

Leveraging Line-point Consistence to Preserve Structures for Wide Parallax Image Stitching

Qi Jia¹, ZhengJun Li¹, Xin Fan¹, Haotian Zhao¹, Shiyu Teng¹, Xincheng Ye¹, Longin Jan Latecki²

¹International School of Information Science & Engineering, Dalian University of Technology

²Computer & Information Sciences, Temple University

Abstract

Generating high-quality stitched images with natural structures is a challenging task in computer vision. In this paper, we succeed in preserving both local and global geometric structures for wide parallax images, while reducing artifacts and distortions. A projective invariant, Characteristic Number, is used to match co-planar local sub-regions for input images. The homography between these well-matched sub-regions produces consistent line and point pairs, suppressing artifacts in overlapping areas. We explore and introduce global collinear structures into an objective function to specify and balance the desired characters for image warping, which can preserve both local and global structures while alleviating distortions. We also develop comprehensive measures for stitching quality to quantify the collinearity of points and the discrepancy of matched line pairs by considering the sensitivity to linear structures for human vision. Extensive experiments demonstrate the superior performance of the proposed method over the state-of-the-art by presenting sharp textures and preserving prominent natural structures in stitched images. Especially, our method not only exhibits lower errors but also the least divergence across all test images. Code is available at <https://github.com/dut-media-lab/Image-Stitching>.

1. Introduction

Image stitching, that combines multiple images into a larger image with a wider field of view [25], is widely used in photogrammetry [24], robot navigation [6] and panorama on smart phones [29]. It is still challenging to produce high quality stitched images for the state-of-the-art as they suffer from severely unpleasant effects such as artifacts and distortions, especially for wide parallax images.

Feature matching is the key to aligning multiple images for producing artifact-free stitching as the matched features act as anchors in alignment. The SIFT features [23] are

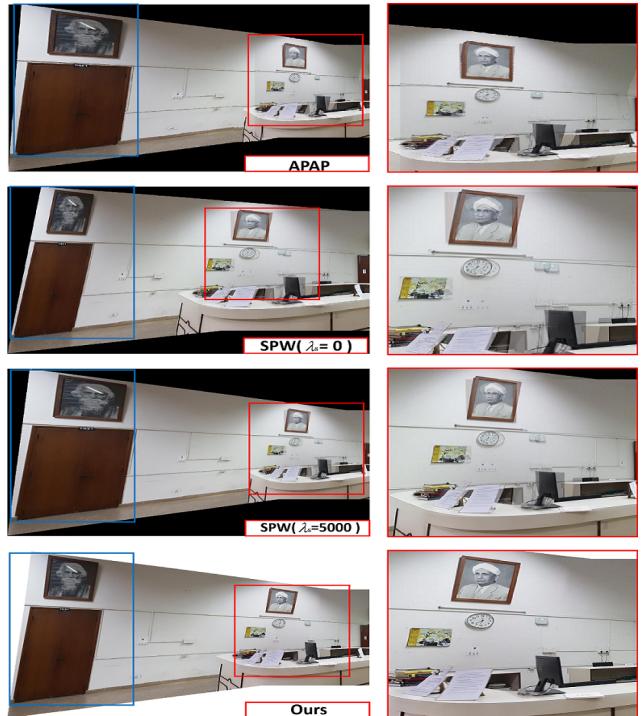


Figure 1: Comparisons of stitching methods. Evident artifacts and distortions appear in the results of the existing methods shown in the zoomed-in rectangles, but ours is free of these unpleasant effects.

widely used in many traditional methods for feature points detection and matching [4, 5, 30]. Some recent works also introduce line features to obtain robust matching in the cases of large parallax and/or low textures where points are prone to mismatched [11]. Lin *et al.* exploit both point and line features by different weights in an objective function [16]. Unfortunately, these methods separately match points and lines, and thus the local surrounding areas may be inconsistently and non-uniformly stretched or compressed when mismatches inevitably occur, presenting artifacts in the stitched images. Liao *et al.* employ the RANSAC strategy to refine point and line pairs by using the homography between images [17]. It is worth noting

that the homography relationship *only* holds for points and lines in the same projective plane [12]. Therefore, those refinements upon the homography but neglecting the co-planar constraint fail to give accurate matches. As shown in the red rectangles of the first three rows in Fig. 1, the magnified overlapping area on the right exhibits artifacts on the picture frame, clock and computer. It is highly desirable to explore co-planar areas and refine corresponding matching pairs of points and lines.

Image stitching has to preserve linear structures while alleviate distortions since human visual perception is very sensitive to these structures. The as-projective-as-possible (APAP) method adopts parametric warps by local constraints [30], but suffers from severe shape distortions especially in non-overlapping areas, as shown in the blue rectangle of the first row in Fig. 1. Shape-preserving half-projective (SPHP) [4] and global similarity prior (GSP) [5] share a similar idea to adapt different warps for different image areas. Geodesic-preserving[13] and line-structure-preserving[3] involve collinearity preservation, but their image resizing takes one panoramic image as input, already including correct global geometric structures as reference. Recently, Liao *et al.* propose single-perspective warps (SPW) [17] to protect linear structures while suppress distortions. These methods can well preserve local structures but fail to resolve the conflict when maintaining both local and global linear structures. Global collinear structures can be either a long line across the major part of an image, *e.g.*, the long line under the two picture frames in Fig. 1, or several separate collinear line segments. Current line detectors [26] cannot detect or connect these long lines. Consequently, local shapes may be well preserved by setting appropriate parameters, but the global linear structure is out of shape in the second row of Fig. 1. In the third row, preserving linear structure results in severe distortions for local shapes. It still remains unresolved to preserve both local and global collinear structures.

Meanwhile, the existing metrics to evaluate the stitching quality are not comprehensive enough. These metrics including the distance between matched points [30] and average geometric error (SSIM) [27] on local patterns of pixel intensities can only quantify performance on point matching. None of them can reflect the alignment of points on linear structures or the collinearity of matched line segments. Quantitatively evaluating the preservation on linear structures for image stitching is also an open issue.

This paper leverages the line and point consistence to preserve linear structures that are essential geometries for image stitching. We divide input images into co-planar regions upon the neighborhoods of lines, and match the regions from different views using a series of geometric invariants reflecting the intrinsic nature of lines and points. Hence, the homography between these co-planar regions

can accurately generate matches of both lines and points. Subsequently, an line-guided objective function for warping is designed to preserve both local and global linear structures and suppress distortions. The fourth row in Fig. 1 demonstrates that our method yields a significant gain in image quality. Moreover, a quantitative evaluation measure for lines is proposed to analyze the quality of stitched images more comprehensively. Our contributions are summarized as follows:

- We design a new matching strategy to obtain consistent point and line pairs by exploring co-planar sub-regions using projective invariants. This matching follows the essential co-planar requirement for homography so that it can provide accurate pre-alignment while eliminating artifacts and non-uniform distortions.
- To the best of our knowledge, we are the first to incorporate global collinear structures as a constraint that significantly alleviates unnatural distortions.
- We propose a comprehensive metric to quantify the preservation of linear structures for image stitching.

We compare the proposed method with the state-of-the-art on challenging natural image pairs with prominent linear structures covering variations on camera motions, scenes and fields of view. Our method can produce visually appealing stitching and our average RMSE for point matching is 31% lower than that of SPW [17]. Meanwhile, ours works the most accurate and stable for preserving the linear structures in terms of the proposed metric. Sections 3, 4 and 5 elaborate our contributions, respectively.

2. Related Works

This paper brings the line-guided image stitching method that preserves both local and global structures. Hence, this section reviews previous works related to warps for less distortion and warps with line structure constraints.

Traditional stitching methods usually estimate an optimal global transformation for each input image. They can only work well for ideal near planar scenes, and the resulting images often suffer from local artifacts and projective distortion [2]. Therefore, some methods try to make warps adaptive to different areas of images. Lin *et al.* [21] propose a smoothly varying affine (SVA) transformation for better local adaptation. Li *et al.* use Bayesian model to remove outliers and the thin plate spline for analytical warp [14]. Gao *et al.* divide image into ground plane and distant plane, and propose a dual-homography warp (DHW) [8] to reduce the distortion. Shape-preserving half-projective (SPHP) warps [4] combine the projection transformation in the overlapping and non-overlapping area. Adaptive as-natural-as-possible (AANAP) warps [18] share similar idea, which transforms the homography transformation in the overlapping area to the whole image. Herrman *et al.*

[9] introduce multiple registrations to capture greater accuracy instead of a single registration. Li *et al.* propose a quasi-homography (QH) warp [15], which relies on a global homography while squeezing the non-overlapping areas. However, they are not flexible enough to decrease distortion for scenes with large parallax.

In order to get better alignment with less distortion, APAP [30] fine-tunes the global homography warp to accommodate location dependent alignment. Chen *et al.* propose a global similarity prior (GSP) based warps by minimizing an energy function consisting of alignment, local and global similarity terms [5]. Their method aims at solving the distortion in non-overlapping area, but linear structures are not well protected. Zhang *et al.* achieve a better performance by setting a series of prior constraints and manual guidance [32]. Lin *et al.* take the difference of pixel intensity into consideration, which works well in low texture images [20]. Lee *et al.* partition images into super-pixels and warp them adaptively using the computed feature matches according to the warping residuals for parallax scenes [12].

In addition, there are some seam-based approaches for less local distortion. A parallax-tolerant warp is proposed that combines homography and content-preserving-warps (CPW) [22] to control distortion [31]. However, their method still leads to shape distortion in large parallax. Lin *et al.* iteratively improves the seam-guided local alignment by adaptive feature weighting and introduces a novel term to preserve salient line structures approach [19]. However, global distortion still exists in non-overlapping areas.

In order to achieve better stitching quality with less distortion and preserving linear structures, Li *et al.* introduce line features into image stitching, which improves content-preserving-warps by introducing linear alignment terms [16]. Xiang *et al.* propose a line-guided local warping with global similarity constraint [28]. Liao *et al.* simultaneously emphasizes different characteristics of the single-perspective warp, including alignment, distortion and saliency [17]. However, global collinear structure has seldom been addressed, and the conflict between local and global structure preserving still exists in these approaches.

3. Pre-alignment Based on Consistent Line-Point Constraints

In this section, a dual-feature (lines and points) based pre-alignment algorithm is designed, which is demonstrated in Fig. 2. First, image is divided into coplanar sub-regions based on line detection, and one of them is illustrated in green rectangle of Fig. 2. Then, sub-regions are matched by the similarity calculated from a series of projective invariants. The third, matched point pairs are increased and refined, and lines are matched by the homography between matched regions. Finally, a global pre-alignment is con-

structed based on dual features.

3.1. Sub-region Division Based on Line Detection

Local homography between coplanar regions is more accurate than global one. As many lines are formed by the intersection of planes, we make a rough assumption that the neighborhood determined by the length of the line can be regarded as a local coplanar sub-region of the image. LSD [26] is used to obtain the original line segments, then the neighborhood of lines is split into the left one and the right one according to the gradient direction, as points located on different sides of a line may not be coplanar. The gradient of a line is defined as the average gradient of all points on it. As shown in Fig. 3, in the neighborhood of a line, the distance from any pixel to the line is less than $\alpha \cdot \text{len}(l)$ and less than $\beta \cdot \text{len}(l)$ to the perpendicular bisector line. In our experiments, α and β are set as 2.0 and 0.5, respectively [10].

3.2. Sub-regions Matching by Line-point Invariant

In order to match coplanar sub-regions and finally more line and point pairs, a projective invariant Characteristic Number (*CN*) is introduced to construct a line-point invariant, and the similarity between sub-regions is defined based on it. The *CN* is defined as follows:

Let \mathbb{K} be a field and $\mathbb{P}^m(\mathbb{K})$ be m-dimension projective space over \mathbb{K} , and $\{P_i\}_{i=1,2,\dots,R}$ be distinct points in $\mathbb{P}^m(\mathbb{K})$ that construct a close loop ($P_{R+1} = P_1$). There are distinct points $\{Q_i^{(j)}\}_{j=1,2,\dots,S}$ on the line segment $\{P_iP_{i+1}\}_{i=1,2,\dots,R}$ such that each point $Q_i^{(j)}$ can be linearly represented by P_i and P_{i+1} as $Q_i^{(j)} = a_i^{(j)}P_i + b_i^{(j)}P_{i+1}$. Let $\mathcal{P} = \{P_i\}_{i=1,2,\dots,R}$ and $\mathcal{Q} = \{Q_i^{(j)}\}_{i=1,2,\dots,R}^{j=1,2,\dots,S}$, then the

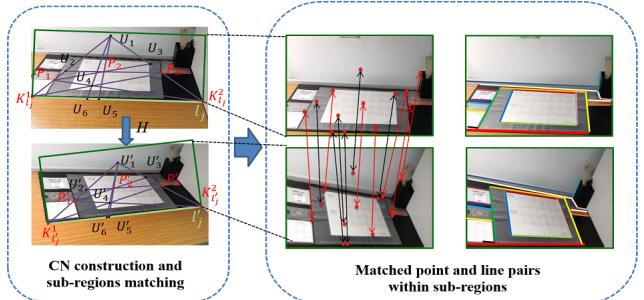


Figure 2: Overview of pre-alignment. In the left box with dotted lines, line-point invariant is illustrated by the red matching points and lines in two different views, labeled as l_j and l'_j . New added point pairs are corresponded by black double-headed arrows, and original point pairs are labeled in red double-headed arrows on the right dotted box. Output matched line pairs are labeled by different colors on right.

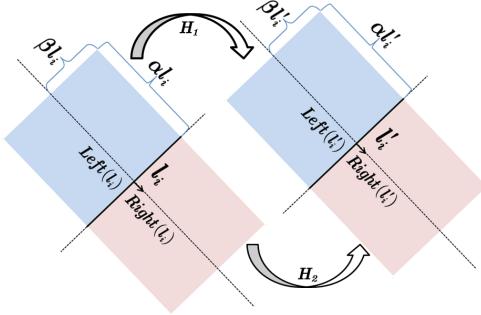


Figure 3: Coplanar sub-regions division and matching.
quantity

$$CN(\mathcal{P}, \mathcal{Q}) = \prod_{i=1}^R \left(\prod_{j=1}^S \frac{a_i^{(j)}}{b_i^{(j)}} \right) \quad (1)$$

is called the Characteristic Number of \mathcal{P} and \mathcal{Q} [10].

As the construction of CN requires a close loop and equal number of points on each edge, we use five points to construct a triangle and equal intersections on each edge [10]. As shown in the upper left image of Fig. 2, K_l^1 and K_l^2 are two endpoints on the red line l . P_1, P_2 and P_3 are three non-collinear feature points on the same side of the line that are marked as red dots. Any three of the points ($K_l^1, K_l^2, P_1, P_2, P_3$) are not collinear.

We denote the line through two points, P_i and P_j , as $\overline{P_i P_j}$ and the intersection of two lines, $\overline{P_i P_j}$, and $\overline{P_k P_m}$, as $\langle \overline{P_i P_j}, \overline{P_k P_m} \rangle$. We can obtain several intersection points (blue points), including $U_1 = \langle \overline{K_l^1 P_1}, \overline{K_l^2 P_3} \rangle$, $U_2 = \langle \overline{K_l^1 P_1}, \overline{P_2 P_3} \rangle$, $U_3 = \langle \overline{P_1 P_2}, \overline{K_l^2 P_3} \rangle$, $U_4 = \langle \overline{K_l^1 P_3}, \overline{P_1 K_l^2} \rangle$, $U_5 = \langle \overline{K_l^1 K_l^2}, \overline{U_1 P_2} \rangle$ and $U_6 = \langle \overline{K_l^1 K_l^2}, \overline{U_1 U_4} \rangle$.

Thus, we have $\triangle K_l^1 U_1 K_l^2$, and we are able to calculate CN with $\mathcal{P} = \{K_l^1, U_1, K_l^2\}$ and $\mathcal{Q} = \{P_1, U_2, U_3, P_3, U_5, U_6\}$. We denote the CN value constructed in this way as $CN(l, P_1, P_2, P_3)$. Thereafter, in the other view shown in the down left figure of Fig. 2, we can construct $\triangle K_l'^1 U'_1 K_l'^2$ in the same way, and $CN(l, P_1, P_2, P_3)$ is equal to $CN(l', P'_1, P'_2, P'_3)$ with corresponding matched line and point pairs. A series of CN values can be obtained by different feature points.

Let I and I' denote the target and reference images, respectively. We use SURF [1] to detect and match features, and use LSD [26] to detect lines. Then, we can calculate the similarity between candidate sub-regions based on a series of CN values within the corresponding regions [10]. For the matching sub-regions $Reg \in I$ and $Reg' \in I'$ with the highest similarity, the existing matching point pairs within the matching region are used to construct CN as demonstrated in the left image of Fig. 2. The intersection points $U_1, U_2, U_3, U_4, U_5, U_6$ on $\triangle K_l^1 U_1 K_l^2$, and the corresponding points $U'_1, U'_2, U'_3, U'_4, U'_5, U'_6$ on $\triangle K_l'^1 U'_1 K_l'^2$ are added into the matching points set to increase the an-

chor points for warping. Finally, we use RANSAC [7] to refine matching points and estimate its local homography H , which is used to obtain more matching lines in sub-regions [10].

3.3. Pre-alignment Based on Dual Features

Let $p_i = (x_i, y_i, 1)$ and $p'_i = (x'_i, y'_i, 1)$ be matched point pairs $\{(p_i, p'_i)\}_{i=1,2,\dots,N}$ in homogeneous coordinates, where N is the number of matched point pairs, for the set of matching line pairs $\{(l_j, l'_j)\}_{j=1,2,\dots,L}$, $l_j \in I$ and $l'_j \in I'$, where L is the number of matching line pairs. Line l_j is represented as (l_j^s, l_j^e) , where l_j^s and l_j^e are two end points. In order to achieve a better registration, the Euclidean distance between matched points and lines after warps should be minimized. We denote H as the initial homography, H_* is the vector expression of H , and \hat{H}_* is the desired homography. Therefore, a global homography based on dual features can be expressed as

$$\begin{aligned} \hat{H}_* &= \arg \min_H \left(\sum_{i=1}^N \|p'_i \times H p_i\|^2 + \sum_{j=1}^L \|dis(l'_j, H l_j^{s,e})\|^2 \right) \\ &= \arg \min_H \left(\sum_{i=1}^N \|\mathcal{U}_i H_*\|^2 + \sum_{j=1}^L \|\mathcal{V}_j H_*\|^2 \right), \end{aligned} \quad (2)$$

where $dis(l'_j, H l_j^{s,e})$ denotes the distance between the endpoints $H l_j^{s,e}$ and line l'_j , $H_* \in \mathbb{R}^{9 \times 9}$. $\mathcal{U}_i \in \mathbb{R}^{2 \times 9}$ and $\mathcal{V}_j \in \mathbb{R}^{2 \times 9}$. We can easily minimize the function $[\mathcal{U}_i \mathcal{V}_j]^T H_* = 0$ via SVD. Further, we use normalization and coordinate axis rotation to improve the stability and accuracy of the model [4] and [17].

4. Global Line-guided Mesh Deformation

The global homography estimated by pre-alignment only provides an approximate transformation, but there are still distortions and bent lines. The longer the salient lines, the greater the straight line bending. In order to address this issue, we explore global collinear structures for line preserving constraints and combine it with point-line alignment, and distortion terms in an energy function.

4.1. Energy Function Definition

First, we construct rectangular meshes for each image pair. Let I and I' denote the target and reference images, respectively. Suppose the index for the mesh grid vertices is from 1 to n . Vector $V = [x_1 \ y_1 \ x_2 \ y_2 \ \dots \ x_n \ y_n]^T$ ($V \in \mathbb{R}^{2n}$) is used to describe the coordinates of original vertex, and vector $\hat{V} = [\hat{x}_1 \ \hat{y}_1 \ \hat{x}_2 \ \hat{y}_2 \ \dots \ \hat{x}_n \ \hat{y}_n]^T$ ($\hat{V} \in \mathbb{R}^{2n}$) represents the coordinates of warped vertex. For any sample point $p \in I$, we represent it as $\tau(p)$ by a bilinear interpolation of its four enclosing grid vertices. Then, the corresponding warped point \hat{p} is represented as $\tau(\hat{p})$. The total

Algorithm 1 Extraction of local and global lines.

Input: original set of lines $S_l = l_1, l_2, \dots, l_n$.
Output: the local lines set S_{lo} and global lines set S_{gl}

```

1: procedure "LINE-MERGING AND CLASSIFICATION"
2:   for  $i = 1 \rightarrow n$  do
3:     for  $j = 1 \rightarrow n - 1$  do
4:       if  $flag(l_m^{lo}) == 0$  and  $flag(l_n^{lo}) == 0$ 
5:         and  $i \neq j$  then
6:           if  $\theta = \arctan \left| \frac{k(l_1) - k(l_2)}{1 + k(l_1)k(l_2)} \right| < \gamma_1$  and
7:              $dis(l_j, p_{l_i}^e) - dis(l_j, p_{l_i}^s) < \gamma_2$  and
8:              $dis(p_{l_i}^e, p_{l_j}^s) < \gamma_3 \cdot dis(p_{l_i}^s, p_{l_j}^e)$  then
9:               merge  $l_i$  and  $l_j$  as a new line  $\hat{l}_{ij}$  and
10:               $S_l \leftarrow S_l \cup \{\hat{l}_{ij}\}$ ,  $n = n + 1$  and
11:               $flag(l_m^{lo}) = 1$ ,  $flag(l_n^{lo}) = 1$ 
12:            end if
13:          end if
14:        end for
15:      end for
16:      for each  $l \in S_l$  do
17:        if  $len(l) > \mu$  then
18:           $S_{gl} \leftarrow l$ 
19:        else
20:           $S_{lo} \leftarrow l$ 
21:        end if
22:      end for
23:      return  $S_{gl}$  and  $S_{lo}$ 
24: end procedure

```

energy function $E(\hat{V})$ is

$$E(\hat{V}) = E_{lp}(\hat{V}) + E_a(\hat{V}) + E_d(\hat{V}), \quad (3)$$

where $E_{lp}(\hat{V})$ addresses line preserving term by protecting both local and global lines, $E_a(\hat{V})$ addresses point-line alignment term by improving the correspondences between matching points and lines, and $E_d(\hat{V})$ addresses distortion control term by preserving the slope of grid lines and warping the adjacent grids evenly.

4.2. Line Preserving Term

Both local individual salient lines and global collinear line segments are crucial for warps. Thus, line preserving term is defined as

$$E_{lp}(\hat{V}) = \lambda_{lo} E_{lo}(\hat{V}) + \lambda_{gl} E_{gl}(\hat{V}), \quad (4)$$

where $E_{lo}(\hat{V})$ and $E_{gl}(\hat{V})$ are constraints on local lines and global co-linear line segments, respectively. λ_{lo} and λ_{gl} are the weights of each term.

As shown in Fig. 4, the local red lines $l_1 \dots, l_5$ are separated in space, but they are co-linear as illustrated by the global blue line l_6 . As demonstrated in the second image in Fig. 4, the local line structures can be kept but their collinearity are easily destroyed during warping. Since such distortions are very disturbing for human perception of image quality, we design a merging strategy of local warp-

ings to preserve global linear structure by evaluating the collinearity of line segments, which is one of our main contributions. The merging process is detailed in Algorithm 1. We evaluate pairs of lines each time, and the merged lines should meet three constraints. First, the slope of two lines $slope(l_i)$ and $slope(l_j)$ should be close. Second, the distances from the endpoints to another line, which are $dis(l_j, p_{l_i}^e)$ and $dis(l_j, p_{l_i}^s)$ should be small. Third, the distance of adjacent endpoints $dis(p_{l_i}^e, p_{l_j}^s)$ of two lines should be small, shown in Fig. 4. Note that we introduce $flag$ to avoid infinite loop of merging lines, which is set to 0 initially and set to 1 after merging. While local lines are most likely obtained by the original line fitting, the global lines usually result from the merging process. It is worth noticing that the length of each line $len(l)$ is used to classify local and global line segments by threshold μ .

Let us take lines $\{l_g\}_{g=1,2,\dots,Q}$ in global line set S_{gl} as an example, where Q is the number of lines. Each line is uniformly sampled with M_g points $\{p_k^g\}_{k=1,2,\dots,M_g}$. Then,

$$E_{gl}(\hat{V}) = \sum_{g=1}^Q \sum_{k=1}^{M_g-1} \|(\tau(\widehat{p_{k+1}^g}) - \tau(\widehat{p_k^g})) \cdot \vec{n}^g\|^2 \\ = \|W_{gl}\hat{V}\|^2, \quad (5)$$

where \vec{n}^g is the normal vector of l_g , and $W_{gl} \in \mathbb{R}^{(\sum_{g=1}^Q (M_g-1)) \times 2n}$. We use the same method to construct the constraint term for local lines.



Figure 4: A diagram demonstrating line merging.

4.3. Point-Line Alignment and Distortion Control

The point-line alignment term is defined based on an intuitive constraint that matched point and line pairs should be coincident with each other after warps, which is defined as

$$E_a(\hat{V}) = \lambda_p E_p(\hat{V}) + \lambda_l E_l(\hat{V}), \quad (6)$$

where $E_p(\hat{V})$ and $E_l(\hat{V})$ are point and line alignment term, respectively. λ_p and λ_l are the weights of each term.

In order to control the distortion of the target image I , a series of horizontal and vertical lines are constructed, which

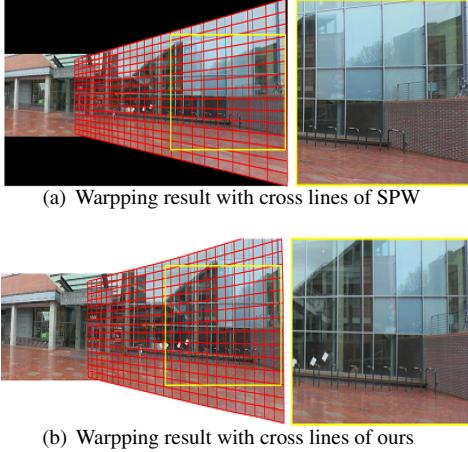


Figure 5: Comparison of warping result between SPW and our method showing by the same number of cross lines.

are called cross lines. These constructed lines are regarded as inherent linear structure of image I , as demonstrated as red lines in Fig. 5. The slope of the lines and space between their intersections are used to control the distortion. The distortion term is defined by global term E_{dg} and non-overlapping term E_{dn}

$$E_d(\hat{V}) = \lambda_{dg} E_{dg}(\hat{V}) + \lambda_{dn} E_{dn}(\hat{V}), \quad (7)$$

where λ_{dg} and λ_{dn} denote the weights of each term.

As all constraint terms are quadratic, they can be reformulated and minimized by a sparse linear solver. More details can be found in [17]. The warping result is compared with SPW [17], which also has line constraints. Both methods have the same number of evenly spaced cross lines for target image I in original. The warping result is demonstrated in Fig. 5, and our method exhibits dense cross lines and moderate transition from overlapping area to non-overlapping area, showing good control to distortions, while evident distortions appear in the results of SPW [17] shown in the zoomed-in rectangle.

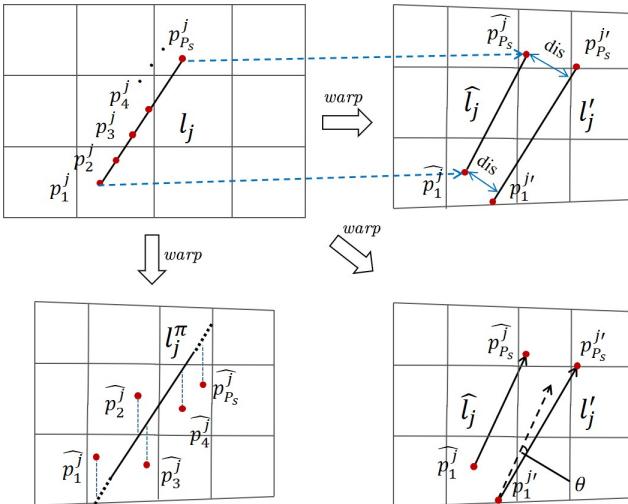


Figure 6: Quantitative evaluation for linear structures.

5. Quantitative Evaluation for Collinearity

In order to quantify the stitching performance on linear structures, we design a new evaluation method that considers three aspects: the collinearity of points, the distance of matched lines, and the discrepancy in the direction of matched lines.

We sample P_s points $\{p_k^j\}_{k=1,2,\dots,P_s}$ on line l_j uniformly, which are labeled in red in Fig. 6. The bottom-left figure in Fig. 6 demonstrates the fitted line $l_j^\pi = \pi(\{p_k^j\}_{k=1,2,\dots,P_s})$ to warped points $\{\widehat{p}_k^j\}_{k=1,2,\dots,P_s}$ by least square method. The error term E_{err} for L lines is defined as

$$E_{err} = \sqrt{\frac{1}{L} \sum_{j=1}^L \left\| \sum_{k=1}^{P_s} (l_j^\pi(y)_{x=\widehat{p}_k^j(x)} - \widehat{p}_k^j(y)) \right\|^2}, \quad (8)$$

where $\widehat{p}_k^j(y)$ is the y-coordinate of \widehat{p}_k^j , and $l_j^\pi(y)_{x=\widehat{p}_k^j(x)}$ is the y-coordinate on l_j^π with the same x-coordinate as \widehat{p}_k^j .

As demonstrated in the upper-right figure of Fig. 6, the distance term E_{dis} represents the average distance between two warped endpoints and the matching line, which is defined as

$$E_{dis} = \sqrt{\frac{1}{L} \sum_{j=1}^L \left\| \frac{dis(l'_j, \widehat{p}_1^j) + dis(l'_j, \widehat{p}_{P_s}^j)}{2} \right\|^2}, \quad (9)$$

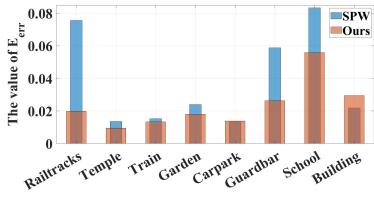
where \widehat{p}_1^j and $\widehat{p}_{P_s}^j$ are two endpoints of \widehat{l}_j .

The direction term E_{dir} estimates the direction difference between the warped line and the matching line. We represent line \widehat{l}_j and line l'_j as direction vectors \overrightarrow{l}_j and \overrightarrow{l}'_j , respectively. The endpoint with smaller x-coordinate is used as the start point of each vector. As demonstrated in the bottom-right figure of Fig. 6, the cross product of two lines $\overrightarrow{l}'_j \times \overrightarrow{l}_j = len(\overrightarrow{l}'_j) \times len(\overrightarrow{l}_j) \times \sin(\theta)$ can reflect the direction difference of two vectors, which also takes the length of two lines $len(\overrightarrow{l}'_j)$ and $len(\overrightarrow{l}_j)$ into consideration. Hence, E_{dir} for all lines is denoted as

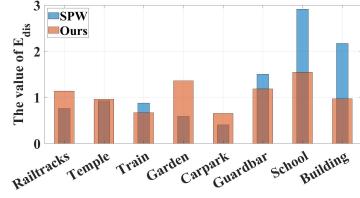
$$E_{dir} = \sqrt{\frac{1}{L} \sum_{j=1}^L \left\| \overrightarrow{l}'_j \times \overrightarrow{l}_j \right\|^2}. \quad (10)$$

6. Experiments

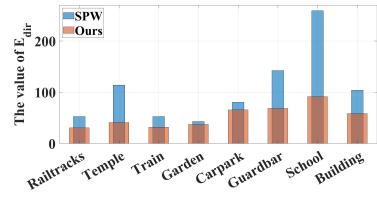
We demonstrate the effectiveness of the proposed method with an ablation study and with quantitative and qualitative comparisons to the state-of-the-art on 15 testing images pairs, which covered different types of datasets with respect to camera motions, scenes and fields of view. Input images pairs are resized to 1000×800 pixels, and the size for each mesh grid is 40×40 . This allows us to keep



(a) Evaluation on collinearity of warped lines.

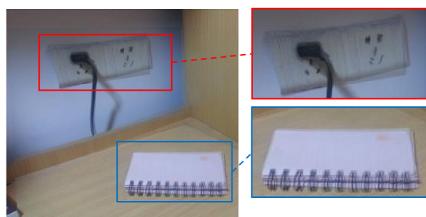


(b) Distance of matched line pairs.

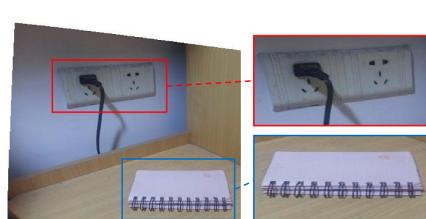


(c) Angle difference of matched line pairs.

Figure 7: Quantitative comparison on linear structure preservation between SPW [17] and our method.



(a) Matching lines and points separately.



(b) Joint matching with the proposed consistent point and line pairs.
Figure 8: Ablation study on consistent lines and points matching.

all the parameters constant. The threshold μ to divide local and global line segments is set to three times the diagonal length of the grid. In energy function, λ_{lo} and λ_{gl} are set to 50 and 100 for lines preserving, λ_p and λ_t are set to 1 and 5 for point-line alignment, λ_{dg} and λ_{dn} are set to 50 and 100 for distortion minimization.

Dataset	SVA [21]	CPW [22]	APAP [30]	GSP [5]	SPW [17]	Ours
APAP-railtracks	7.30	6.77	4.51	4.58	2.28	1.78
DH-temple	12.21	2.54	2.04	2.21	2.31	1.78
APAP-conssite	11.36	7.06	5.16	5.01	4.4	2.02
APAP-train	9.16	6.33	5.24	4.47	2.11	1.89
APAP-garden	8.98	6.36	5.19	4.15	2.85	1.44
DH-carpark	4.05	3.60	1.38	1.78	1.54	1.31
SVA-chessgirl	20.78	9.45	2.96	2.88	1.88	1.69

Table 1: RMSE on matched feature points.

6.1. Ablation Study

Consistent lines and points constraints can provide accurate alignment and suppress artifacts. We substitute our joint matching strategy with a separate points and lines matching strategy [17] with other parts unchanged. Sample results are shown in Fig. 8. There are three stitching instances, the enlarged overlapping areas are shown on the right of each result. As we can see, our joint points and lines matching produces clear stitching result in Fig. 8(b), while there are obvious artifacts in Fig. 8(a).

Linear structure preserving can keep both local and global linear structures and suppress the distortion as shown in Fig. 9. In Fig. 9(a), local shapes are well preserved but the lines are bent without linear constraints. When we roughly use lines detected by LSD, the short lines on buildings and the ground are much better, but they are still not straight enough. Especially, the bike has severe deformation in Fig. 9(b). By contrast, the local and global linear constraints are well balanced in our method, as shown in Fig. 9(c).

6.2. Comparison with the state-of-the-arts

Couple of experiments are conducted to compare the stitching results with the existing methods including



(a) Without linear constraints.

(b) With only local linear constraints.

(c) With our local and global linear constraints.

Figure 9: Ablation study on structure preserving.

SVA [21], CPW [22], APAP [30], ELA [14], SPHP [4], GSP [5], and SPW [17]. The quantitative and qualitative results are from their papers or the released code.

We adopt quantitative evaluation on both points and lines. The alignment accuracy on points is measured by the root mean squared error (RMSE) [30] on a set of matched points. Table 1 depicts the RMSE values on 7 image pairs named in the first column. Our method yields the lowest errors on all 7 pairs. Our average error is 1.7014, which is 31% lower than that of SPW. Moreover, the variance of our errors is 0.05, which is 93% lower than the value 0.75 of SPW.

Further, we evaluate linear structures using our three measures proposed in Sec. 5. In Fig. 7, we compare our method to SPW [17], which has local but not global line constraints. As shown in Fig. 7, the x-coordinate indicates the name of the image pairs, and vertical coordinate indicates the errors. Our method outperforms SPW on collinearity and difference of angles, as shown in Fig. 7(a) and Fig. 7(c). In Fig. 7(b), the variance of our error is 0.0855, which is 87.11% less than 0.6630 of SPW, showing less divergence errors for all pairs. SPW shifts severely with about twice errors of us on the image pairs 'School' and 'Building'.

A comprehensive visual comparison is demonstrated in Fig. 10. Our method outperforms all the other methods in preserving linear structure and in producing clear and artifacts-free overlapping areas. The linear structures in the first four rows all exhibit severe bent, labelled in red. Evident artifacts appear in the results of the existing methods shown in the zoomed-in rectangles, such as the flowers, tables, and parasols.

7. Conclusion

We propose a structure preserving image stitching method based on line-guided warping and line-point constraint. We partition input images into sub-regions, and match them by line-point invariants. The local matching provides accurate line and point pairs for pre-alignment, showing no blur or artifacts in overlapping areas. We pro-

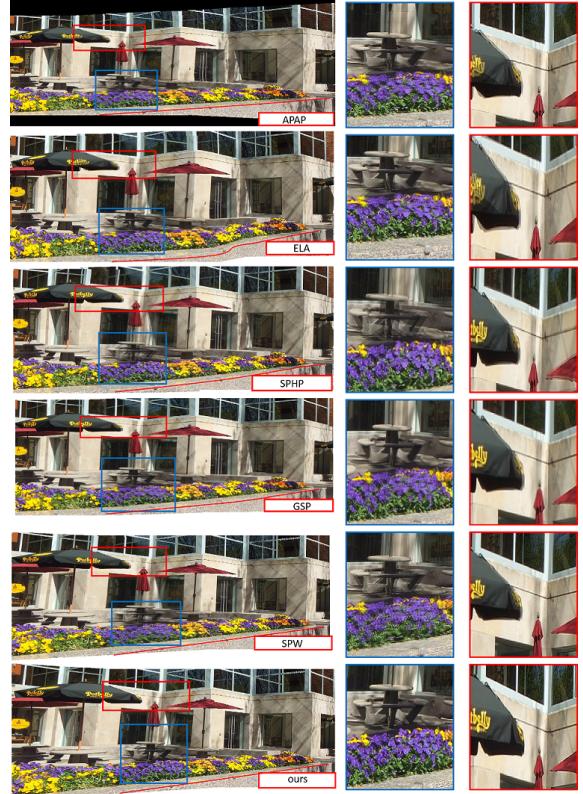


Figure 10: Comparison of different stitching methods. pose a line-guided warp to preserve both local and global structures while eliminating distortion for non-overlapping area. Furthermore, we design a new quantitative evaluation measure for linear structures, which is consistent with human perception in that human vision is very sensitive to distortions in linear structures. Experimental results demonstrated that the proposed method accurately aligns overlapping and non-overlapping areas on challenging test images, and yields a significantly better performance compared with the state-of-the-arts.

Acknowledgements. This work is partially supported by the NSF of China under grant Nos. 61876030, 61733002, and 61572105, Fundamental Research Funds for the Central University under grant No. DUT19RC(3)004, and by the NSF under Grant No. IIS-1814745.

References

- [1] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008. 4
- [2] Matthew Brown and David G Lowe. Automatic panoramic image stitching using invariant features. *International journal of computer vision*, 74(1):59–73, 2007. 2
- [3] Che-Han Chang and Yung-Yu Chuang. A line-structure-preserving approach to image resizing. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1075–1082. IEEE, 2012. 2
- [4] Che-Han 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, 2, 4, 8
- [5] Yu-Sheng Chen and Yung-Yu Chuang. Natural image stitching with the global similarity prior. In *European conference on computer vision*, pages 186–201. Springer, 2016. 1, 2, 3, 7, 8
- [6] Abhishek Kumar Dewangan, Rohit Raja, and Reetika Singh. An implementation of multi sensor based mobile robot with image stitching application. *Int J Comput Sci Mobile Comput*, 3(6):603–609, 2014. 1
- [7] Martin A Fischler and Robert C Bolles. A paradigm for model fitting with applications to image analysis and automated cartography (reprinted in readings in computer vision, ed. ma fischler,”). *Comm. ACM*, 24(6):381–395, 1981. 4
- [8] Junhong Gao, Seon Joo Kim, and Michael S Brown. Constructing image panoramas using dual-homography warping. In *CVPR 2011*, pages 49–56. IEEE, 2011. 2
- [9] Charles Herrmann, Chen Wang, Richard Strong Bowen, Emil Keyder, Michael Krainin, Ce Liu, and Ramin Zabih. Robust image stitching with multiple registrations. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 53–67, 2018. 3
- [10] Qi Jia, Xinkai Gao, Xin Fan, Zhongxuan Luo, Haojie Li, and Ziyao Chen. Novel coplanar line-points invariants for robust line matching across views. In *European Conference on Computer Vision*, pages 599–611. Springer, 2016. 3, 4
- [11] Kyungdon Joo, Namil Kim, Tae-Hyun Oh, and In So Kweon. Line meets as-projective-as-possible image stitching with moving dlt. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 1175–1179. IEEE, 2015. 1
- [12] Kyu-Yul Lee and Jae-Young Sim. Warping residual based image stitching for large parallax. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8198–8206, 2020. 2, 3
- [13] Dongping Li, Kaiming He, Jian Sun, and Kun Zhou. A geodesic-preserving method for image warping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 213–221, 2015. 2
- [14] 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. 2, 8
- [15] Nan Li, Yifang Xu, and Chao Wang. Quasi-homography warps in image stitching. *IEEE Transactions on Multimedia*, 20(6):1365–1375, 2017. 3
- [16] 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, 3
- [17] 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, 8
- [18] 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. 2
- [19] 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. 3
- [20] Kaimo Lin, Nianjuan Jiang, Shuaicheng Liu, Loong-Fah Cheong, Minh Do, and Jiangbo Lu. Direct photometric alignment by mesh deformation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2405–2413, 2017. 3
- [21] Wen-Yan Lin, Siying Liu, Yasuyuki Matsushita, Tian-Tsong Ng, and Loong-Fah Cheong. Smoothly varying affine stitching. In *CVPR 2011*, pages 345–352. IEEE, 2011. 2, 7, 8
- [22] Feng Liu, Michael Gleicher, Hailin Jin, and Asseem Agarwala. Content-preserving warps for 3d video stabilization. *ACM Transactions on Graphics (TOG)*, 28(3):1–9, 2009. 3, 7, 8
- [23] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. 1
- [24] Xiaoyuan Luo, Yang Li, Jing Yan, and Xinping Guan. Image stitching with positional relationship constraints of feature points and lines. *Pattern Recognition Letters*, pages 431–440, 2020. 1
- [25] Richard Szeliski. Image alignment and stitching. In *Handbook of mathematical models in computer vision*, pages 273–292. Springer, 2006. 1
- [26] 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. 2, 3, 4
- [27] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 2
- [28] Tian-Zhu Xiang, Gui-Song Xia, Xiang Bai, and Liangpei Zhang. Image stitching by line-guided local warping with global similarity constraint. *Pattern Recognition*, 83:481–497, 2018. 3
- [29] Yingen Xiong and Kari Pulli. Sequential image stitching for mobile panoramas. In *2009 7th International Conference on Information, Communications and Signal Processing (ICICS)*, pages 1–5. IEEE, 2009. 1

- [30] 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](#), [7](#), [8](#)
- [31] 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](#)
- [32] Guofeng Zhang, Yi He, Weifeng Chen, Jiaya Jia, and Jun Bao. Multi-viewpoint panorama construction with wide-baseline images. *IEEE Transactions on Image Processing*, 25(7):3099–3111, 2016. [3](#)