Geometric Structure Preserving Warp for Natural Image Stitching

Peng Du[1], Jifeng Ning[1†], Jiguang Cui[1], Shaoli Huang[2], Xinchao Wang[3] and Jiaxin Wang[1]
[1]Northwest A&F University, China, [2]Tencent AI-Lab, [3]National University of Singapore, Singapore
{peng-du, jiaxinwang123}@nwafu.edu.cn, njf@nwsuaf.edu.cn,
cuijiguang1997@163.com, shaolihuang@tencent.com, xinchao@nus.edu.sg

Abstract

Preserving geometric structures in the scene plays a vital role in image stitching. However, most of the existing methods ignore the large-scale layouts reflected by straight lines or curves, decreasing overall stitching quality. To address this issue, this work presents a structure-preserving stitching approach that produces images with natural visual effects and less distortion. Our method first employs deep learning-based edge detection to extract various types of large-scale edges. Then, the extracted edges are sampled to construct multiple groups of triangles to represent geometric structures. Meanwhile, a GEometric Structure preserving (GES) energy term is introduced to make these triangles undergo similarity transformation. Further, an optimized GES energy term is presented to reasonably determine the weights of the sampling points on the geometric structure, and the term is added into the Global Similarity Prior (GSP) stitching model called GES-GSP to achieve a smooth transition between local alignment and geometric structure preservation. The effectiveness of GES-GSP is validated through comprehensive experiments on a stitching dataset. The experimental results show that the proposed method outperforms several state-of-the-art methods in geometric structure preservation and obtains more natural stitching results. The code and dataset are available at https://github.com/flowerDuo/GES-GSP-Stitching.

1. Introduction

With the popularity of multimedia devices such as smartphones and digital cameras, the requirement to obtain high-quality panoramic images is increasing [9, 20]. Although, image stitching [24] has made tremendous progress, it is still a challenge to produce high-quality panoramic images [23] due to the wide baseline, large parallax, and low-texture under complex stitching scenes [24].

The overall naturalness is an important factor affecting the quality of image stitching. The existing stitching methods can be roughly classified into single feature-based alignment and multiple features-based alignment. The former relies on the homography transformation estimated by point features. The AutoStitch [1] uses global homography transformation for image mapping, but it cannot handle multi-plane scenes. In the dual homography warp (DHW) [4], the scene is simply considered composed of a perspective plane and a ground plane. Furthermore, the as-projective-as-possible (APAP) warps [27] divides the image into meshes and estimates a set of smooth transformations for each grid to improve local alignment. The robust elastic warping (ELA) [10] applies the Bayesian model to improve local alignment for images with parallax. However, the use of only one or more homographic transformations may result in excessive perspective transformation and affect the overall naturalness of the stitching result. Therefore, some works such as the smoothly varying affine (SVA) stitching [16], the shape-preserving half-projective (SPHP) [2], the adaptive as-natural-as-possible (AANAP) warps [14] and the global similarity prior (GSP) model [3] try to obtain more natural stitching results by exploring the advantage of local or global similarity transformation.

On the other hand, joint alignment of point features and line features [8, 12, 13, 25] at the same time can better estimate homography transformation attributed to its strong constraint for image stitching. Li *et al*. first proposed dual-feature including point and line features for warping-based motion model estimation (DFW) [12] to handle the lack of features during stitching. The single-perspective warp (SPW) [13] solves projective distortion to a certain extent, and Xiang *et al*. [25] proposed a line-guided local warping method with a global similarity constraint. Jia *et al*. [8] proposed to leverage line-point consistence (LPC) to preserve structures, which introduces global collinear structures to enhance the desired characters for image warping. Then, Zhang *et al*. [30] applied LPC to the regularization [6] of the stitched result.

Generally, the feature registration method mainly de-

†Corresponding author: Jifeng Ning, njf@nwsuaf.edu.cn

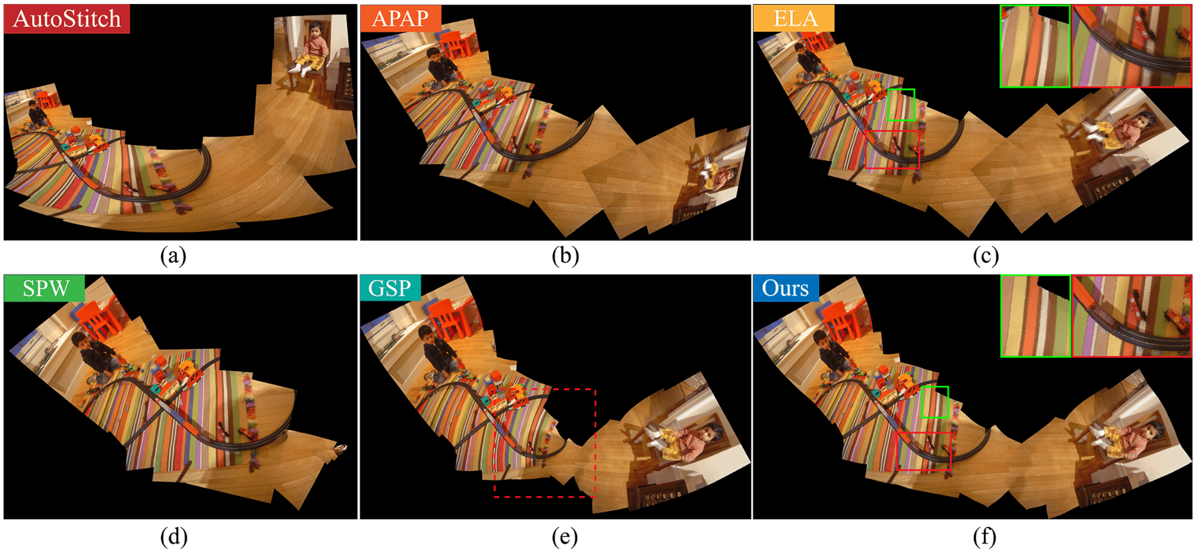

Figure 1. An example of stitching 10 images. (a) The AutoStitch's [1] result is severely distorted. (b) The person on the right side is distorted in the APAP's [27] result. (c) Several misalignments (red and green closeup) in the ELA's [10] result. (d) The SPW's [13] result exhibits significant wrong scale at the right end. (e) There are some distortions, *e.g.*, the floor and carpet in the red box become curved in the result obtained by GSP [3]. (f) Our result preserves the salient geometric structures in the scene.

pends on more homography transformations and local or global similarity constraints to improve stitching quality. However, the *curve constraints* effectively reflecting the critical structure of the scene are not considered. Although the stitching method based on dual features [8, 13, 25] can better keep the scene structure compared with the method based on a single feature, it needs to find the corresponding relationship between the point and the line features among the stitched images. Also, when it is extended to multiple images, finding the correspondence becomes more complicated. Besides, the geometric structure of the scene contains not only straight lines but also curves with better integrity. Interestingly, the seam-guided local alignment for parallax-tolerant image stitching (Seagull) [15] performs similarity transformation on the edges to improve stitching quality [5], indicating the importance of geometric structure information.

In this work, curve structures are introduced into single feature-based stitching model to preserve the geometric structure better than the methods based on point and line features [8, 12, 13, 25]. Different from Seagull [15] only handling a local alignment near the seam line, this work not only ensures the alignment accuracy but also protects the geometric structure in all overlapping and non-overlapping regions. Our method is built on GSP [3] because of the whole naturalness of its stitching result. This work designs a special scheme to protect line and curve structures in overlapping regions and non-overlapping regions to obtain more natural stitching results.

Figure 1 illustrates the stitching result of a challenging example in [19]. It can be seen that the stitching result of AutoStitch [1] is severely distorted. In the stitching result of APAP [27], the person on the right side is distorted. In the stitching result of ELA [10], the salient geometric structures are not destroyed, but there are several misalignments (red and green closeup). In the stitching result of SPW [13], there is a significant wrong scale at the right end. GSP [3] solves the problem of limited field of view, but there are distortions in the local geometric structure, *e.g.*, the floor and carpet in the red box become curved and the toy track is too small in scale than the real scene. The stitching result of our method maintains the geometric structures of carpets, toys, and floors well, and it also looks more natural locally and globally. Our method performs significantly better than AutoStitch, APAP, SPW, and GSP and slightly better than ELA. Generally, the contributions of this work are summarized as follows:

- This work fully utilizes the line and edge features extracted from the stitched images to represent the large-scale geometric structure to obtain high-quality paranoiac images.
- A geometric structure-preserving energy term is added to the GSP stitching model, and the weights of the sampling points on geometric structures are set reasonably to ensure a smooth transition between local alignment and geometric structure preservation to achieve natural stitching results.
- The experiments on 50 sets of images demonstrate that

the proposed method outperforms several state-of-the-art structure-preserving methods.

2. Related Work

Mainly for Alignment. It can usually be considered in terms of global alignment and seam-guided stitching. DHW [4] divides the scene into a perspective plane and a ground plane and represents the transformation model through a combination of two homography transformations. SVA [16] aligns images using affine transformation and smooth transformation to enhance the local alignment accuracy. APAP [27] improves the ability of local alignment by estimating a set of smooth transformations for each grid. ELA [10] applies the Bayesian model to remove incorrect local matches to improve alignment. The above methods belong to single feature-based alignment. In addition, some works [8, 12, 13, 25] improve the stitching quality by using point and line features and the estimate transformation matrix from different aspects when the images lack point features.

Meanwhile, the seam-guided image stitching method only needs to find a parallax-free local region to perform partial alignment and then produce a good seam for stitching, which is more advantageous for large parallax image stitching. Gao *et al*. [5] selected a subset of sparse feature matches that facilitate finding local regions for stitching. Zhang and Liu [28] used an efficient randomized feature selection algorithm to hypothesize homography candidates to obtain good stitching seams. Seagull [15] uses the estimated seam to guide the process of optimizing local alignment so that the seam quality can be improved iteratively.

Mainly for Naturalness. SPHP [2] divides the image into three parts so that the non-overlapping areas have less distortion. AANAP [14] stitches better than SPHP by combining the local homography and the global similarity transformation. GSP [3] uses mesh vertex alignment as an alignment term and combines the local similarity term and global similarity term to optimize the mesh deformation. Li *et al*. [11] proposed a quasi-homography warp to balance perspective distortion and projective distortion in the non-overlapping region. Zhang *et al*. [29] proposed a warp to produce an orthogonal projection of a wide-baseline scene.

Liu *et al*. [17] implemented content-preserving by constraining grid similarity transformation [7]. Guided by dual-feature, DFW [12], SPW [13] and LPC [8] achieves a good image stitching effect under the lack of feature points and solves projective distortion to a certain extent. LPC introduces global collinear structures to balance the desired characters for image warping, which can preserve both local and global straight-line structures. However, the methods guided by dual-feature [8, 12, 13] need to accurately find the corresponding line features, which is challenging for a scene with parallax [15]. Moreover, they ignore the *curve structure* in the scene, which is more common than the straight-line structure in complex real scenes.

In general, the preservation of curve structure in the scene is less studied, although curve structures are more common than straight-line structures. In this work, a stitching model is proposed to preserve different types of geometric structures extracted from images to obtain high-quality stitching results.

3. The Proposed Method

In this section, the limitation of GSP [3] is first analyzed, and then the proposed method is presented in detail, including large-scale edge extraction, geometric structure-preserving term, and our stitching model.

3.1. Limitation of GSP Stitching Method

The GSP [3] is a stitching method based on mesh optimization [7], and it constructs an energy function with multiple constraints. Let V_i and E_i denote the set of vertices and edges in the mesh for the image I_i, respectively. V denotes the set of all vertices in all images. The GSP method attempts to find a set of deformed vertex positions $\hat{V}$ such that the energy function $\psi(\hat{V})$ is minimized. The energy function consists of three terms: the alignment term $\psi_a(\hat{V})$, the local similarity term $\psi_l(\hat{V})$, and the global similarity term $\psi_g(\hat{V})$. It is defined as

$$\hat{V} = \arg\min_{\hat{V}} \psi_a(\hat{V}) + \lambda_l \psi_l(\hat{V}) + \psi_g(\hat{V}). \tag{1}$$

In Equation (1), alignment term $\psi_a(\hat{V})$ ensures the alignment accuracy after image transformation, local similarity term $\psi_l(\hat{V})$ ensures that each grid undergoes similarity transformation, and the global similarity term $\psi_g(\hat{V})$ ensures that each image undergoes a whole similarity transformation to obtain a natural stitching result. Please refer to [3] for the details.

It can be seen that the local similarity term and global similarity term of the GSP [3] method protect the scene structure to a certain extent. Also, GSP takes the grid as the optimization unit, and the local geometric structure in each grid can be protected. However, *when a local geometric structure crosses multiple grids, its structure may be destroyed because the transformation of each grid is different*. As shown in Figure 1(e), the lack of large-scale geometric structure preservation affects the naturalness of the result compared with our stitching result shown in Figure 1(f) with geometric structure constraints.

3.2. Large-scale Geometric Edge Extraction

There are obvious edge structures such as straight lines and smooth curves in real scenes, and smooth curves are more common. If these salient structures are deformed during stitching, the naturalness of the stitching result could not be guaranteed. Therefore, extracting and protecting large-scale edges in images have an important effect on stitching performance.

The geometric structure of an object is not entirely composed of straight lines, and the curve is also an important part. There are a large number of short lines and relatively complete curve structures in the scene shown in Figure 2(a). So, ignoring the complete geometric structure of curves in the image warping will lead to non-optimal results. Although image stitching with content-preserving has been studied [6, 8, 13, 17, 30], the image structure extraction methods in these works do not make full use of the non-linear structure in the image, and the importance of geometric structural integrity is ignored.

Inspired by the progress of deep learning in computer vision, we use the Holistically-nested Edge Detection (HED) [26] method based on the Convolutional Neural Network (CNN) to extract the large-scale edges of images. HED resolves the challenging ambiguity in edge and object boundary detection, and it filters many weak edges and highlights the contour edge structures of the objects. Then, after the branches are cut and the corner disconnection is processed, the outline edge structures representing the geometric structure of the scene initially are obtained.

The large-scale edges of the scene extracted from Figure 2(a) by HED [26] are shown in Figure 2(b), which represent the obvious curves and lines in the image. However, there are still discontinuities in the edges and partial missing of line structures. To solve this problem, we add the lines extracted by LSD [22] into the obtained edge structures. The large-scale edges of the scene are further refined by post-processing such as reconnection of broken lines and collinear constraints to improve the integrity of geometric structures of objects. As shown in Figure 2(c), the geometric structures of most objects are well represented by several concise and complete contours.

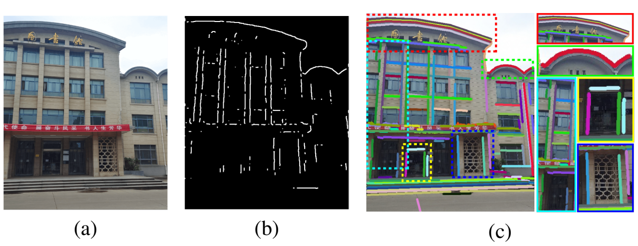

Figure 2. An example of large-scale edge extraction. (a) The original image. (b) The extracted large-scale edge structure based on HED [26]. (c) After the line structure is added to (b), the more concise and complete geometric structure is obtained through line reconnection and collinear processing where the line segment of the same color belongs to the same geometric structure. The curve structure of the roof is completely extracted and the geometric structure of the door, windows, horizontal, and longitudinal structures of the building are well represented by several contours.

3.3. Geometric Structure Preservation Based on Triangle-sampling

After the continuous large-scale edge structure reflecting the geometric structure in the image is obtained, it should be represented effectively in the stitching model. Similar to [15], the sampling points are set equidistantly on the geometric edges. Each sampling point forms a triangle with the endpoint of the geometric structure. The transformation of the geometric structure is indirectly constrained by the similarity transformation constraint [7] on the triangle corresponding to each sampling point, which is also suitable for straight-line structures. As shown in Figure 3(a), a continuous curve representing a geometric structure is sampled at equal intervals, and a group of triangles is formed by the two endpoints and sampling points on the curve. Obviously, if these triangles undergo similarity transformation only in image warping, then the geometric structure can also be effectively protected, as shown in Figure 3(d).

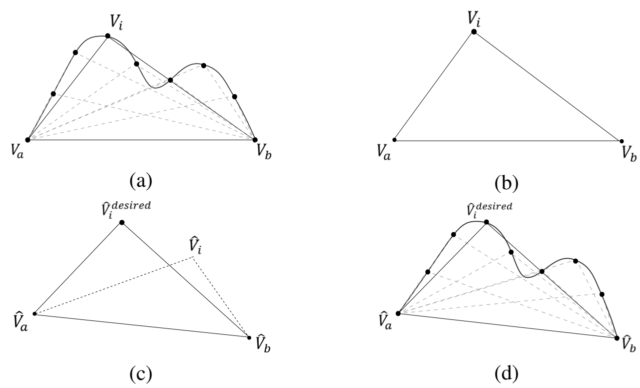

Figure 3. Illustration of triangle sampling in a curve edge. (a) A curve edge is sampled equidistantly. (b) A triangle in (a). (c) The triangle in (b) undergoes a similarity transformation and projection transformation to triangle $(\hat{V}_i^{desired}, \hat{V}_a, \hat{V}_b)$ and triangle$(\hat{V}_i, \hat{V}_a, \hat{V}_b)$ respectively. (d) The same similarity transformation is performed for all triangles in (a) to obtain the similarity transformation result of this contour, then the structure can be protected, which is available for the protection of line structure.

In Figure 3(a), the coordinates of V_i can be represented by the other two endpoints V_a,V_b and (u_i, h_i),

$$V_i = V_a + u_i\left(V_b - V_a\right) + h_i \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \left(V_b - V_a\right), \quad (2)$$

where (u_i, h_i) are the known coordinates within the local coordinate system, and its value will not change after the triangle undergoes similarity transformation [7]. (u_i, h_i) can be calculated by three vertex coordinates,

$$h_i = H\left(\frac{\|(V_i - V_a)\times(V_b - V_a)\|}{\|V_b - V_a\|}\right), u_i = U\left(\frac{\sqrt{\|V_i - V_a\|^2 - {h_i}^2}}{\|V_b - V_a\|}\right) \quad (3)$$

where

$$H(x)=\begin{cases} x & if\ (V_i-V_a)\times(V_b-V_a)<0 \\ -x & else \end{cases}$$
$$U(x)=\begin{cases} -x & if\ \frac{(V_i-V_a)\cdot(V_b-V_a)}{\|V_i-V_a\|\|V_b-V_a\|}<0 \\ x & else \end{cases} \quad . \tag{4}$$

As shown in Figure 3(c), supposing V_i, V_a, V_b are transformed into $\hat{V}_i, \hat{V}_a, \hat{V}_b$ and $\hat{V}_i^{desired}, \hat{V}_a, \hat{V}_b$ after the projection transformation and the similarity transformation, respectively. $\hat{V}_i^{desired}$ is calculated as,

$$\widehat{V}_i^{desired}=\widehat{V}_a+u_i\left(\widehat{V}_b-\widehat{V}_a\right)+h_i\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\left(\widehat{V}_b-\widehat{V}_a\right). \tag{5}$$

To indirectly protect the structure as much as possible, one can minimize the following cost to encourage similarity transformation on a given triangle,

$$E_i=\left\|\widehat{V}_i-\widehat{V}_i^{desired}\right\|^2. \tag{6}$$

The error equation of all the geometric structure sampling points can be obtained as,

$$\psi_{ges}(\widehat{V})=\sum_{j=1}^{N_c}\sum_{i=1}^{N_s}E_i^j, \tag{7}$$

where N_c is the total number of all geometric structures in the image, and N_s is the number of all sampling points in the geometric structure i.

3.4. Geometric Structure Preservation based GSP Stitching Method

It is found that the improvement effect by adding Equation (7) as a new constraint term to the GSP [3] model is not good. This is because the same similarity transformation is used for each sample triangle regardless of whether they are in overlapping or non-overlapping regions. After adding geometric structure constraints, in the non-overlapping area, strong geometric constraints need to be maintained; in the overlapping area, geometric constraints need to be maintained, and image alignment needs to be ensured. Therefore, the weights for the sampling points on geometric edges need to be set reasonably and will play an important role in keeping the best balance between image deformation and structure preservation.

To ensure the overall naturalness of the stitching result, the weights of the sampling points in the non-overlapping area and overlapping area are set to 1 and less than 1, respectively. Figure 4(a) shows an example of the weight map of sampling points when two images are pre-stitched. P_0, P_1, and P_2 denote the three sampling points on a curve structure in the figure on the right (green). P_0 is located in the non-overlapping area of the right image, and the geometric structure constraints are completely dominated by the right image, so the weight of P_0 is 1. P_1 tends to concentrate on the center position of the right image, but it is located at the edge of the left image (red). Thus, it is considered that the geometric structure constraints at the position of P_1 are dominated by the right image. And similarly, the weight of P_2 is less than that of P_1.

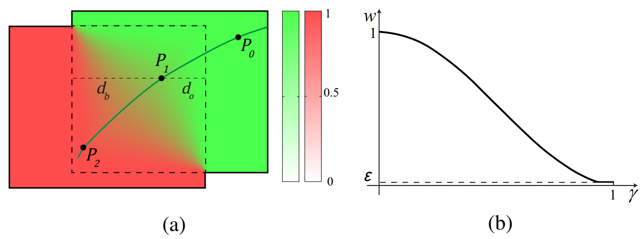

Figure 4. (a) Illustration of the weight map of the sampling points in a geometric curve for two stitched images. (b) The function curve of the weight equation.

Therefore, we calculate the weight by using the minimum distances from the sampling points to the overlapping region boundary and the image boundary, respectively. Note that when the overlapping region of two images is a square, an optimal solution is obtained, and it is expected that the sum of the weights of the two sampling points in the same position in the overlapping region is 1. So, we use a centrosymmetric function, *e.g.* cosine (Figure 4(b)), and the formula for calculating the weight of the sampling points in the overlapping area is

$$w_i=\max\left(\frac{1}{2}\left(\cos\left(\gamma_i\times\pi\right)+1\right),\varepsilon\right), \tag{8}$$

where $\gamma_i=d_o\left(V_i\right)/\left(d_b\left(V_i\right)+d_o\left(V_i\right)\right)$, $d_o\left(V_i\right)$ is the minimum distance between the grid where the sampling point is located in and the boundary of the overlapping region, and $d_b(V_i)$ is the minimum distance between the grid where the sampling point is located in and the image boundary, as shown in Figure 4(a). In all of our experiments, ε=0.01.

Finally, geometric structure preservation is integrated into the GSP model to obtain our GES-GSP stitching model,

$$\widehat{V}=\underset{\widehat{V}}{\operatorname{argmin}}\left(\psi_a\left(\widehat{V}\right)+\lambda_l\psi_l\left(\widehat{V}\right)+\psi_g\left(\widehat{V}\right)+\boldsymbol{\lambda_{ges}\psi_{ges}}\left(\widehat{V}\right)\right) \tag{9}$$

where, $\psi_{ges}(\widehat{V})=\sum_{j=1}^{N_c}\sum_{i=1}^{N_s}w_i^jE_i^j$. In all of our experiments, $\lambda_l=0.75$, λ_{ges}=1.5.

Our stitching method attempts to find a set of deformed vertex positions $\hat{V}$ such that the total energy term is minimized. Locally, each grid undergoes similarity transformation so that the panoramic image has better detail information. Globally, the appropriate scale and rotation are found for each image to maintain a good structure. Geometrically, the salient geometric structures in the image are protected from distortion. Thus, the energy function consists of four terms: the alignment term $\psi_a(\widehat{V})$, the local similarity term $\psi_l(\widehat{V})$, the global similarity term $\psi_g(\widehat{V})$ and the geometry-preserving term $\psi_{ges}(\widehat{V})$.

The proposed method can be regarded as a linear optimization problem, and the optimal mesh vertex set can be obtained as a closed-form solution by solving a sparse linear matrix equation. The detailed process is given in the supplementary material.

4. Experiments

The experiments were performed on a computer equipped with a 2.9GHz CPU and 16GB memory, and running Windows10 and Ubuntu18 operating systems. SIFT [18] features are extracted using VLFeat [21], and edges are extracted by HED [26] and LSD [22]. The grid size is 40×40 pixels for mesh-based methods. By default, the minimum sampling point interval is set as the grid size, and as many sampling points as possible are obtained according to the total length of the geometric structure.

To comprehensively test the effect and stability of the proposed method, we constructed 50 diversified and challenging datasets (26 from [2–4, 8, 14, 19, 27] and 24 collected by ourselves). The numbers of images range from 2 to 35, and the spatial relations among the images are 1D and 2D. Compared with GSP [3], our method takes some time on image pre-processing, but the time consumption is still acceptable. For the resolution of 800×600, the GSP method takes 2.37 s for stitching two images (Figure 5) and 20.27 s for stitching 21 images (Figure 7), while the proposed method takes 4.418s and 31.168s, respectively.

Due to the limited space, please see the supplementary material for more detailed comparisons and discussions.

4.1. Comparison with the State-of-the-art Methods

Our GES-GSP method is compared with AutoStitch [1], APAP [27], ELA [10], SPW [13], LPC [8] (Only supports stitching two images) and GSP [3] on a variety of real scenes.

Figure 5 illustrates the result of stitching two images in the scene of the office (data from LPC [8]). In the result obtained by APAP, SPW, and LPC, the door exhibits artifacts with obliqueness and non-uniform deformation (blue box). In the result obtained by APAP and SPW, the clock exhibits misalignment (red closeup). In the result obtained ELA, there are distortion and misalignment (red circle). In the result obtained by LPC, the door exhibits artifacts, and the overall perspective is not natural. In the results obtained by GSP, the wall indicated by the red, yellow, and green lines is bent and does not conform to the visual perspective effect. In the result obtained by our GES-GSP, neither the table nor the door on the left is stretched, the wall is not bent, and the overall visual effect is more natural.

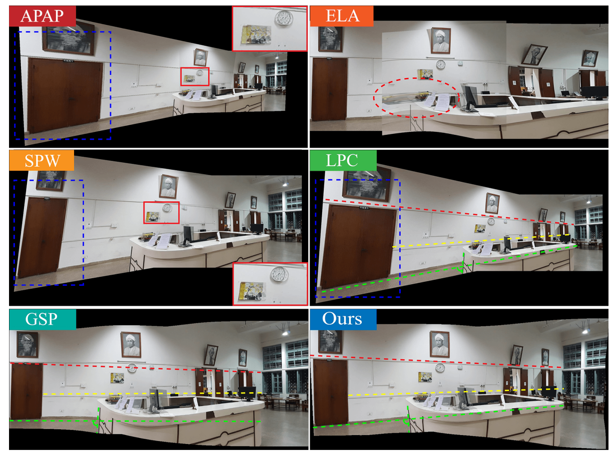

Figure 5. An example of stitching two images in the scene of the office.

Figure 6 shows the result of stitching five images in the scene of the mall with smooth curve geometry. To better represent the differences between the results, green curves are added to the results with the same curvature, which are closer to the real scene. In the results obtained by AutoStitch and GSP, as indicated by the green curve, the arc at the top exhibits an unnatural expansion, and the part indicated by the red line is bent. In the results obtained by APAP, ELA, and SPW, the arc at the top is significantly shrunk, and the parts of the image are severely oblique and stretched (blue box) due to over-projection. In the result obtained by our method, the scale and radian of the top circle are more natural, and the geometric structures of other salients are also well protected in warping.

Figure 7 shows the result of stitching 21 images, and the spatial relations among them are 2D (data from GSP [3]). In the result obtained by AutoStitch, there are obvious distortions in the horizontal and vertical directions of the building (red box). In the result obtained by ELA, there is distortion caused by the spherical projection. There are misalignments in the result obtained by APAP and SPW (red circle). In the result obtained by GSP, the edges of the buildings in the middle (red, green, yellow lines) are bent, and the billboard on the right (red box) is distorted into a fan shape. In the result obtained by our method, the artifacts of the building (red, green, yellow lines) are well preserved, and the geometric structure of the billboard on the right is protected.

In summary, AutoStitch [1] and ELA (spherical projection) [10] suffer from the distortion caused by the spherical

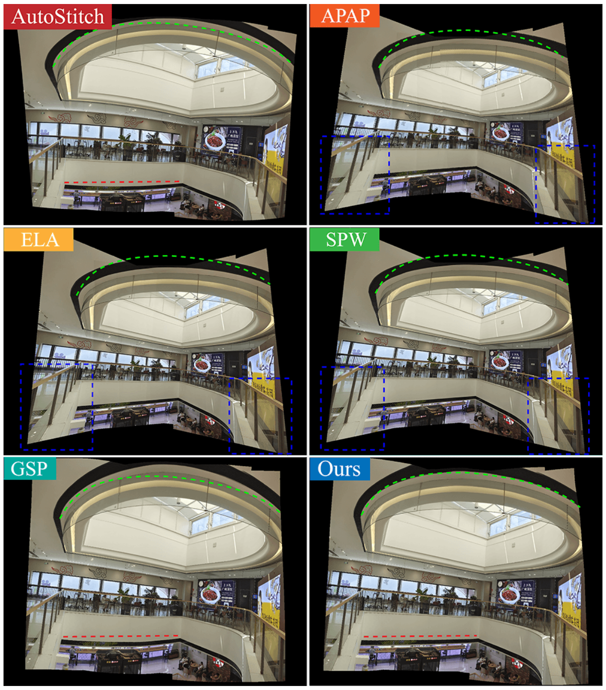

Figure 6. An example of stitching five images in the scene of mall.

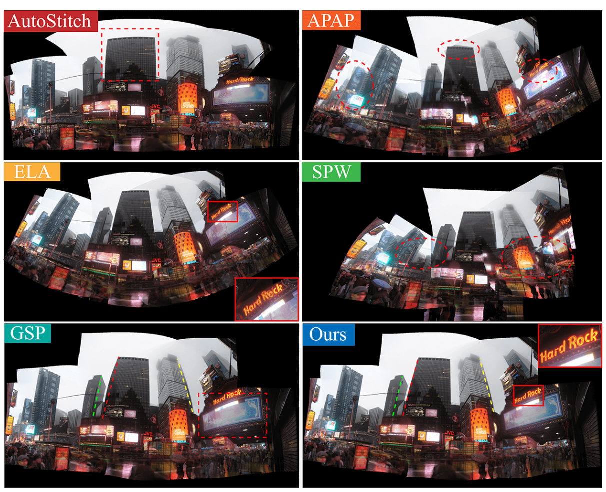

Figure 7. An example of stitching 21 images in the scene.

projection. APAP [27] and ELA [10] suffer from severe shape and area distortions, especially in non-overlapping areas. SPW [13] and LPC [8] protect line structures and suppress distortions, and the shape and area are stretched and non-uniformly enlarged to a certain extent. However, the result is still not very good because using the straight-line feature only is not enough for the complex real scene. GSP [3] selects the proper scale and rotation for each image. Though it solves the above problems nicely, the geometric structures in the image are destroyed.

As for our method, the type of structure to be preserved is not limited, so it can maintain different geometric structures extracted and obtain the best balance between image deformation and structure preservation. Therefore, the results obtained by our method are more natural.

4.2. Discussions

In this section, the influence of sampling interval and the weight of sampling point and the evaluation of distortion on the stitching result are analyzed.

4.2.1 Sampling Interval

As mentioned above, geometric structure often crosses multiple grids. Therefore, we set different numbers of sampling points for each grid on the geometric structure to evaluate the influence on geometric structure preservation. Specifically, five schemes are set up, *e.g*. sampling four points, two points, and one point for every grid, and sampling one point every two grids and every four grids, respectively.

As shown in Figure 8, there are approximately horizontal edges spanning multiple grids on the building. When the number of sampling points is greater than one for each grid, a better geometric structure preservation effect can be obtained; otherwise, the geometric structure preservation effect will be reduced. Then, to simplify the sampling process, we only sample one point per grid in the experiment.

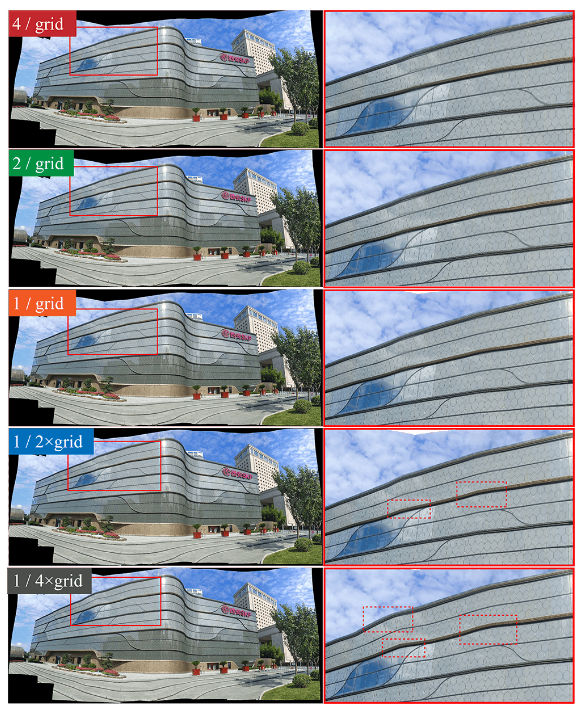

Figure 8. An example of geometric structure preservation under different sampling schemes.

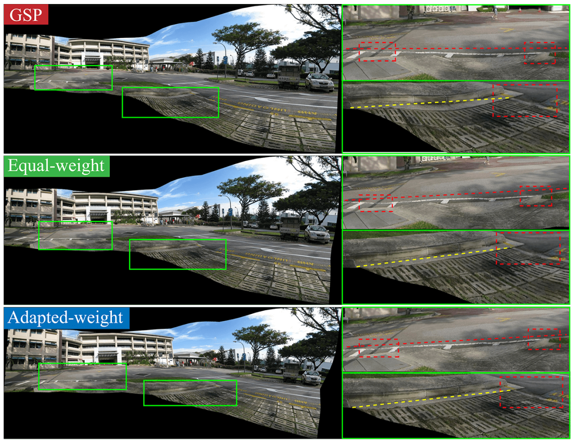

Figure 9. An example of the influence of with/without adapted weight.

4.2.2 The Weight of Sampling Point

Setting the weight of the sampling point on geometric structures is crucial for alignment accuracy and naturalness during stitching. The influence of the adapted-weight and equal-weight is discussed here.

Figure 9 shows an example of three stitching methods including GSP [3], the proposed method with the equal-weight and the adapted-weight. The result obtained by GSP exhibits distortion on the ground (green closeup). The proposed method with equal weight has a certain effect (yellow line), but the part in the red box is bent compared with the corresponding part of the GSP. Such shape bending is due to the same weight of sampling points on the geometric structure, which makes the stitching in the overlapping region difficult to balance between local alignment and geometric structure preservation. Finally, as shown in the red box and line, our method can obtain more natural stitching results with less distortion via adapted weight for each sampling point.

4.2.3 Comparison with GSP in Geometric Structure Preservation

Two examples of geometric structure preservation by our method and the GSP [3] method are illustrated here. Figure 10 shows the local deformation of the original image in Figure 6. The mesh deformation demonstrated by the curve parts (red box) in the image indicated that the result obtained by the GSP method without geometric structure preservation exhibits arched bending. Meanwhile, the distortion in the result obtained by our GES-GSP is much less than that of the result obtained by GSP. Similarly, in Figure 11, the mesh in the vertical street lamp (red box) shows obvious bending, and the bridge (green box) is also arched in the result obtained by GSP.

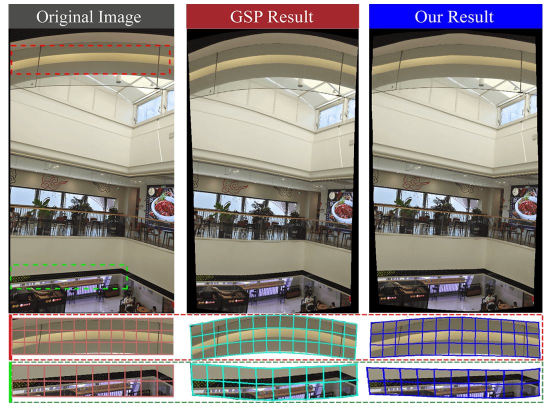

Figure 10. The warping of an image from Figure 6.

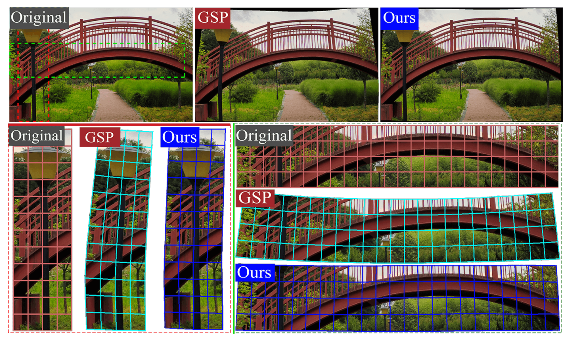

Figure 11. The warping of an image with obvious curve structures.

5. Conclusion

This paper proposes an image stitching method guided by geometric structure not limited to specific edge types. First, the deep learning-based large-scale edge detection method and the traditional line detection method are used to extract various types of edges reflecting the structure information of the scene. Then, triangle sampling is performed on the structures to obtain a set of triangles representing the corresponding structure. Finally, the obtained triangles are used to construct a geometric structure preservation term to perform similarity transformation for content preservation.

In the proposed GES-GSP method, the adapted weights for sampling points balance between the alignment and geometric structure preservation to obtain more natural stitching results. Compared with the state-of-the-art methods, the proposed GES-GSP can preserve different types of geometric structures as much as possible, thus obtaining a high-quality panoramic image. In the future, we will explore the spatial constraints between different geometric structures and further obtain more compelling stitching results.

Acknowledgements. This project is supported by the National Natural Science Foundation of China under Grant No.:61876153 and NUS Faculty Research Committee (FRC) Grant (WBS:A-0009440-00-00).

References

[1] Matthew Brown and David G Lowe. Automatic panoramic image stitching using invariant features. *International journal of computer vision*, 74(1):59–73, 2007. 1, 2, 6

[2] Chehan Chang, Yoichi Sato, and Yung-Yu Chuang. Shape-preserving half-projective warps for image stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3254–3261, 2014. 1, 3, 6

[3] Yusheng Chen and Yung-Yu Chuang. Natural image stitching with the global similarity prior. In *Proceedings of the European conference on computer vision*, pages 186–201. Springer, 2016. 1, 2, 3, 5, 6, 7, 8

[4] Junhong Gao, Seon Joo Kim, and Michael S Brown. Constructing image panoramas using dual-homography warping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 49–56. IEEE, 2011. 1, 3, 6

[5] Junhong Gao, Yu Li, Tat-Jun Chin, and Michael S Brown. Seam-driven image stitching. In *Eurographics (Short Papers)*, pages 45–48, 2013. 2, 3

[6] Kaiming He, Huiwen Chang, and Jian Sun. Rectangling panoramic images via warping. *ACM Transactions on Graphics*, 32(4):1–10, 2013. 1, 4

[7] Takeo Igarashi, Tomer Moscovich, and John F Hughes. As-rigid-as-possible shape manipulation. *ACM transactions on Graphics*, 24(3):1134–1141, 2005. 3, 4

[8] Qi Jia, Zhengjun Li, Xin Fan, Haotian Zhao, Shiyu Teng, Xinchen Ye, and Longin Jan Latecki. Leveraging line-point consistence to preserve structures for wide parallax image stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12186–12195, 2021. 1, 2, 3, 4, 6, 7

[9] Kyu-Yul Lee and Jae-Young Sim. Warping residual based image stitching for large parallax. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8198–8206, 2020. 1

[10] Jing Li, Zhengming Wang, Shiming Lai, Yongping Zhai, and Maojun Zhang. Parallax-tolerant image stitching based on robust elastic warping. *IEEE Transactions on multimedia*, 20(7):1672–1687, 2017. 1, 2, 3, 6, 7

[11] Nan Li, Yifang Xu, and Chao Wang. Quasi-homography warps in image stitching. *IEEE Transactions on Multimedia*, 20(6):1365–1375, 2017. 3

[12] Shiwei Li, Lu Yuan, Jian Sun, and Long Quan. Dual-feature warping-based motion model estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4283–4291, 2015. 1, 2, 3

[13] Tianli Liao and Nan Li. Single-perspective warps in natural image stitching. *IEEE Transactions on Image Processing*, 29:724–735, 2019. 1, 2, 3, 4, 6, 7

[14] Chung-Ching Lin, Sharathchandra U Pankanti, Karthikeyan Natesan Ramamurthy, and Aleksandr Y Aravkin. Adaptive as-natural-as-possible image stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1155–1163, 2015. 1, 3, 6

[15] Kaimo Lin, Nianjuan Jiang, Loong-Fah Cheong, Minh Do, and Jiangbo Lu. Seagull: Seam-guided local alignment for parallax-tolerant image stitching. In *European conference on computer vision*, pages 370–385. Springer, 2016. 2, 3, 4

[16] Wenyan Lin, Siying Liu, Yasuyuki Matsushita, Tian-Tsong Ng, and Loong-Fah Cheong. Smoothly varying affine stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 345–352. IEEE, 2011. 1, 3

[17] Feng Liu, Michael Gleicher, Hailin Jin, and Aseem Agarwala. Content-preserving warps for 3d video stabilization. *ACM Transactions on Graphics*, 28(3):1–9, 2009. 3, 4

[18] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. 6

[19] Yoshikuni Nomura, Li Zhang, and Shree K Nayar. Scene collages and flexible camera arrays. In *Proceedings of the Eurographics conference on Rendering Techniques*, pages 127–138, 2007. 2, 6

[20] Richard Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 2(1):1–104, 2006. 1

[21] Andrea Vedaldi and Brian Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1469–1472, 2010. 6

[22] Rafael Grompone Von Gioi, Jérémie Jakubowicz, Jean-Michel Morel, and Gregory Randall. Lsd: a line segment detector. *Image Processing On Line*, 2:35–55, 2012. 4, 6

[23] Zhaobin Wang and Zekun Yang. Review on image-stitching techniques. *Multimedia Systems*, 26(4), 2020. 1

[24] Lyu Wei, Zhou Zhong, Chen Lang, and Zhou Yi. A survey on image and video stitching. *Virtual Reality & Intelligent Hardware*, 1(1):55–83, 2019. 1

[25] Tianzhu Xiang, Guisong Xia, Xiang Bai, and Liangpei Zhang. Image stitching by line-guided local warping with global similarity constraint. *Pattern Recognition*, 83:481–497, 2018. 1, 2, 3

[26] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1395–1403, 2015. 4, 6

[27] Julio Zaragoza, Tat-Jun Chin, Michael S Brown, and David Suter. As-projective-as-possible image stitching with moving dlt. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2339–2346, 2013. 1, 2, 3, 6, 7

[28] Fan Zhang and Feng Liu. Parallax-tolerant image stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3262–3269, 2014. 3

[29] Guofeng Zhang, Yi He, Weifeng Chen, Jiaya Jia, and Hujun Bao. Multi-viewpoint panorama construction with wide-baseline images. *IEEE Transactions on Image Processing*, 25(7):3099–3111, 2016. 3

[30] Yun Zhang, Yukun Lai, and Fanglue Zhang. Content-preserving image stitching with piecewise rectangular boundary constraints. *IEEE transactions on visualization and computer graphics*, 27(7):3198–3212, 2020. 1, 4