IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 15, 2017, accepted March 22, 2018, date of publication April 27, 2018, date of current version June 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2829082

An Accurate Multi-Row Panorama Generation Using Multi-Point Joint Stitching

JIN ZHENG[1,2], (Member, IEEE), ZHI ZHANG[1], QIUHAO TAO[3], KAI SHEN[1], AND YUE WANG[1]

[1]Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China
[2]State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China
[3]Shanghai CiGu Business Consulting Company, Ltd., Shanghai 200438, China

Corresponding author: Yue Wang (wangyue64@126.com)

This work was supported in part by the National Key Research and Development Plan under Grant 2016YFC0801002, in part by the NSFC under Grant 61370124, Grant 61632001, and Grant 61772054, and in part by the Army Equipment Research Project under Grant 301020203.

ABSTRACT Most of the existing panorama generation tools require the input images to be captured along one direction, and yield a narrow strip panorama. To generate a large viewing field panorama, this paper proposes a multi-row panorama generation (MRPG) method. For a pan/tilt camera whose scanning path covers a wide range of horizontal and vertical views, the image frames in different views correspond to different coordinate benchmarks and different projections. And the image frame should not only be aligned with the continuous frames in timeline but also be aligned with other frames in spatial neighborhood even with long time intervals. For these problems, MRPG first designs an optimal scanning path to cover the large viewing field, and chooses the center frame as the reference frame to start to stitch. The stitching order of multi-frame is arranged in first-column and second-row to ensure a small alignment error. Moreover, MRPG proposes a multi-point joint stitching method to eliminate the seams and correct the distortions, which makes the current frame accurately integrated into the panoramic canvas from all directions. Experimental results show that MRPG can generate a more accurate panorama than other state-of-the-art image stitching methods, and give a better visual effect for a large viewing field panorama.

INDEX TERMS Multi-point joint stitching, panorama, reference frame, scanning path, SIFT registration.

I. INTRODUCTION

There are many panoramic image stitching technologies in use nowadays [1]–[7]. For example, we can easily use our mobile phones or digital cameras to generate a panoramic photo, or we can use some commercial applications, such as Autostitch [8], [9], Kolor Autopano,[1] Microsoft ICE,[2] Realviz[3] and Microsoft Photosynth,[4] to synthesize several images and generate a panorama. These experiences give us an illusion that image stitching is a mature technology.

Although panoramic image stitching technologies have been widely used, the existing image stitching algorithms still have a lot of deficiencies. The most obvious problem is the limited coverage of the viewing field. As far as we know, most mobile phones including iPhone can only generate one-dimensional single-row panoramas, which means that the scanning path of camera only covers a straight line between the start point and the terminal point (shown in Fig.1) rather than a wide viewing field, and only a narrow strip panorama (shown in Fig.2(a)) is generated. There are some other mobile phones supporting image stitching in horizontal and vertical direction simultaneously, such as Moto X, which takes five photos located at the left, right, top, bottom and the center positions, to generate a panorama. Although the process extends the viewing field, the misalignment phenomenon generally exists (shown in the red rectangle in Fig.2(b)). In addition, it is true that we can generate a panorama based on multiple images using some image stitching tools, such as Autostitch [8] (the generated panorama is shown in Fig.2(c)). These tools usually use bundle adjustment to refine the 3D coordinates which describe the scene geometry, and then generate panoramas. However, bundle adjustment requires the support of camera pose parameters, and the computational cost is high. Meanwhile, it still introduces unconvincing

[1]http://www.kolor.com/
[2]https://www.microsoft.com/en-us/research/project/image-composite-editor/
[3]https://luminous-landscape.com/realviz-stitcher-4-0-review/
[4]https://photosynth.en.softonic.com/#app-softonic-review

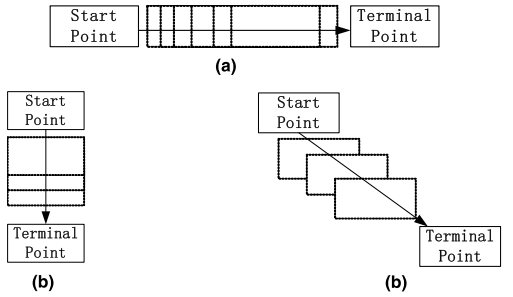


FIGURE 1. The limited coverage illustration in 1D scanning. (a) horizontal scanning. (b) oblique scanning. (c) vertical scanning.

FIGURE 2. The generated panoramas using the existing image stitching tools or applications. (a) 1D scanning based panorama using iPhone. (b) 5 images based panorama using Moto X. (c) multi-image based panorama using Autostitch [8].

results, such as the misalignment shown in the red rectangle in Fig.2(c). Here, the misalignment results in the image blur and dislocation.

Hence, in this paper, the work focuses on an accurate multi-row panorama generation method, which uses multiple images to generate a wide viewing field panorama. For the given start point and terminal point, this paper firstly proposes an optimal scanning path to cover the whole viewing field between the two points. According to the path, the center frame is selected as the reference frame, and the coordinate of the reference frame is used as the benchmark to produce the coordinate of the panorama. It helps to avoid strabismus and reduce cumulative errors. And then, the stitching order is arranged in first-column and second-row to ensure a small alignment error. Moreover, a multi-block and multi-point joint stitching method is proposed. Multi-block refers to calculate the projective transformation model for each image block to solve the problem that the different blocks in a large size panorama have the different projections. Multi-point refers to neighbor constrained points and relative constrained points, and these points help to align multi-frame from all directions. Neighbor constrained points are use to eliminate the cutting seams existing in neighboring blocks, and relative constrained points are used to maintain the correct relative positions between the blocks, which make the wide viewing panorama look coincident and accurate.

The rest of paper is organized as follow: Section II surveys related works. Section III introduces the proposed optimal scanning path, the reference frame and the stitching order. Section IV introduces multi-block and multi-point joint stitching. Section V evaluates the proposed method comparing with the state-of-the-art stitching methods. Finally, the paper is concluded in Section VI.

II. RELATED WORK

Producing a large viewing field panorama involves two questions: Which frame is regarded as the coordinate benchmark of the panorama? How does the current frame align with other frames which have overlapping region? The former refers to the reference frame selection, and the latter refers to the multi-frame stitching method.

A. REFERENCE FRAME SELECTION

Generally, for multi-image stitching, the coordinate of the reference frame is taken as the coordinate benchmark of the panorama. It means the selected reference frame is directly stuck onto a panorama canvas. Then, taking this reference frame as a benchmark, the remaining frames are corrected by image alignment and warped onto the panorama canvas in a certain order. The key is to choose one frame as the reference frame, and design an optimal stitching order to prevent the accumulated and amplified errors.

There are two categories of methods for the reference frame selection. One category is to choose a frame directly from a video sequence without considering the accumulative errors of the warping too much [10], [11]. Commonly, the first frame of a video sequence is selected as the reference frame [10]. Obviously, it is easy to cause strabismus and introduce the accumulative errors. Additionally, Kang *et al.* [11] selects a common frame as the reference frame. The common frame refers to the frame which has more overlapping region with other frames. The existence of a common frame means that the viewing field is limited. For the accumulative errors are not considered adequately, these methods do not behave very well.

The second category of method is to select the reference frame based on the evaluation of the accumulative errors using bundle adjustment [12]. The bundle adjustment method

uses the least square to optimize the re-projection errors between the video frames to be stitched and the reference frame. It can be defined as follows.

To estimate the required homography models $\{H^k\}_{k=1}^K$, bundle adjustment minimizes the sum of the re-projection errors of all correspondences by minimizing the cost in an iteration manner

$$E(\Theta) == \sum_{i=1}^{N}\sum_{k=1}^{K}\left\|x_i - f(p_i, H^k)\right\|^2 \tag{1}$$

Where $\Theta = [H^1, \ldots, H^K, p_1, \ldots, p_N]$. $\{H^k\}_{k=1}^K$ is a set of dependent homography models. K is the number of the homography models, and N is the number of the correspondences. x_i is the i^{th} arbitrary location in image I, p_i is the matched location of x_i in image I'. $f(p_i, H^k)$ is the projective warp (in homogeneous coordinates) defined as

$$f(p_i, H^k) = \left[\frac{r_1[p_i^T\,1]^T}{r_3[p_i^T\,1]^T}\ \frac{r_2[p_i^T\,1]^T}{r_3[p_i^T\,1]^T}\right]^T \tag{2}$$

where r_1, r_2, r_3 are the three row vectors of homography model H^k.

Thus, a 3D coordinate system is established using formula (1), and the coordinate system of the panorama is also established. However, for a continuous scanning video, in view of the characteristics of camera movement, the positions of the adjacent frames are very close and regular. Thus, the process of reconstructing a 3D coordinate system and minimizing the warping error by iteration, is a waste of computing resources. In addition, the method pays more attention to the re-projection error and the global alignment, but is less concerned with the similar motions of the adjacent scenery. Thus, it introduces inaccurate alignment (like Fig.2(c)).

B. MULTI-FRAME STITCHING

In many scenarios, it is supposed that the panorama is the multi-frame stitching result for a video camera scanning along one direction, that means only the adjacent frame on the timeline has overlapping region with the current frame [13]. So the matching relationship only exists between the current frame and its adjacent frame on the timeline. The only one frame along the scanning direction needs to be aligned. We define the stitching result along one direction as a single-row panorama. The two frames alignment in these scenarios is relatively simple. The commonly used methods include pixel-based matching [3], [13] and feature-based matching [14]–[17]. Among them, feature-based matching, especially SIFT (Scale-Invariant Feature Transform) feature matching [17], is very popular.

In contrast, 2D scanning refers to the movement along horizontal and vertical directions. According to the scanning path, 2D scanning usually has multi-row scanning frames with the different vertical perspectives (as shown in Fig.3(a)), especially for a pan/tilt camera in the ground surveillance. The essential characteristic of a 2D scanning is the current frame not only has overlapping region with the adjacent frame

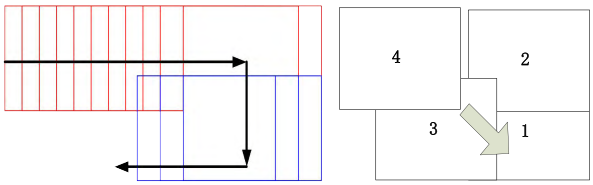


FIGURE 3. 2D scanning path along horizontal and vertical directions. (a) 2D scanning path. (b) multi-frame stitching.

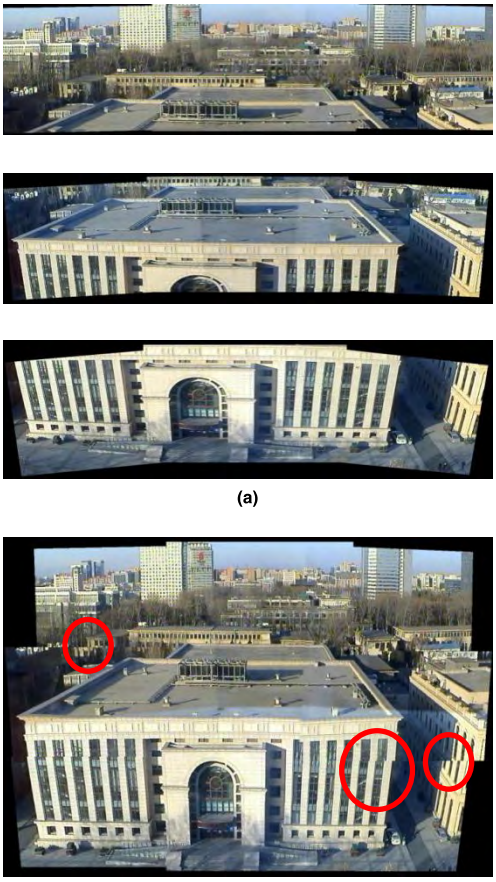


FIGURE 4. The misaligned panorama for a wide viewing field. (a) three single-row stitching panoramas. (b) a multi-row stitching panorama.

on the timeline, but also has overlapping region with the previous frames even with long time intervals. For a 2D multi-row panorama generation, multiple frames with overlapping region in vertical and horizontal neighborhoods need to be warped onto the current frame (as shown in Fig.3(b)). The determination of the multiple frames with overlapping region, and the calculation of the projective transformation models between them, as well as the detailed warping for each pixel, all need to be considered.

Moreover, in a wide viewing field, the different frames, even the different regions of an image, probably correspond to the different projections. The direct stitching using only one projective transformation model for a wide viewing field image is easy to introduce errors. For example, Fig.4(b) is

a multi-row panorama which is based on three single-row panoramas shown in Fig.4(a). The stitching between two single-row panoramas is modeled using only one projective transformation model. Obviously, even the results of single-row panoramas are good, there are misalignment for the multi-row panorama (shown in the red circles in Fig.4(b)). The reason of misalignment is that the projective discriminations exist in different regions for a wide single-row panorama. Moreover, the frames with overlapping region in neighborhood are probably captured at different times, and some frames even undergo the changes of illumination. These further exacerbate the difficulty of accurate alignment. Hence, the generation of an accurate multi-row panorama is more difficult than a single-row panorama.

In summary, the main problems for a wide panorama generation are the determination of multi-frame with overlapping region, the stitching order of multi-frame and the stitching method. The first problem is related to the scanning path. For a pan/tilt camera in ground surveillance, the regular scanning paths produce overlapping frames at regular positions. In this application, the determination of multi-frame which has overlapping region with the current frame is simple. For the second problem, Davis [18] searches a linear transformation model from the alignment set, and then decides the stitching order, but the calculation is complex. Kang *et al.* [11] uses graph-based reconstruction to determine the stitching order, but the method requires that the common frame is overlapped with all other frames.

For the stitching method, it concerns: (1) the stitching of multi-frame in neighborhood along different directions; (2) the stitching for generating a large size panorama. For (1), the existing methods align the current frame with multi-frame, respectively. It is based on the rule that each frame has its overlapping region with the current frame. Thus, multiple local projective transformation models are obtained. Then, for each pixel in the current frame, the final projective transformation model is a weighted result of multiple local projective transformation models [1], [10], [15]. The weight of each local projective transformation model is computed according to the distance between this pixel and the center of the overlapping region for the corresponding frame. The smaller the distance is, the bigger the weight is. However, the weighting process introduces image blur.

For (2), because the different regions of a wide panorama have different projections, many works propose to divide all the pixels in one image into some groups firstly [2], [10], [15], and for each group, one local projective transformation model is computed using the matches located in this group to model the local projection. For the pixel grouping methods, Zaragoza *et al.* [10] proposes As-Projective-As-Possible stitching (APAP), Wang *et al.* [19] and Chin *et al.* [20] propose multiple structures segmentation, and Gao *et al.* [14] proposes dual-homography warping method, et.al. Most notably, APAP divides the frame into many cells whose size is 100×100 pixels, and calculates the projective transformation model for each cell using SIFT features. Depending on these projective transformation models, each cell is aligned with the reference frame based on a local warping process. The multi-cell alignment strategy partially solves the problem that the different projections exist in the different image regions. However, because each cell is considered separately and the constraints of positional relation among these cells are not considered, the distortions still exist in some regions of the panorama, e.g., the edge of the building is distorted (shown in the blue box of Fig.5). In addition, the trivial partition makes the generated panorama blur when multiple cells are stitched (shown in the red box of Fig.5).

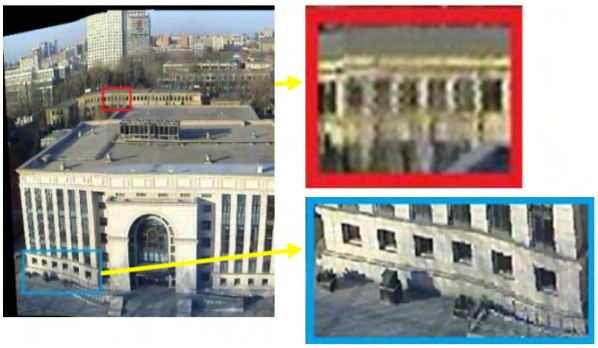


FIGURE 5. Blur and distortion phenomena illustration for APAP [10].

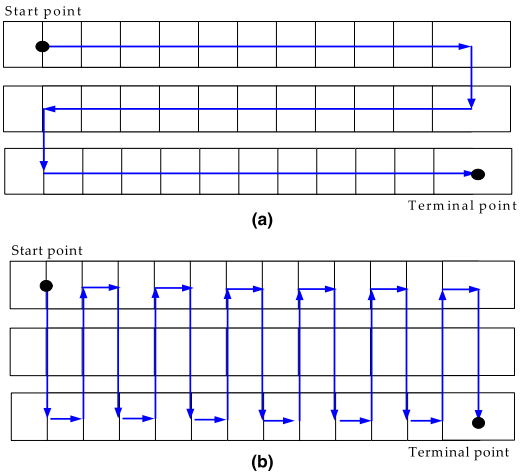


FIGURE 6. Two kinds of scanning path: row priority vs. column priority.

III. INPUT REGULARIZATION

The process of MRPG mainly includes two parts: input regularization and multi-point joint stitching. For a multi-row panorama, when the input images are captured using a pan/tilt camera around a fix point, the scanning path can help figure out the position relationship of these spatial overlapping frames. And then, the position relationship can determine the reference frame, the stitching order, as well as the alignment strategy.

A. OPTIMAL SCANNING PATH

To generate a large view panorama, the camera should cover the scene as large as possible. For the given start point and

terminal point, the camera should scan along horizontal and vertical directions, not from the start point to the terminal point directly. In addition, because the horizontal scanning range is often larger than the vertical scanning range, row priority (Route A in Fig.6) is more appropriate than column priority (Route B in Fig.6) for less turning points of the camera. Thus, an optimal scanning path for multi-row scanning is designed, which is row priority.

To guarantee a good stitching, the different scanning rows should satisfy a certain overlapping rate. Meanwhile, instead of using all the frames for stitching, the frames in the same row should be extracted in accordance with the overlapping rate requirement to ensure the computation speed. The number of the extracted frames in the same row satisfies formula (3).

$$n = \frac{W - w}{w \times (1 - o)} + 1 \tag{3}$$

Here, w refers to the width of one image frame, W refers to the horizontal coverage when the camera is moving, o refers to the overlapping ratio of two extracted adjacent frames, and n refers to the number of extracted frames in each row. Thus, for each scanning row, the numbers of the extracted frames are equal. The frames located in the different scanning rows and at the same vertical positions, can establish the one-to-one column correspondences. For a good stitching, the overlapping ratios of the computed frames, including the one-to-one column correspondence frames located in the different rows, as well as the extracted frames located in the same row and adjacent column, are both greater than 50% empirically.

B. CENTER REFERENCE FRAME

As the captured video is regularized, one frame is directly chosen as the reference frame. In consideration of the accumulative errors, the center frame is selected as the reference frame instead of the first frame of video or some other frames. Suppose there are m scanning rows, and n extracted frames in one scanning row, that means, there are $m \times n$ image frames. The center frame, which is located in the $\lfloor m/2 \rfloor^{\text{th}}$ row and the $\lfloor n/2 \rfloor^{\text{th}}$ column, is regarded as the only reference frame.

In detail, as the camera is doing a uniform rectilinear movement, it can be assumed that the error of calculated projective model between two frames is d. Suppose there are $m \times n$ frames (m and n are both odd numbers, and $m \leq n$) to be stitched, two cases, including the center frame is served as the reference frame and the first frame is served as the reference frame, are illustrated in Fig.7, respectively. The accumulated errors of two cases can be calculated by the follow formulas (4)-(5).

$$E_{cf} = d^0 + \sum_{i=1}^{(m-1)/2} 4id^i + \sum_{i=(m+1)/2}^{(n-1)/2} 2md^i + \sum_{i=1}^{(m-1)/2} 4id^{(m+n)/2-i} \tag{4}$$

$$E_{ff} = \sum_{i=1}^{m-1} (id^{i-1} + id^{m+n-1-i}) + \sum_{i=m-1}^{n-1} md^i \tag{5}$$

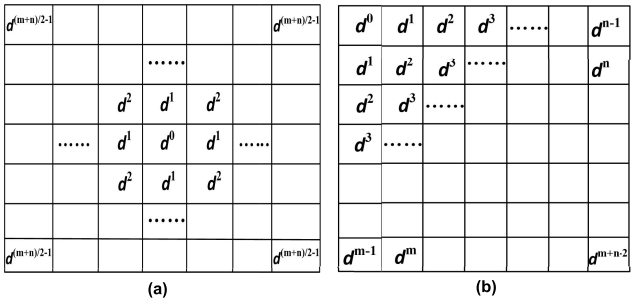


FIGURE 7. The comparisons of accumulated errors. (a) the errors of the proposed center frame method. (b) the errors of the existing first frame method.

E_{cf} refers to the accumulated errors when the center frame is set as the reference frame, and E_{ff} refers to the accumulated errors when the first frame is set as the reference frame. It can be proved that E_{cf} is smaller than E_{ff}. Also, choosing the center frame as the reference frame can certainly avoid strabismus, because the center frame is usually just facing the observation point.

C. STITCHING ORDER

Generally, if the reference frame is determined, all the other frames around the reference frame can be stitched onto the panorama canvas. But it is difficult to consider the transformations along horizontal and vertical directions simultaneously, for the frames are captured at the different times and different perspectives. Thus, this paper rearranges the stitching order. The one-to-one column correspondence frames in the different scanning rows but in the same column are first aligned, which is called as first-column stitching order. Then, some column-panoramas are gotten. Based on these column-panoramas, the stitching process from the center column to both sides along left directions and right directions are carried out, which is called as second-row, and finally a wide panorama is generated.

The stitching along the same column is a 1D process, and thus only two frames, which are the spatial adjacent frames along the vertical direction, are stitched each time. Because these image frames with limited size have been calibrated in a same column, and the overlapping ratio is greater than 50%, the stitching operation based on SIFT feature matching and single projective transformation model is used to align them. The column-panorama is generated using SIFT feature matching [9] and RANSAC outlier remove [21], [22]. Hence, the transformation matrix between two adjacent frames located in the same column can be calculated, which describes the transformation along vertical direction. The two images in Fig.8 are two column-panoramas with three scanning rows, and each of them shows a good alignment and visual effect.

IV. MULTI-POINT JOINT STITCHING

Because a column-panorama is stitched based on multiple frames located in multiple scanning rows, the height of the column-panorama is large. In addition, the different

FIGURE 8. Two column-panoramas: the different regions have the different transformation models.

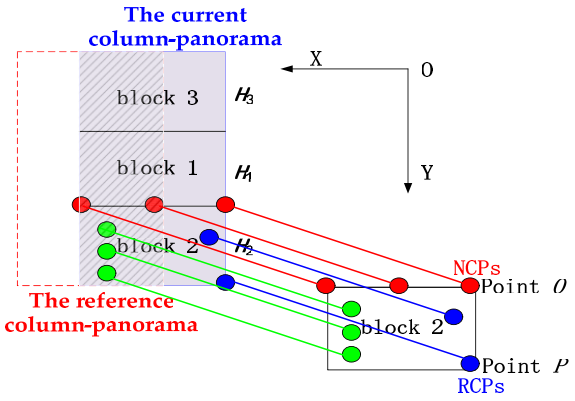


FIGURE 9. NCPs & RCPs illustration.

perspectives and capture times result in the different regions in a column-panorama have different projections. The direct stitching of two large size column-panoramas is easy to introduce errors. Thus, this paper proposes the multi-block and multi-point joint stitching method.

A. MULTIPLE BLOCKS PARTITION

In Fig.8, the red lines located on the top of the column-panoramas are parallel, whereas, the red lines below, as well as the green lines are not parallel. The reason is that these positions correspond to the different scanning rows and columns, and the parallax results in the different projections contained in the different regions. Constructing a single projective transformation model to simulate the various warps in different regions existing in two large size column-panoramas is inconceivable.

In the following second-row processing, considering the different regions in a column-panorama have different projections, each column-panorama is divided into several blocks firstly, and then a projective transformation model is computed for each block. The number of blocks can be simply set as the number of scanning rows, and these blocks are non-overlapped.

If the projective transformation model in each block is computed separately, and the positional constraints among these blocks which have been computed in first-column process are not considered, what we have done aforementioned to stitch those one-to-one column correspondence frames is meaningless. To maintain the consistency and accuracy of the whole panorama, this paper proposes neighbor constrained points and relative constrained points for the multi-block stitching, which ensure a good alignment for multiple frames in all directions.

B. NEIGHBOR CONSTRAINED POINTS

For the reference column-panorama and the column-panorama to be stitched, the latter is divided into several blocks. Each block has a common border with its adjacent block, which is determined by first-column process.

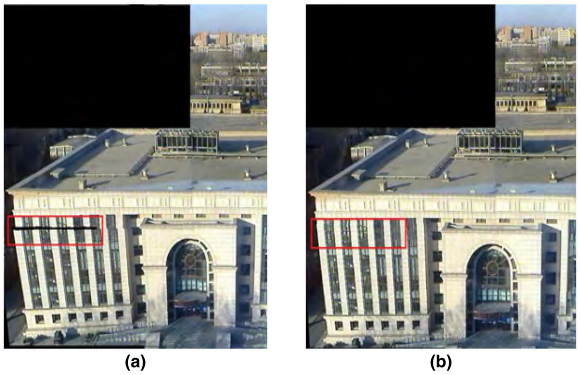
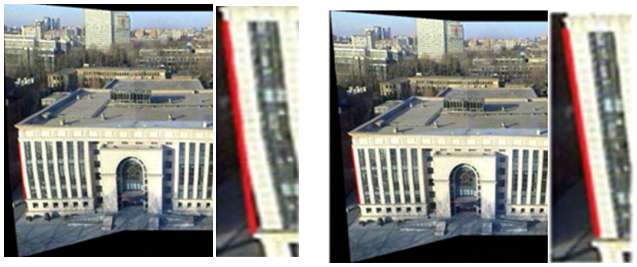


FIGURE 10. Multi-block stitching using NCPs. (a) panorama without NCPs. (b) panorama with NCPs.

Therefore, after the first block is stitched, the position of the common border with the second block is also determined. Some points on the common border are chosen to record their coordinates and called Neighbor Constrained Points (NCPs, shown as the red points in Fig.9). When the projective transformation model of the second block is calculated, these recorded NCPs, together with the inliers in the overlapping region of block 2 (shown as the green points in Fig.9) are all used.

For a column-panorama, if each block is stitched separately via the projective transformation model calculated just using the inliers in this block, the result is shown in Fig.10(a). Obviously, this process just considers the overlapping region between the current block and the reference column-panorama along the horizontal direction, and ignores the positional constraints between the neighboring blocks along the vertical direction. Thus, there is a cutting seam between the two blocks, which means the consistency of the column-panorama is broken. In contrast, Fig.10(b) shows the stitching result using NCPs. In this case, not only the inliers in the overlapping region, but also the NCPs located on the common border, are all participated in the calculation of the projective transformation model. Particularly, NCPs maintain the consistency of the two neighboring blocks, and thus, the cutting seam is removed.



(a) (b)

FIGURE 11. Panorama stitching using RCPs. (a) the panoramas without RCPs. (b) the panoramas with RCPs.

C. RELATIVE CONSTRAINED POINTS

Even NCPs are considered, and the obvious cutting seam is removed, the inaccurate phenomena still probably exists. For example, the edge of the building is distorted in Fig 11(a), and actually it should be a straight line (see the enlarged drawing on the right). The aforementioned stitching strategy takes into account the matching of the overlapping region, as well as the matching of the common border. However, the inaccuracies still exist in those non-overlapping region.

To solve this problem, Relative Constrained Points (RCPs) are proposed when the projective transformation models of the subsequent blocks are computed. RCPs refer to those points in the non-overlapping region of the pending block (such as the blue point in Fig.9), and their positions are determined by the first-row process and the projective transformation model computed by the previous aligned block.

The computation of projective transformation models for NCPs and RCPs are illustrated in Fig.9. A typical NCP point O is assumed to be the origin of the coordinate system and its coordinate is (0,0). A typical RCP point P is the right bottom point of the pending block, which is located in non-overlapping region. If the height of a block is h, the coordinate of P is (0, h). The projective matrix of the previous block (block 1) is supposed to be H_1, and thus, the warping should follow formula (6).

$$X' = H_1 X \tag{6}$$

Here, X refers to the original coordinate, and X' refers to the coordinate after warping $\boldsymbol{H}_1$. According to the positional constraints, the coordinate of point O and point P after warping can be achieved, and supposed to be (x_1, y_1), (x_2, y_2). In such way, $(0,0) \leftrightarrow (x_1, y_1)$ and $(0, h) \leftrightarrow (x_2, y_2)$ are two pairs of matched feature and can be added in calculating the new projective matrix $\boldsymbol{H}_2$ for the current block (block 2). The whole process is constantly done according to the successive blocks, and $\boldsymbol{H}_1$, $\boldsymbol{H}_2$,... are computed for each subsequent block.

Fig.11(b) shows the result using RCPs. For RCPs maintain the consistency of the relative position between two blocks, the result illustrates the correct straight line along the edge of the building.

Algorithm 1 Multi-Row Panorama Generation Algorithm(MRPG)

Input: A sequence of images captured by multi-row scanning. The number of scanning rows is m, and the threshold of overlapping rate is o.
Output: The generated panorama.
1: Input regularization:
1.1: Extract the frames to be stitched based on o. n is the number of extracted frame in each row through formula (3). The input frames are regulated as $F_{1,1}, F_{1,2},\ldots, F_{1,n}, F_{2,1}, F_{2,2},\ldots F_{2,n},\ldots, F_{m,1}, F_{m,2},\ldots F_{m,n}$.
1.2: The center frame located in the $\lfloor m/2 \rfloor^{\text{th}}$ row and the $\lfloor n/2 \rfloor^{\text{th}}$ column, is regarded as the reference frame.
2: Column-panorama generation:
for i =1 to n
Stitching $F_{1,\text{i}}, F_{2,\text{i}},\ldots, F_{m,\text{i}}$ to generate CP_i: starting from $F_{\lfloor m/2 \rfloor,i}$, two adjacent frames are stitched towards the up and down directions until reaching $F_{1,i}$ and $F_{m,i}$. The stitching method adopts SIFT feature matching and RANSAC outlier remove.
end for
Output: column-panoramas CP_1,CP_2, ...,CP_n.
3:Two column-panoramas stitching:
3.1: Suppose the reference column-panorama is CP_i, and the adjacent column-panorama to be stitched is $CP_j(j = i\text{-}1$ or $j = i + 1$, $1 \leq i, j \leq n)$.CP_j is divided into $m$blocks, which are $CP_{1,j}$, $CP_{2,j},\ldots,CP_{m,j}$.
3.2: Starting from $CP_{\lfloor m/2 \rfloor,j}$, the projective transformation model between $CP_{\lfloor m/2 \rfloor,j}$ and CP_i is computed using the inliers in the overlapping region, which is $\boldsymbol{H}_1$.
3.3: for $k = \lfloor m/2 \rfloor + 1 : 1 : m$
The projective transformation model $\boldsymbol{H_k}$ is computed using the inliers in the overlapped regions, NCPs and RCPs.
end for
3.4: for $k = \lfloor m/2 \rfloor - 1{:}{-}1{:}1$
The projective transformation model $\boldsymbol{H_k}$ is computed using the inliers in the overlapped regions, NCPs and RCPs.
end for
3.5 for k =1:m;
$CP_{k,j}$ is stitched with CP_iusing $\boldsymbol{H_k}$.
end for
4: A wide panorama generation.
Starting from$CP_{\lfloor n/2 \rfloor}$, two adjacent column-panoramas are stitched towards the left and right directions until reaching CP_1 and CP_n.

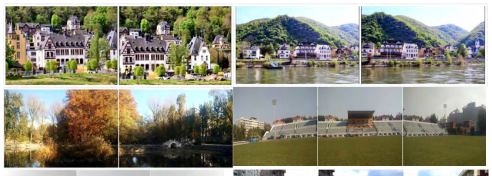
D. ALGORITHM SUMMARY

Based on the above principles, the proposed MRPG algorithm can be summarized in Algorithm 1.

V. EXPERIMENTAL EVALUATION

In this section, we first introduce the collection data set for multi-row panorama generation. Then, we present the experimental settings as well as comparison methods, followed

TABLE 1. The data set for multi-row panorama generation.

Scenarios	Illustration of the typical scenarios
Building: 9 sequences, 1140 frames	
Scenery: 8 sequences, 1108 frames	
Street view: 8 sequences, 1129 frames	

by the performance comparisons of different methods and analyses.

A. DATASET AND EXPERIMENTAL SETTINGS

Since there are no benchmark data sets for the evaluation of multi-row panorama generation at present, this paper first establishes a relevant data set. The video data is collected on Internet or captured by some surveillance cameras and personal mobile phones. There are three kinds of image resolution, 1920 × 1080 pixels, 704 × 576 pixels and 640 × 480 pixels. The established data set includes three types of scenarios, a total of 25 video sequences, and 3377 image frames. The maximum horizontal scanning coverage is 360 degree, and the maximum vertical scanning coverage almost reaches 70 degree. The typical scenarios are shown in TABLE 1.

The process of MRPG is totally implemented in C++ as the method described above. In addition, OpenCV[5] is used to read and write video frames and images, and OpenSIFT Library[6] is used to extract SIFT features in the images. The default parameters of SIFT are adopted except SIFT_CONTR_THR and SIFT_IMG_DBL. The former is set to 0.04 and the later is set to 1 to extract more features for the weak-texture images.

B. NCPs AND RCPs ANALYSIS

1) NCPs ANALYSIS

NCPs are distributed on the common border of two adjacent blocks. For the width of frame is w, there are w candidates. We test two sampling modes, one mode is sequential dense sampling, and the other mode is uniformly sparse sampling (shown in Fig.12). The former obtains NCPs point by point along the common border from the side of the overlapping region, and the latter obtains evenly spaced NCPs on the common border.

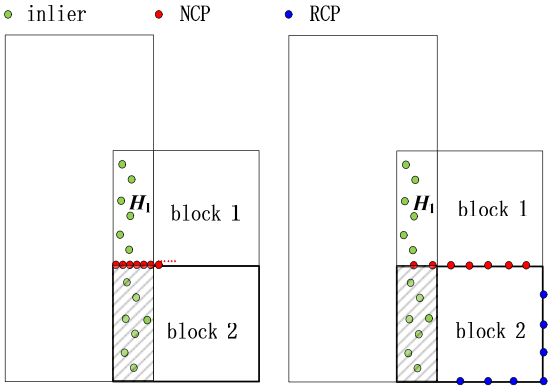


FIGURE 12. Two sampling modes. (a) sequential dense sampling. (b) uniformly sparse sampling.

Suppose the projective transformation model of block 1 has gotten, which is $\boldsymbol{H_1}$. When the projective transformation model of block 2 is computed, the inliers (green points in Fig.12) in the overlapping region (shadow area) are extracted firstly. Assuming that there are k inliers, thus, k points (red points) using two sampling modes are scattered on the common border, respectively (see Fig.12(a) and Fig.12(b)).

Next, we test the effect of the number of NCPs on the accuracy of the transformation model. Here, l NCPs are extracted from the k candidates. There are also two modes: for mode (a), the l points are extracted point by point from the side of the overlapping region; for mode (b), the l points are evenly extracted from the center of the common border to both sides. l is from 0 to $k(k \leq w)$. And then, the l points are added to the inliers set to compute the projective transformation model χ_l together. For each specified l, the generated χ_l is deemed as $\boldsymbol{H_2}$.

To measure the results, the re-projection error is adopted. For the inliers, they have the warped coordinates according to SIFT feature matching; for NCPs, $\boldsymbol{H_1}$ can determine the warped coordinates. Thus, (x, x') represents one correspondence, and x is the referenced feature coordinate, and x' is the warped coordinate. Moreover, all the inliers and NCPs can get the final warped coordinate based on $\boldsymbol{H_2}$. That is, x is warped to $\hat{x}$ by $\boldsymbol{H_2}$. The re-projection error is defined as

$$\mathrm{Re}Err = |x' - \hat{x}| \tag{7}$$

For N correspondences, the average re-projection error is defined as

$$Ave\_\mathrm{Re}Err = \frac{\sum_{i=1}^{N} |x_i' - \hat{x}_i|}{N} \tag{8}$$

For χ_l, which is computed based on the k inliers and the l NCPs, the average re-projection error curve for inliers and NCPs are drawn in Fig.13 (in this example, the processed image is shown in Fig.10). For fairness, the number of inliers and the number of NCPs involved in the error calculation

[5]https://opencv.org/
[6]http://robwhess.github.io/opensift/

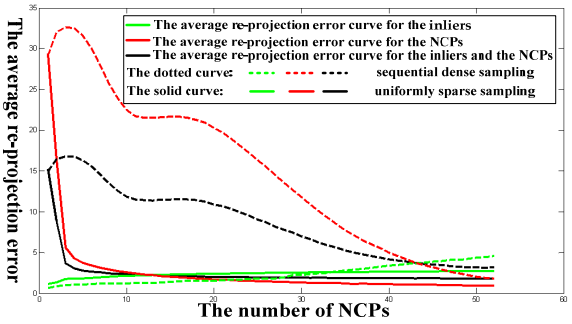


FIGURE 13. The re-projection error for NCPs.

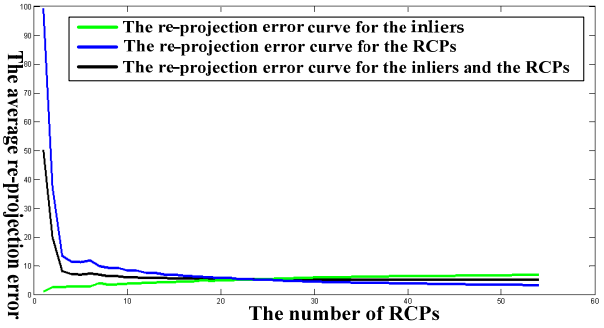


FIGURE 14. The re-projection error for NCPs.

are equal. That is, the green curve denotes the average re-projection error of the k inliers, and the red curve denotes the average re-projection error of k candidate points on the common border. The average re-projection error for the k inliers and the k candidate points on the common border is drawn in black curve. The dotted curve and solid curve correspond to two sampling modes, respectively. The horizontal coordinate axis is the number of NCPs involved in calculating the projective transformation model, which is l. The vertical coordinate axis is the average re-projection error.

With the increase of the number of NCPs, the average re-projection error of the k inliers also increases, whereas the average re-projection error of the k points on the common border decreases. It shows that the two types of pixels have an obvious projective discrimination. A large number of experiments show that the average re-projection error of the inliers increases slowly, while the average re-projection error of the points on the common border decreases sharply, even only several NCPs are added.

Meanwhile, this result also demonstrates that the re-projection error in uniformly sparse sampling mode converges faster than that in sequential dense sampling mode. Thus, 5 NCPs are added empirically using uniformly sparse sampling mode in the following experiments.

2) RCPs ANALYSIS

RCPs are extracted from the non-overlapping region. Since there are no correspondences in the non-overlapping region, we only extract the points on the uncommon border and the corner (such as the blue points in Fig.12(b)) as RCPs to simplify the computation.

Similarly, the average re-projection error of the inliers in overlapping region and RCPs are computed respectively, and the curves are drawn in Fig.14. With the increase of the number of RCPs, the re-projection error of the inliers increases gradually, whereas the re-projection error of the RCPs decreases rapidly. A large number of experiments illustrate that only several RCPs can guarantee a small re-projection error. Thus, 5 RCPs are used in the following experiments empirically. One RCP is taken at the corner, and other four RCPs are evenly located on two uncommon borders.

C. PANORAMAS COMPARISONS

1) SUBJECTIVE EVALUATION

APAP [10] is a representative method considering the alignment of parallax images. In APAP, the image has been divided into many fixed-size cells, and a local transformation model is computed based on the inliers in each cell. For each pixel in the image, the final transformation model is a weighted result of multiple local transformation models [2]. The idea of region segmentation and local transformation model computation in this paper is derived from APAP. In addition, Autostitch [8], Kolor Autopano[1] and Microsoft ICE[2] are popular commercial tools for image stitching. Therefore, the proposed method is compared with APAP, Autostitch, Kolor Autopano and Microsoft ICE.

In view of the three types of scenarios in the data set, Fig.15-17 show the experimental results for Building scenarios. From Fig.15-17(a)-(e), the generated panoramas using APAP, Autostitch, Microsoft ICE, Kolor Autopano and our method are displayed in turn. It can be seen that there are obvious misalignment for APAP, such as the seams in Fig.15(a), the distortion in Fig.16(a), the tilted flagpole and the blurry region in Fig.17(a). The results of Autostitch and Microsoft ICE are different from the real ones, such as the distortion of the straight edge of the building in Fig.15-16(b),(c). The result of Kolor Autopano may lose some image contents, such as Fig.16(d). In addition, the result of Kolor Autopano may also show misalignment, such as Fig.17(d). Comparatively speaking, there is no obvious seams or distortions in our results. It is worth mentioning that there are illumination changes and obvious scene depth differences in the case in Fig.17, our method stitches multiple images accurately, and the generated panorama has no obvious seams and distortions.

Fig.18 shows the experimental results for Scenery scenarios. It can be seen there are obvious seams and blurring in Fig.18(a). There is a ghost in Fig.18(b) and distortion in Fig.18(d). Comparatively, there is no obvious misalignment in Fig.18(c) and Fig.18(e).

In view of Street View scenarios, the results are similar to that of the Building and Scenery scenarios, but the premise is that there are not too many moving targets.

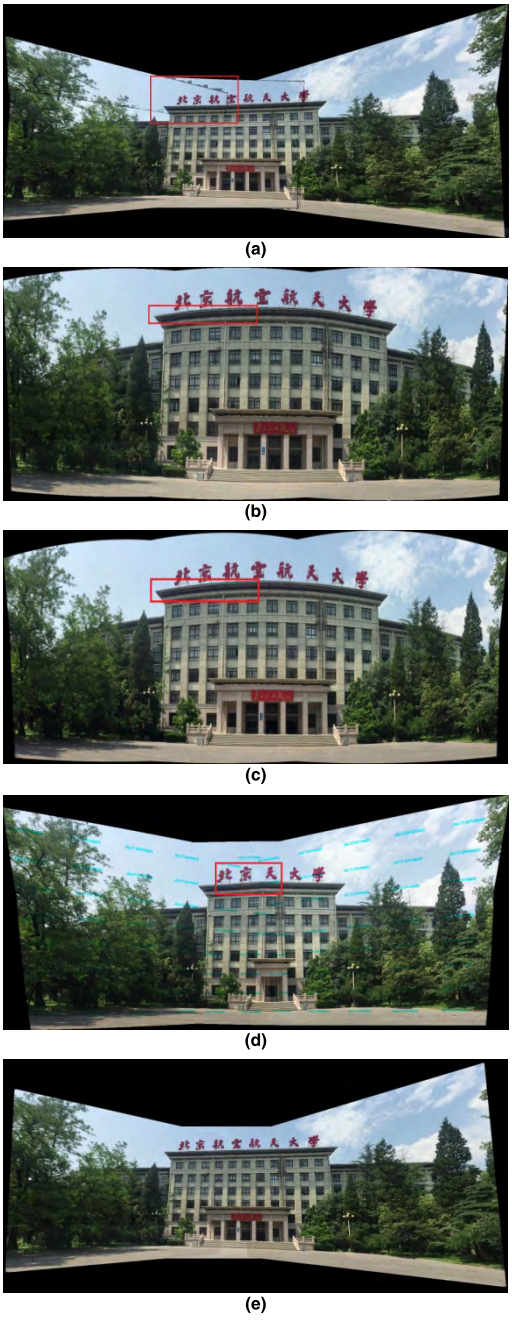


FIGURE 15. The generated panoramas for Main Building sequence. (a) APAP. (b) AutoStitch. (c) Microsoft ICE. (d) Kolor Autopano. (e) the proposed method.

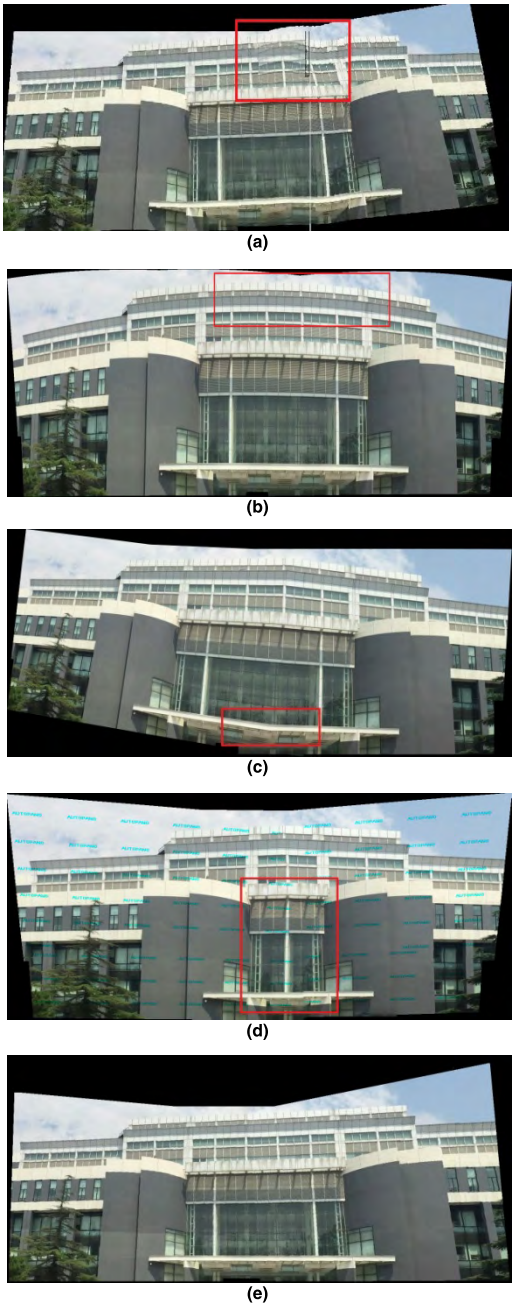


FIGURE 16. The generated panoramas for Library sequence. (a) APAP. (b) AutoStitch. (c) Microsoft ICE. (d) Kolor Autopano. (e) the proposed method.

In summary, APAP chooses the first frame as the reference frame, so the strabismus phenomenon always exists. Meanwhile, APAP divides the frame into some small cells, and the spatial constraint is broken, thus, distortion happens. In addition, the weighted process is easy to introduce image blurry. For those commercial tools, it is difficult to obtain technical data. Thus, the reason for the bad visual effect is not very clear. Comparatively speaking, the proposed method focuses on each step of the multi-row panorama generation, and the generated panorama has good visual effect.

2) OBJECTIVE EVALUATION

This paper also makes an objective evaluation on the data set. Because the code of APAP can be obtained from the authors of the papers, and the commercial tools AutoStitch, ICE and Kolor Autopano only provide the final stitching images, we only compare the proposed method with APAP. The average re-projection error is adopted to evaluate the accuracy of the alignment.

For each aligned image pairs, 20 matching points are marked artificially, which are regarded as the reference

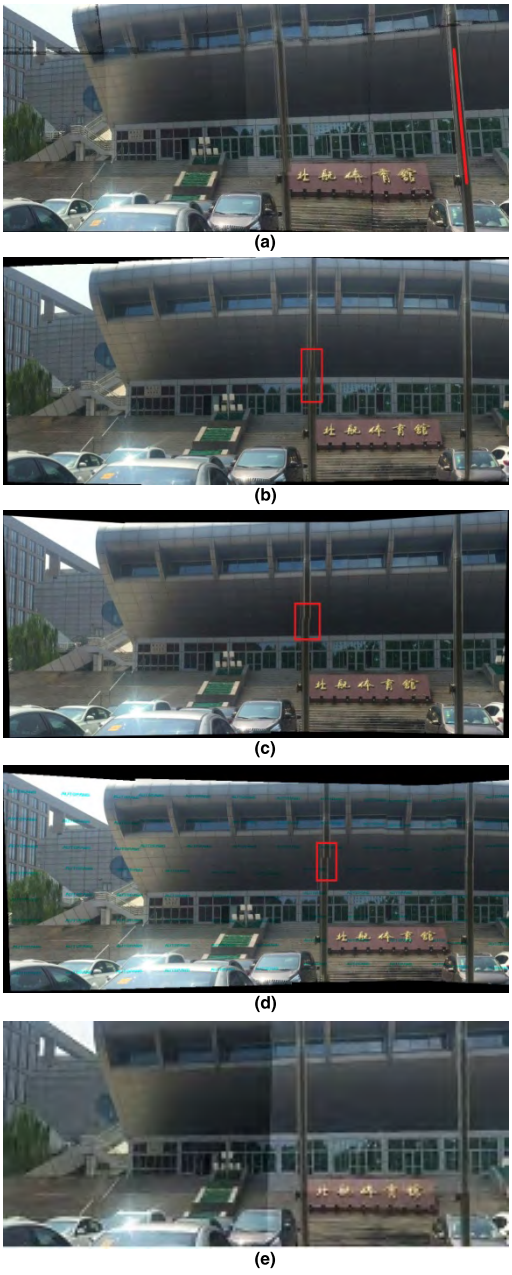


FIGURE 17. The generated panoramas for Gym sequence. (a) APAP. (b) AutoStitch. (c) Microsoft ICE. (d) Kolor Autopano. (e) the proposed method.

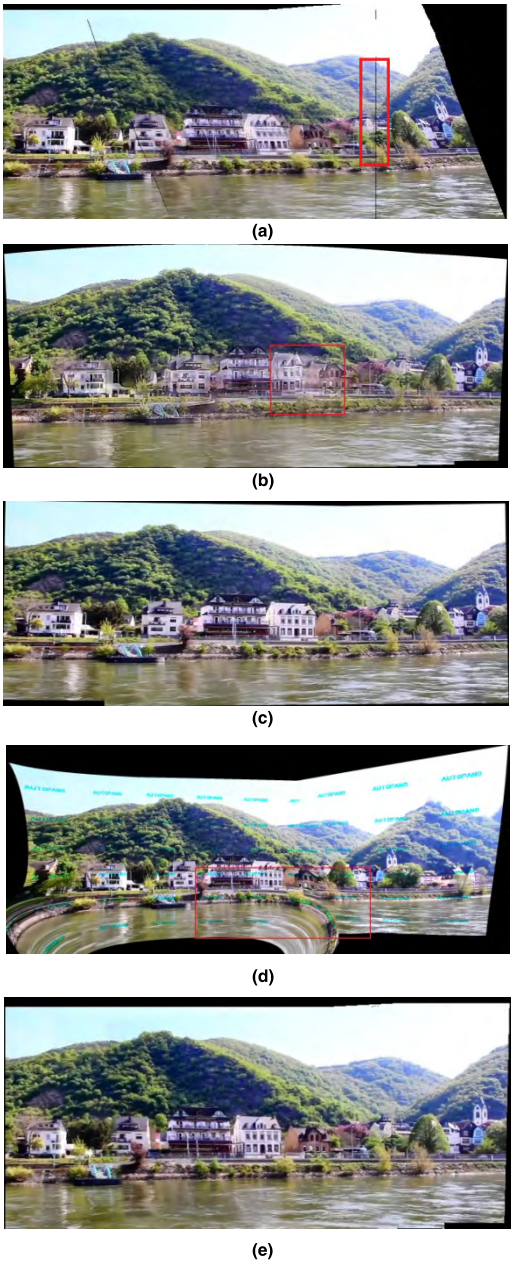


FIGURE 18. The generated panoramas for Rhine sequence. (a) APAP. (b) AutoStitch. (c) Microsoft ICE. (d) Kolor Autopano. (e) the proposed method.

positions. The reference positions are compared with the warped positions to get the re-projection error. TABLE 2 depicts the average re-projection error (over 20 repetitions) on the sequences of Building, Scenery, Street View scenarios for APAP and the proposed method. Obviously, the proposed method outperforms better.

TABLE 2. Average re-projection error (in pixel) comparisons.

Scenarios	APAP	The proposed method
Building	1.8	0.6
Scenery	2.9	0.8
Street View	3.4	0.8

D. TIME EFFCIENCY ANALYSIS

The computational cost of the proposed MRPG mainly focuses on image registration and image resampling.

In the process of image registration, the most time-consuming parts are the Gaussian pyramids building, the computation of the feature description and the feature matching. Suppose the size of one frame or one block is $h \times w$, and the SIFT features number is N. Suppose there are

TABLE 3. Processing time.

methods \ resolution	1920×1080	704×576	640×480
APAP	48.6s	44.4s	36.3s
AutoStitch	5s	1.2s	1s
ICE	5s	1.2s	1s
Kolor Autopano	5s	1.2s	1s
The proposed method	450ms	380ms	300ms

M group and L level in each group. Then, for one stitching, the cost for the image registration is

$$O(\sum_{i=0}^{M-1} \frac{h \times w}{4^i} \times L + N + N) \tag{9}$$

In the process of the image resampling, the cost is $O(h \times w)$. Thus, for a video sequence with u scanning row and v scanning column, the cost is

$$O((\sum_{i=0}^{M-1} \frac{h \times w}{4^i} \times L + N + N + h \times w) \times u \times v) \tag{10}$$

For comparisons, all the methods are tested using a PC with a Intel(R) Core^{TM}i3-3110M CPU (2.4GHz) and 4 GB RAM. For the three commercial tools AutoStitch, ICE and Kolor Autopano, the processing and optimization have been integrated into the executable programs, which are released on their official websites. Thus, we can only count the running time of the executable programs. For each tool, the running time includes three parts: reading two images, stitching two images and outputting the stitching result. Therefore, for the sake of fairness, the other methods also count the time of the three parts as the final running time.

For APAP, the code is provided by the authors of the papers, and they are run in MATLAB. For the proposed method, it is implemented in C++ code. Considering the different compiling environments, the comparison of the running time for these methods is unfair. We just list the running time of these methods in TABLE 3 for reference. Apparently, the processing times are different depending on the image size.

VI. CONCLUSION

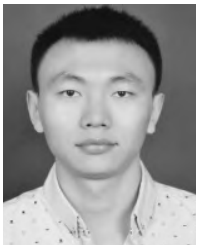
This paper proposes a multi-row panoramic image stitching method. Firstly, it designs an optimal scanning path to cover the large viewing field, and then selects the center frame to start to stitch. This process can cover the viewing field as large as possible, and also avoid the strabismus and accumulative errors. And then, the stitching process uses first-column and second-row manner, rather than uses the common stitching along the scanning direction, or uses the reference frame along horizontal and vertical directions synchronously. The first-column and second-row manner is in favor of handling the accurate alignment. Furthermore, multi-point joint stitching is proposed to guarantee the accurate matching in subtle regions, especially the stitching border and the non-overlapping region.

Experimental results show that the proposed method can provide a faster and more accurate panoramic image than other state-of-the-art image stitching methods, and also give a better visual effect in a large view panorama.

REFERENCES

[1] Y.-S. Chen and Y.-Y. Chuang, "Natural image stitching with the global similarity prior," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands, Oct. 2016, pp. 186–201.
[2] C.-C. Lin, S. U. Pankanti, K. N. Ramamurthy, and A. Y. Ar-avkin, "Adaptive as-natural-as-possible image stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1155–1163.
[3] S. Pravenaa and R. Menaka, "A methodical review on image stitching and video stitching techniques," *Int. J. Appl. Eng. Res.*, vol. 11, no. 5, pp. 3442–3448, 2016.
[4] J. Li, C. Li, T. Yang, and Z. Lu, "A novel visual-vocabulary-translator-based cross-domain image matching," *IEEE Access*, vol. 5, pp. 23190–23203, Oct. 2017.
[5] T. D. Nguyen, A. Shinya, T. Harada, and R. Thawonmas, "Segmentation mask refinement using image transformations," *IEEE Access*, vol. 5, pp. 26409–26418, 2017.
[6] J. Jia and C. K. Tang, "Eliminating structure and intensity misalignment in image stitching," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2. Oct. 2005, pp. 1651–1658.
[7] 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.
[8] 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.
[9] R. Hess. (2015). *OpenSIFT: An Open-Source SIFT Library*. [Online]. Available: http://robwhess.github.io/opensift/
[10] 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.
[11] E. Y. Kang, I. Cohen, and G. Medioni, "A graph-based global registration for 2D mosaics," in *Proc. 15th Int. Conf. Pattern Recognit.*, vol. 1. Sep. 2000, pp. 257–260.
[12] B. Triggs, P. Mclauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment—A modern synthesis," in *Vision Algorithms: Theory and Practice* (Lecture Notes in Computer Science). Berlin, Germany: Springer-Verlag, 2000, pp. 298–375.
[13] R. Kassab, S. Treuillet, F. Marzani, C. Pieralli, and J. C. Lapayre, "An optimized algorithm of image stitching in the case of a multi-modal probe for monitoring the evolution of scars," *Proc. SPIE, Conf. Adv. Biomed. Clin. Diagnostic Syst. XI*, vol. 8572, pp. A1–A10, Feb. 2013.
[14] 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)*, Jun. 2011, pp. 49–56.
[15] 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.
[16] F. Zhang and F. Liu, "Parallax-tolerant image stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, Jun. 2014, pp. 3262–3269.
[17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[18] J. Davis, "Mosaics of scenes with moving objects," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1998, pp. 354–360.
[19] H. Wang, T.-J. Chin, and D. Suter, "Simultaneously fitting and segmenting multiple-structure data with outliers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1177–1192, Jun. 2012.
[20] T. J. Chin, H. Wang, and D. Suter, "Robust fitting of multiple structures: The statistical learning approach," in *Proc. Int. Conf. Comput. Vis.*, Jun. 2009, pp. 413–420.
[21] T. Lai, H. Wang, Y. Yan, and L. Zhang, "A unified hypothesis generation framework for multi-structure model fitting," *Neurocomputing*, vol. 222, pp. 144–154, Jan. 2017.
[22] H. Wang, J. Cai, and J. Tang, "AMSAC: An adaptive robust estimator for model fitting," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2014, pp. 305–309.

JIN ZHENG (M'14) was born in Sichuan, China, in 1978. She received the B.S. degree in applied mathematics and informatics from the College of Science in 2001, the M.S. degree from the School of Computer Science, Liaoning Technical University, Fuxin, China, in 2004, and the Ph.D. degree from the School of Computer Science and Engineering, Beihang University, Beijing, China, in 2009. She joined the School of Computer Science and Engineering, Beihang University, in 2009. In 2014, she visited Harvard University, Cambridge, MA, USA, as a Visiting Scholar for one year. She currently teaches at the Digital Media Laboratory and at the School of Computer Science and Engineering, Beihang University. Her current research interests focus on moving object detection and tracking, object recognition, image enhancement, video stabilization, and video mosaics, among other similar interests.



ZHI ZHANG was born in Shanxi, China, in 1993. He received the B.S. degree from the Civil Aviation University of China in 2016. He is currently pursuing the M.S. degree with the School of Computer Science and Engineering, Beihang University. His current research interests focus on video mosaics, object detection, and so on.

QIUHAO TAO was born in Shanghai, China, in 1991. He received the B.S. degree from the College of Electronics and Information Engineering, Tongji University, in 2014, and the M.S. degree from the School of Computer Science and Engineering, Beihang University, China, in 2017. He is currently an Engineer with Shanghai CiGu Business Consulting Co., Ltd., Shanghai, China. His current research interests focus on video mosaics.

KAI SHEN was born in Hubei, China, in 1992. He received the B.S. degree from the Hubei University of Technology, Hubei, in 2015. He is currently pursuing the M.S. degree with the School of Computer Science and Engineering, Beihang University. His current research interests focus on video mosaics and moving object detection and tracking.

YUE WANG was born in Liaoning, China, in 1981. She received the B.S. and M.S. degrees from the School of Computer Science, Liaoning University, in 2004 and 2007, respectively, and the Ph.D. degree from the School of Computer Science and Engineering, Beihang University, Beijing, China, in 2017. Her current research interests focus on video mosaics, moving object detection and tracking, and video stabilization.