-1} \tag{8}$$

where $\boldsymbol{H}_{i,j}^{mdlt}$ is local homography at (i, j), $\boldsymbol{S}_{\mathrm{g}}$ is the global similarity transformation, and α linearly varies from 1 to 0 in ANAP [11]. $\boldsymbol{T}_{i,j}^t$ is the transformation matrix of the target image, and $\boldsymbol{T}_{i,j}^r$ is the transformation matrix of the reference image.

The warped coordinates (i', j') of an arbitrary position (i, j) in the image are calculated by the corresponding local transformation matrix. Specifically, the position (i, j) of the image I1 is projected onto the corresponding warped mesh $\boldsymbol{W}1$ by equation $[i', j', 1]^{\mathrm{T}} = \boldsymbol{T}_{i,j}^r [i, j, 1]^{\mathrm{T}}$. Similarly, the warped mesh $\boldsymbol{W}2$ of the image I2 is calculated by $\boldsymbol{W}_2(i, j) = \boldsymbol{T}_{i,j}^t \mathbf{I}_2(i, j)$.

C. LOCAL ALIGNMENT

Radial basis functions (RBFs) are powerful tools for image alignment [24]. A variety of RBFs functions are designed to better align images. The spline tool TPS was used earlier for aligning images [17]. The TPS can be decomposed into global affine transformation and local bending function respectively controlled by global affine matrix and local non-rigid warping function [20]. TPS can perform a physical deformation with minimal bending energy to create a smooth surface containing all given anchor points. Bookstein [17] defined the energy function $E(f)$ based on the space integral of the second order derivatives of the mapping function, which expresses the relationship between the physical bending energy of the thin metal plate and the point constraint.

$$E(f) = \iint_{R^2} \left(\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right) dxdy \quad (9)$$

The TPS satisfies these conditions, which minimizes the bending energy function and contains all control points. Given N control points $(\widetilde{x}_i, \widetilde{y}_i)$ and corresponding function values v_i in a plane. The form of TPS function is as shown in equation (10).

$$f(x, y) = a_1 + a_2 x + a_3 y + \sum_{i=1}^{N} \omega_i U \left(\|(\widetilde{x}_i, \widetilde{y}_i) - (x, y)\| \right) \quad (10)$$

where $[a_1\ a_2\ a_3]$ and ω_i are parameters. $U(r) = r^2 \ln r$ is the RBF. Note that the spline $f\ (x, y)$ has a squared indefinite second derivative, so the constraint of the function is

$$\sum_{i=1}^{N} \omega_i = 0 \text{ and } \sum_{i=1}^{N} \omega_i x_i = \sum_{i=1}^{N} \omega_i y_i = 0.$$

The parameters of TPS can be calculated by solving the linear system (11).

$$\begin{bmatrix} \boldsymbol{K} + \lambda \boldsymbol{E} & \boldsymbol{Q} \\ \boldsymbol{Q}^{\mathrm{T}} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{W} \\ \boldsymbol{A} \end{bmatrix} = \begin{bmatrix} \boldsymbol{V} \\ \boldsymbol{0} \end{bmatrix} \quad (11)$$

where $K_{i,j} = U\left(\left\| (\widetilde{x}_i, \widetilde{y}_i) - (\widetilde{x}_j, \widetilde{y}_j) \right\| \right)$. The i-th row of $\boldsymbol{Q}$ is $(1, \widetilde{x}_i, \widetilde{y}_i)$. $\boldsymbol{A} = [a_1\ a_2\ a_3]^T$ and $W = [\,\omega_1\ \omega_2\ \ldots\ \omega_{\mathrm{N}}\,]^T$. v_i is the element of $\boldsymbol{V}$. $\boldsymbol{E}$ is the N-order identity matrix. λ is the regular parameter. Specifically, Bookstein proposes to map each point $(\widetilde{x}_i, \widetilde{y}_i)$ to its corresponding $(\widetilde{x}'_i, \widetilde{y}'_i)$ using $f(x, y) = [f_x(x, y), f_y(x, y)]$ and $v_i = [\widetilde{x}'_i, \widetilde{y}'_i]^{\mathrm{T}}$.

TPS works well for interpolation of 3D surfaces, which can interpolate pixels of an image to a specific location by a given common anchor to align the image. We use TPS to calculate the deviation at an arbitrary position in the overlap region. It is not directly using TPS to warp an image to another image.

We substitute the projection deviation $\boldsymbol{PD}$ into equation (11) to calculate the corresponding parameters by $v_i = [Dx_i, Dy_i]^{\mathrm{T}}$. The projection deviation is calculated from local homography rather than global homography. We then use (10) to obtain the deviation function at (x, y). Specifically, the deviation function $f^{\mathrm{x}}(x, y)$ describes the projection deviation in the x direction, and $f^{\mathrm{y}}(x, y)$ represents the projection deviation in the y direction. The alignment item $f_{\mathrm{A}}(x, y) = [f^{\mathrm{x}}(x, y), f^{\mathrm{y}}(x, y)]$ consists of the deviation functions $f^{\mathrm{x}}(x, y)$ and $f^{\mathrm{y}}(x, y)$. In particular, the pixel alignment matrix $\boldsymbol{F}_{\mathrm{A}}$ is calculated by $\boldsymbol{F}_{\mathrm{A}}(i, j) = f_{\mathrm{A}}(i, j)$ at the arbitrary position (i, j).

Given the index $[\boldsymbol{M}^{\mathrm{x}}, \boldsymbol{M}^{\mathrm{y}}]$ of the warped mesh in the x and y directions. The projection deviation of the mesh $\boldsymbol{W}1$ is eliminated by subtracting the alignment matrix $\boldsymbol{F}_{\mathrm{A}}$ from the matrix $[\boldsymbol{M}^{\mathrm{x}}, \boldsymbol{M}^{\mathrm{y}}]$ as shown in equation (12).

$$\widehat{\boldsymbol{W}1} = [\boldsymbol{M}^{\mathrm{x}} - \boldsymbol{F}_{\mathrm{A}}^{\mathrm{x}}, \boldsymbol{M}^{\mathrm{y}} - \boldsymbol{F}_{\mathrm{A}}^{\mathrm{y}}] \quad (12)$$

where the $\widehat{\boldsymbol{W}1}$ is the index of aligned pixel. $\boldsymbol{F}_{\mathrm{A}}^{\mathrm{x}}$ and $\boldsymbol{F}_{\mathrm{A}}^{\mathrm{y}}$ are the deviation matrices of the image. The superscript x refers to the x direction and the superscript y indicates the y direction. Note that equation (12) is the TPS-based local deviation correction relation. And its role is to locally adjust the image to achieve local alignment.

The warped images I^{r} and I^{t} calculated from the images I1 and I2 are obtained by backward interpolation of the $\widehat{\boldsymbol{W}1}$ and $\boldsymbol{W}2$ by the bilinear interpolation function. The warped images I^{r} and I^{t} are linearly fused by the method provided by ANAP [11]. The weight of I^{t} is calculated by equation (13) in the overlapping area of images I^{r} and I^{t}. Specifically, the non-overlapping region weights are 1 in the image I^{t}. The sum of the weights of the images I^{r} and I^{t} at (i, j) is 1, (i, j) in the overlapping area of the images I^{r} and I^{t}.

$$\beta_{i,j} = \frac{L_{i,j} - (L_{\min} + 0.001)}{(L_{\max} - 0.001) - (L_{\min} + 0.001)} \quad (13)$$

where the $p_{i,j}$ is the coordinate of the overlapping area of the warped images. p_r and p_t are the center points of the warped image of the reference image and the target image, respectively. $L_{i,j} = \overrightarrow{p_r p_{i,j}} \cdot \overrightarrow{p_r p_t}$ is an element of L, which linearly maps each 2-dimensional point into a 1-dimensional projection point in the $\overrightarrow{p_r p_t}$ direction. $L_{\min}$ and $L_{\max}$ are the minimum and maximum values of matrix L, respectively.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we first introduce the relevant datasets and experimental settings, and then conduct the comparison analysis of performance in different stages of image stitching including matching point alignment, and stitching image quality. Finally, we also compared the time consumption of the algorithms, including APAP, ANAP, GSP, and REW.

A. DATASETS AND EXPERIMENTAL SETUP

A series of challenging images from various data sets were tested to evaluate the performance of the proposed method. 15 sets of images (1 from [9], 1 from [21], 1 from [15], 2 from [12], 2 from [10], 3 from [4], 4 from [14], and 1 collected by ourselves) listed in this paper consist of indoor, building, street, garden, lake, and other scenes. More comparisons can be found in supplementary material. The compared methods include APAP [9], ANAP [11], GSP [15], and REW [4]. The matching of feature points is done

TABLE 1. Accumulated projection deviation in the x and y directions.

Database	APDiTxD				APDiTyD			
	ANAP [11]	MVPC [16]	REW [4]	Ours	ANAP [11]	MVPC [16]	REW [4]	Ours
building [11]	886.24	406.50	600.07	394.90	1715.71	682.40	1334.06	652.95
castle [14]	1673.43	498.16	1362.97	481.11	561.08	237.05	477.51	219.65
doll [12]	12989.61	4931.06	10277.83	4853.11	2115.57	769.70	1177.39	735.17
forest [13]	134.84	82.07	77.03	62.45	63.36	48.57	42.45	39.39
garden [9]	8447.63	2681.92	4000.85	2387.74	3804.95	998.78	2113.21	956.81
garden [10]	3915.14	1205.58	1926.83	1203.15	5575.66	1112.92	3296.93	1040.72
temple [21]	444.71	177.26	353.45	162.46	290.05	105.73	224.18	85.68
Theater [4]	31984.10	16706.95	24525.03	11553.28	5352.10	1953.96	3513.98	1855.03
street [12]	944.99	413.50	498.21	383.34	794.63	358.75	527.86	319.12
Uffizi [15]	18970.43	8736.26	13467.11	8580.94	9871.19	3972.63	5297.13	3959.86

by the library VLFeat [25] and the RANSAC algorithm. APAP, ANAP, REW and our methods are implemented on MATLAB, while GSP is implemented on C++. The parameters are consistent with the settings of other methods. We use the source code provided by the authors to obtain the comparison results. The parameters in our paper are: $n = 3$, $C1 = C2 = 100$, $\delta = 8.5$, $\gamma = 0.1$, λ is the mean of the projection deviation. All the tested methods run in the same experimental settings with an Intel i5 CPU @ 3.8G-Hz CPU and 8 GB RAM.

B. MATCH POINT ALIGNMENT PERFORMANCE

The accumulated projection deviation of the matching points is the most intuitive embodiment of the alignment accuracy. We define APDiTxD as the accumulated projection deviation in the x directions, and define APDiTyD to represent the accumulated projection deviation in the y direction.

Table 1 depicts the accumulated projection deviation (in pixels) of the different methods on 10 image pairs with a large parallax. Obviously, the results of APDiTxD and APDiTyD obtained by our proposed method are always smaller than the results achieved by [11], [16], and [4]. For example, in the image *castle*, it is obvious that the projection deviation of method [16] is the highest, and our cumulative projection deviation increases by 35.3 % and 46.0 % in the x and y directions compared with [4]. In average, our deviation is slightly better than [16]. That is because ANAP uses traditional matching strategies, which inevitably retain some outliers. For the ANAP-based MVPC and REW, MVPC uses a constant threshold to eliminate outliers, while REW uses the bayesian model to remove outliers. Unfortunately, MVPC and REW inevitably eliminate some correct inliers. It is clear that the geometric projection deviation can effectively distinguish between inliers and outliers. In order to accurately and adaptively remove outliers, we use local homography to obtain projection deviation and introduce the normal distribution to describe the probability of the deviation, which accurately retains more inliers.

C. COMPARISON OF STITCHING QUALITY

The visual effects of stitched images and objective quality assessment indicators are often used to measure the performance of image stitching methods such as APAP, ANAP, GSP, and REW. Furthermore, APAP [9] is a representative method for processing parallax images by local homography. AANA [11] combines local homography with optimal global similarity to enhance the naturalness of the image. In addition, GSP [15] and REW [4] use global similarity prior or global deformation to constraint image warping. Therefore, the proposed method is compared with the methods, including APAP, ANAP, GSP, and REW.

1) SUBJECTIVE EVALUATION

Fig. 3 shows three examples of large parallax image stitching obtained by different methods, including APAP, ANAP, GSP, and REW. The results of different methods are displayed in rows. And the image datasets for the left, middle, and right columns are *lake* (collected by ourselves), *building* [14], and *garden* [10]. For the comparison of stitching visual effects, we use green boxes to mark unnatural scenes such as distortion and misalignment. Moreover, the ghost area is highlighted by red and blue boxes.

Note that there are severe misalignment and heavy shape distortions in the APAP results. The ANAP warp is also subject to local misalignment and ghosting, especially the dislocation of trees in the building. However, the GSP warp under various constraints only suffers from the problem of ghosting, which has a good visual effect. Unfortunately, the results of REW warp based on global homography still has serious misalignment. The result of our method is shown in Fig. 3(e). Unlike previous methods, the proposed method maintains alignment accuracy and shows good visual effects without visible parallax errors and perspective distortions.

The results in Fig. 3 show that the APAP-based local homography warp model can effectively cope with some parallax scenarios, but introduces perspective distortion. As the result of ANAP warp, the combination of local homography and similarity can coordinate the image perspective while preserving the alignment accuracy. However, the ANAP warp still inherits the errors involved in local warping so that misalignment occurs. Moreover, Local alignment and similarity priors help GSP get more natural results. Similarly, the GPS warp ignores the effect of local alignment. The REW warp also achieves better results by deforming constrained global

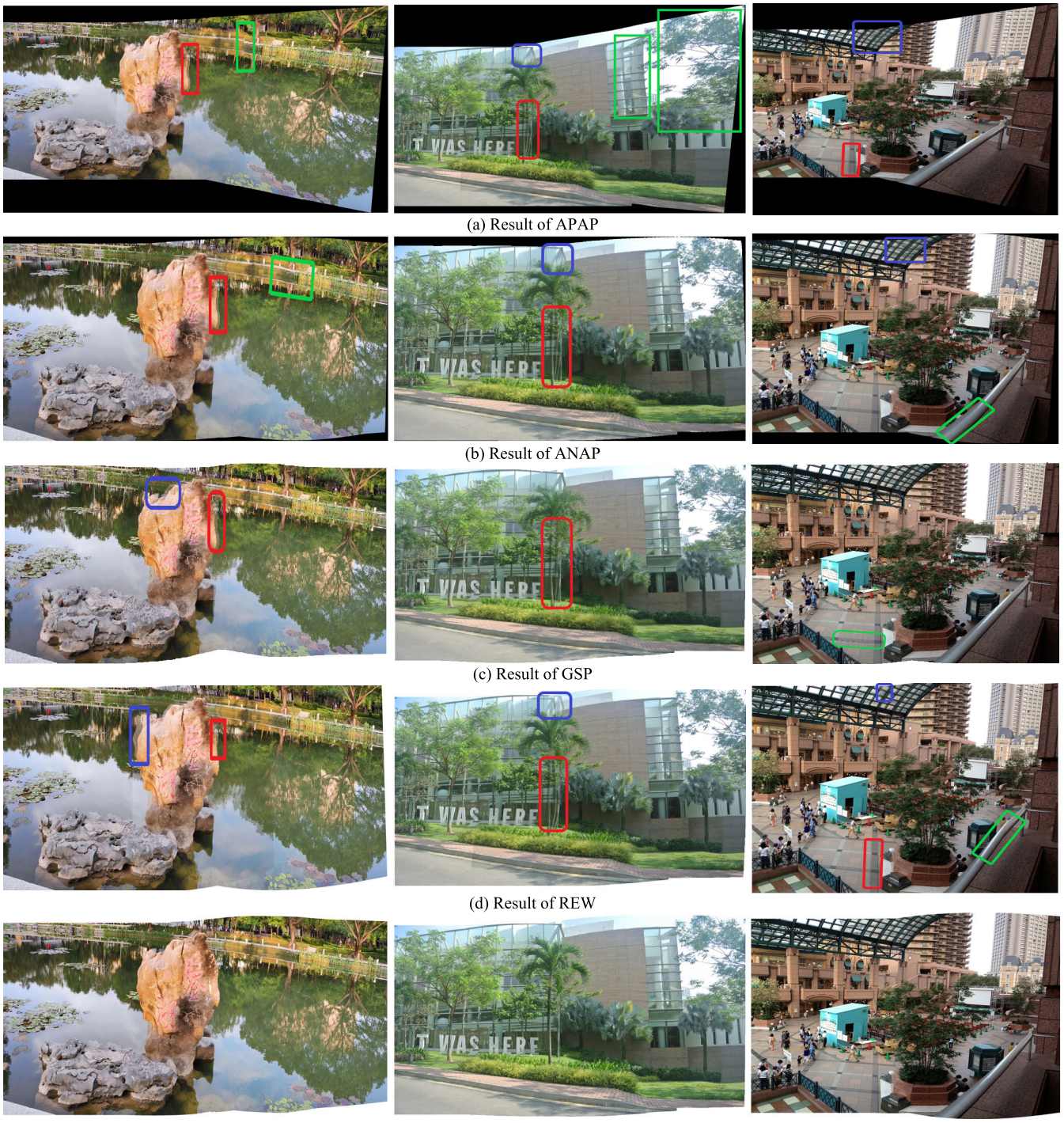


FIGURE 3. **Comparison of stitching results by the APAP [9], ANAP [11], GSP [15], REW [4], and our approach. From left to right, *"lake"*, *"building"* [14], and *"garden"* [10].**

homography, which lacks the application of local warping models. Therefore, we use the normal distribution to remove the outliers and calculate the correction field of the projection bias by the 3-dimensional interpolation function to improve the image alignment accuracy as much as possible.

2) OBJECTIVE EVALUATION

The Naturalness Image Quality Evaluator (NIQE), which calculates non-reference image quality scores for images to be evaluated, was proposed by Mittal *et al.* [26]. The NIQE is close to the human visual system, and the score is inversely

TABLE 2. Comparison of the time consumed of different methods.

Database	Image size(pixel)	Elapsed time (s)				
		APAP [9]	ANAP [11]	REW [4]	Ours	GSP [15]
Pisa [15]	2448 × 3264	61.63	72.39	65.91	68.79	2.8
desk [13]	500 × 375	1.05	2.69	0.81	1.78	0.70
doll [12]	1122 × 1110	7.04	14.43	6.31	8.74	2.60
building [12]	800 ×530	1.87	5.18	1.95	2.99	1.96
building [14]	1000 × 669	4.26	9.71	2.65	5.97	3.06
garden [10]	1440 × 1920	17.12	24.74	16.73	20.34	2.82
lake	4000 × 1844	56.40	74.53	58.92	65.23	2.88
wall [4]	1280 × 960	4.92	11.03	3.60	5.91	2.23
Paint	1600 × 1080	11.02	19.06	7.37	12.28	2.41
road	1080 × 1440	6.12	11.29	6.84	7.30	2.43

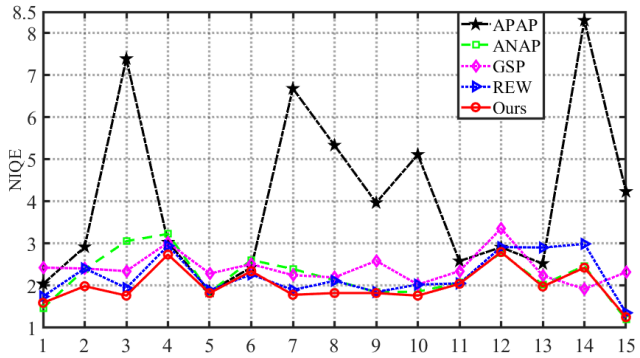


FIGURE 4. Comparison of NIQE scores for different methods in different image sets. The black, green, magenta, blue, and red lines represent the evaluation scores of the APAP, ANAP, GSP, REW, and the proposed method, respectively.

proportional to image quality. Lower NIQE score means the stitched images could be better.

Fig. 4 shows the comparison results of NIQE in 15 sets of images (1 from [9], 1 from [21], 1 from [15], 2 from [12], 2 from [10], 3 from [4], 4 from [14], and 1 collected by ourselves), where in most of the cases, our method generates the smaller score, which means that the stitching quality of our method is higher.

D. COMPARISON OF STITCHING TIME COSTS

To evaluate the computational efficiency of the proposed method, we calculated the time cost of different methods on different databases and compared them with the state-of-the-art stitching methods, namely APAP [9], ANAP [11], GSP [15], and REW [4]. The authors of APAP and ANAP provided acceleration components for higher computational efficiency. The proposed method is implemented on MATLAB, while GSP is implemented on C++. The time consumption of the GSP is only used as a reference due to platform differences.

Table 2 shows the comparison of the time consumption of different methods. Note that APAP and REW are faster, except for GSP. APAP lacks the calculation of similarity, while REW lacks the calculation of local homography. Furthermore, our approach is slower than APAP and REW, which stems from our warping not only combining local homography and global similarity, but also introducing normal distribution, local projection deviation and linear smoothing pixel calculations. However, the proposed method avoids the homography of linearized non-overlapping region, making the method perform faster than ANAP. From the comparison results of Table 2, the proposed method ignores the efficiency of the algorithm.

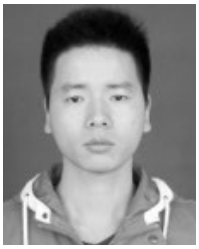
V. CONCLUSION

In this paper, we propose a novel stitching method to solve the problem of local misalignment. On one hand, we use the normal distribution to convert the value of the projection deviation into a probability event, which effectively refines the matching and improves the image alignment accuracy. On the other hand, we use a 3D surface interpolation model to describe the local projection deviation of the image in the local region, which improves the alignment of the local warp. The experimental results show that the proposed method can accurately align the image and eliminate ghosting. However, compared with the state-of-the-arts such as APAP, REW, and GSP, the efficiency of the proposed method is slightly reduced. As a future work, we will improve the efficiency of the proposed method and explore the application of line protection and seam-cut methods in image stitching.

REFERENCES

[1] R. Szeliski, "Image alignment and stitching: A tutorial," *Found. Trends Comput. Graph. Vis.*, vol. 2, no. 1, pp. 1–104, 2006.
[2] T. Xiang, G.-S. Xia, L. Zhang, and N. Huang, "Locally warping-based image stitching by imposing line constraints," in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 4178–4183.
[3] K. Y. Lee and J. Y. Sim, "Stitching for multi-view videos with large parallax based on adaptive pixel warping," *IEEE Access*, vol. 6, pp. 26904–26917, May 2018.
[4] J. Li, Z. Wang, S. Lai, Y. Zhai, and M. Zhang, "Parallax-tolerant image stitching based on robust elastic warping," *IEEE Trans. Multimedia*, vol. 20, no. 7, pp. 1672–1687, Jul. 2018.
[5] A. Levin, A. Zomet, S. Peleg, and Y. Weiss, "Seamless image stitching in the gradient domain," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2004, pp. 377–389.
[6] N. Li, Y. Xu, and C. Wang, "Quasi-homography warps in image stitching," *IEEE Trans. Multimedia*, vol. 20, no. 6, pp. 1365–1375, Jun. 2018.
[7] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 59–73, Aug. 2007.

[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 Recognit. (CVPR)*, Jun. 2011, pp. 345–352.
[9] J. Zaragoza, T.-J. Chin, M. S. Brown, and D. Suter, "As-projective-as-possible image stitching with moving DLT," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 2339–2346.
[10] C.-H. Chang, Y. Sato, and Y.-Y. Chuang, "Shape-preserving half-projective warps for image stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 3254–3261.
[11] 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 Recognit. (CVPR)*, Jun. 2015, pp. 1155–1163.
[12] F. Zhang and F. Liu, "Parallax-tolerant image stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 3262–3269.
[13] S. Li, L. Yuan, J. Sun, and L. Quan, "Dual-feature warping-based motion model estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 4283–4291.
[14] 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. (ECCV)*, 2016, pp. 370–385.
[15] Y.-S. Chen and Y.-Y. Chuang, "Natural image stitching with the global similarity prior," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2016, pp. 186–201.
[16] 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.
[17] 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.
[18] Z. Qiu, H. Tang, and D. Tian, "Non-rigid medical image registration based on the thin-plate spline algorithm," in *Proc. World Congr. Comput. Sci. Inf. Eng.*, Los Angeles, CA, USA, Mar./Apr. 2009, pp. 522–527.
[19] W. Sun, W. Zhou, and M. Yang, "Non-rigid registration of medical images with scale-space corner detection and thin-plate spline," *Biomed. Signal Process. Control*, vol. 7, no. 6, pp. 599–605, Nov. 2012.
[20] H. Zhou, Y. Kuang, Z. Yu, S. Ren, Y. Zhang, T. Lu, and J. Ma, "Image deformation with vector-field interpolation based on MRLS-TPS," *IEEE Access*, vol. 6, pp. 75886–75898, 2018.
[21] J. Gao, S. J. Kim, and M. S. Brown, "Constructing image panoramas using dual-homography warping," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Colorado Springs, CO, USA, Jun. 2011, pp. 49–56.
[22] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[23] O. R. Chum and J. Matas, "Optimal randomized RANSAC," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 8, pp. 1472–1482, Aug. 2008.
[24] K. Rohr, H. S. Stiehl, R. Sprengel, W. Beil, T. M. Buzug, J. Weese, and M. H. Kuhn, "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.
[25] A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," in *Proc. 18th Int. Conf. Multimedia*, Firenze, Italy, Oct. 2010, pp. 1469–1472.
[26] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'Completely Blind' image quality analyzer," *IEEE Signal Process. Lett.* , vol. 20, no. 3, pp. 209–212, Mar. 2013.

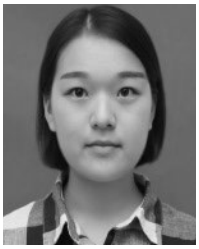


JIALIANG LI was born in Henan, China, in 1995. He received the B.S. degree from the Henan Institute of Science and Technology, Xinxiang, China, in 2017. He is currently pursuing the M.S. degree with the College of Electronic Engineering, Guangxi Normal University. His current research interests include computer vision, image stitching, video mosaics, and so on.

PINQUN JIANG received the Ph.D. degree in theoretical physics from the University of Science and Technology of China, in 2005. He is currently an Associate Professor with the Department of Electronic Science and Technology, College of Electronic Engineering, Guangxi Normal University. His research interests include analog integrated circuit design, artificial intelligence, intelligent robots, and machine vision.

SHUXIANG SONG received the Ph.D. degree from the Department of Electronic and Information Engineering, Huazhong University of Science and Technology, Wuhan, China. He is currently a Full Professor with Guangxi Normal University. His current research interests include intelligent detection, automatic control, and signal and image processing.

HAIYING XIA received the M.S. and Ph.D. degrees from the Department of Electronic and Information Engineering, Huazhong University of Science and Technology, Wuhan, China, in 2007 and 2011, respectively. She is currently an Associate Professor with Guangxi Normal University. Her current research interests include pattern recognition, medical image analysis, and neural networks.



MIN JIANG was born in Guangxi, China, in 1996. She received the B.S. degree from the Yulin Teachers College, China, in 2018. She is currently pursuing the M.S. degree with Guangxi Normal University. Her current research interests include image stitching, computer vision, and so on.