

Feature Correspondences Increase and Hybrid Terms Optimization Warp for Image Stitching

Yizhi Cong ¹, Yan Wang ^{1,2,*}, Wenju Hou ² and Wei Pang ³

¹ Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, School of Artificial Intelligence, Jilin University, Changchun 130012, China

² College of Computer Science and Technology, Jilin University, Changchun 130012, China

³ School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK

* Correspondence: wy6868@jlu.edu.cn

Abstract: Feature detection and correct matching are the basis of the image stitching process. Whether the matching is correct and the number of matches directly affect the quality of the final stitching results. At present, almost all image stitching methods use SIFT+RANSAC pattern to extract and match feature points. However, it is difficult to obtain sufficient correct matching points in low-textured or repetitively-textured regions, resulting in insufficient matching points in the overlapping region, and this further leads to the warping model being estimated erroneously. In this paper, we propose a novel and flexible approach by increasing feature correspondences and optimizing hybrid terms. It can obtain sufficient correct feature correspondences in the overlapping region with low-textured or repetitively-textured areas to eliminate misalignment. When a weak texture and large parallax coexist in the overlapping region, the alignment and distortion often restrict each other and are difficult to balance. Accurate alignment is often accompanied by projection distortion and perspective distortion. Regarding this, we propose hybrid terms optimization warp, which combines global similarity transformations on the basis of initial global homography and estimates the optimal warping by adjusting various term parameters. By doing this, we can mitigate projection distortion and perspective distortion, while effectively balancing alignment and distortion. The experimental results demonstrate that the proposed method outperforms the state-of-the-art in accurate alignment on images with low-textured areas in the overlapping region, and the stitching results have less perspective and projection distortion.

Keywords: image stitching; image alignment; feature correspondences increase; hybrid terms optimization



Citation: Cong, Y.; Wang, Y.; Hou, W.; Pang, W. Feature Correspondences Increase and Hybrid Terms Optimization Warp for Image Stitching. *Entropy* **2023**, *25*, 106. <https://doi.org/10.3390/e25010106>

Academic Editors: KC Santosh, Alejandro Rodríguez-González, Ayush Goyal, Satish Kumar Singh, Djamila Aouada, Aaisha Makkar and Yao-Yi Chiang

Received: 12 November 2022

Revised: 19 December 2022

Accepted: 19 December 2022

Published: 4 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Image stitching is the process of merging a series of images with overlapping regions into a high-resolution image with a wider field of view [1]. It has been widely used in applications such as panorama [2], autonomous driving [3], video stabilization [4], virtual reality [5], and many others. Up till now, there are still great challenges for image stitching to align accurately, eliminate artifacts, and mitigate distortion, especially on low-textured and large-parallax images.

Traditional image stitching methods often use feature points to obtain a global homography matrix, and they map images with overlapping regions to the same plane. However, these approaches only work if the images are in the same plane or only rotated around the camera center. Under the condition of both parallax and rotation, global homography often produces misalignment and/or artifacts in the overlapping region and distortion in the non-overlapping region. Recently, spatially varying warping [6–8] has been widely used in image stitching. It divides an image into several grid cells, and then aligns each grid cell with local homography. It can effectively eliminate misalignment, but it may also produce

distortions in the non-overlapping region. Since homography-based methods inevitably generate distortions, researchers have taken a different approach by using a seam-driven strategy [9–11], which finds the optimal seam according to a customized energy function. Although such an approach can effectively mitigate distortions, it usually cannot achieve the desired alignments. In addition, such a method relies too much on the energy function, and it is more frequently used as a post-processing blending method to eliminate artifacts. In recent years, deep learning-based image stitching frameworks [12–14] have been widely studied by researchers. However, they may still not be able to achieve the desired accurate alignment and distortion preservation performance compared to traditional methods.

Feature detection and correct matching are the basis of the image stitching process. SIFT [15] is a feature point detection and matching method used by almost all image stitching methods. After preliminary matching using SIFT, RANSAC [16] is used to remove outliers and obtain inliers. Although SIFT can extract a large number of feature points, the correct matching points after RANSAC iteration are greatly reduced compared with the detected feature points, which may lead to insufficient feature correspondence in the low-textured or repetitively-textured regions. Figure 1 shows the feature correspondences of SIFT+RANSAC under three different image types. It can be seen that the traditional SIFT+RANSAC pattern has sufficient inliers only under the condition of dense texture in the overlapping region, while obviously insufficient inliers in low-textured areas (see Figure 1a,c).

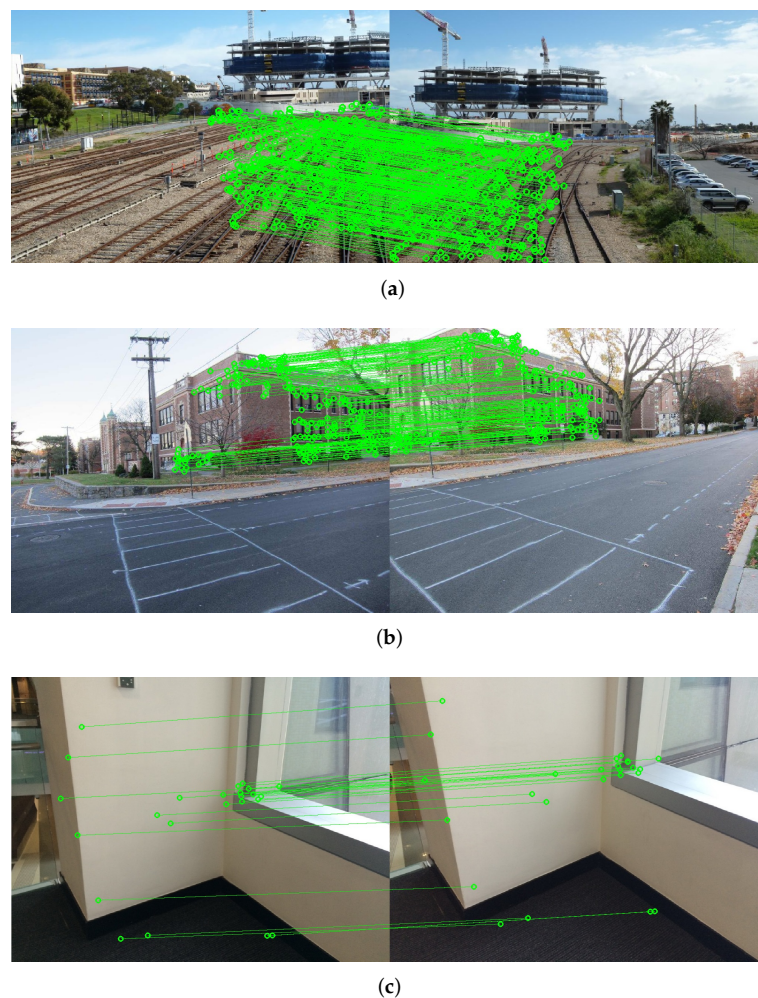


Figure 1. Comparing the number and distribution of inliers on different texture images. (a) Dense texture. (b) Semi-dense texture. (c) Low texture. The images are from [8,17,18], respectively.

Grid-based image stitching methods are heavily dependent on the quantity and quality of feature points, which are crucial to the quality of image stitching results. The as-projective-as-possible (APAP) warp model [8] is a typical mesh-based local warping model. It uses SIFT+RANSAC as pre-processing. Figure 2a shows the inliers of APAP and the stitching result. The input images are with low-textured or repetitively-textured regions, which lead to insufficient inliers after RANSAC. It further causes misalignment of the stitching result. RANSAC can increase the number of inliers by increasing the threshold. Figure 2b shows the inliers and the stitching result after increasing the RANSAC threshold appropriately. Obviously, increasing the number of feature correspondences produces a more accurate alignment. However, limited by the local alignment of the APAP method, some unacceptable jagged misalignments appear unnaturally. Furthermore, limited by the characteristics of the RANSAC algorithm itself, when the RANSAC threshold is too high, it may still generate false matches if the overlapping region has many similar textures.

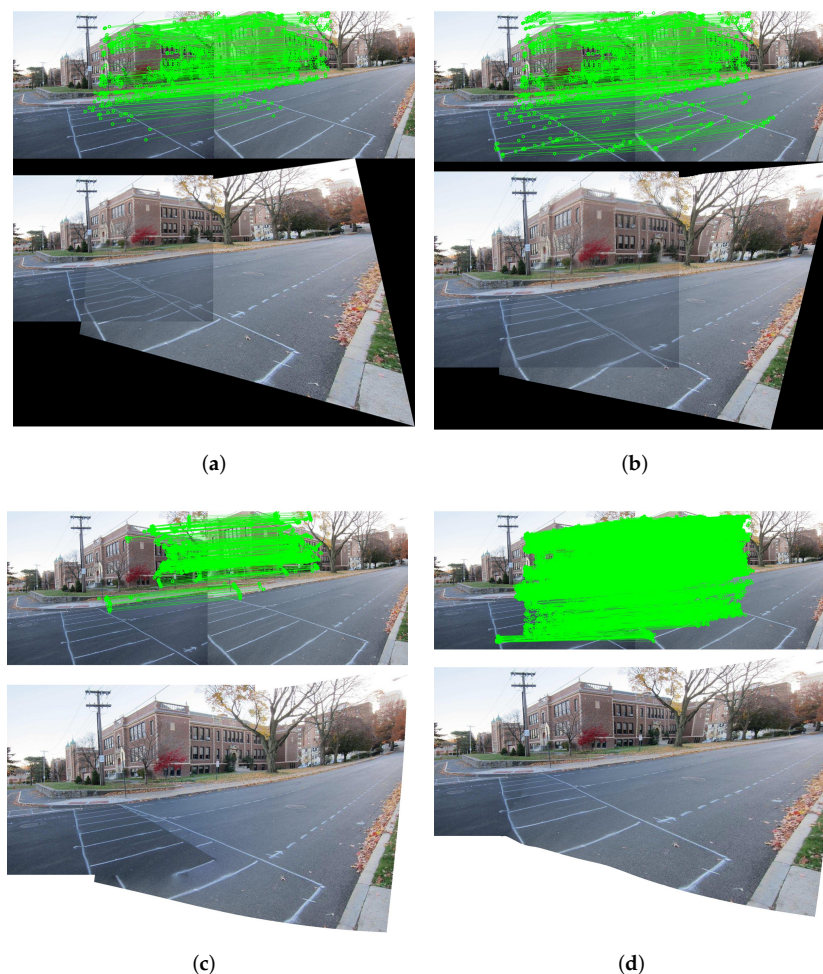


Figure 2. Comparison of inliers and stitching results of different methods. For each subfigure, the feature correspondences are on the top, and the stitching results are on the bottom. (a) APAP with RANSAC threshold 0.1. (b) APAP with RANSAC threshold 0.5. (c) LPC. (d) Ours.

Considering that there may still be the issue of matching errors after RANSAC, elastic local alignment (ELA) [19] introduces Bayesian feature refinement to further remove possible outliers. Although outliers hardly exist after using ELA, ELA also removes some of the original normal values, and this may lead to the issue of insufficient feature points. Leveraging line-point consistence (LPC) [20] establishes a relationship between point and line dual features, and it applies a characteristic number (CN) [21] to add salient points to the feature point set to enrich the number of feature points. However, the feature points added

by CN may not be correct matches, and they need to go through RANSAC, which will also make a low-textured region with insufficient feature points (see Figure 2c). Although some methods [18,22,23] that introduce line features and use LSD [24] to extract line segment features can cope well with a low-textured environment, line segments are mainly matched by endpoints, which are essentially point features, and they are subject to line segment length thresholds. Currently, line segment matching is still a challenging problem.

Grid-based Motion Statistics (GMS) [25] assume that adjacent pixels share similar motions, which is consistent with the principle of image stitching. It uses ORB [26] to extract feature points. Since ORB can set the initial feature point threshold, combined with GMS, it can increase the number of correct matching points in low-textured regions, thereby ensuring that the warped model has sufficient anchor points for local alignment. Compared with the SIFT+RANSAC pattern, the ORB+GMS method can not only make feature correspondences sufficient, but also ensure the correctness of the feature matching, which helps improve the quality of the stitching. The results of the proposed method are shown in Figure 2d.

In this paper, we propose a novel image stitching method based on feature correspondences increase and hybrid terms optimization warp. We first use the ORB+GMS algorithm to extract and match feature points to obtain sufficient feature correspondences, and at the same time we use line features to assist alignment and structure preservation; then, an initial homography combining global homography transformation and global similarity transformation is estimated as the initial homography to be optimized. On this basis, we estimate various optimization terms for the final image warping. Finally, we obtain the final stitching result by linear blending or seam-cutting blending. The overall process architecture of our method is shown in Figure 3.

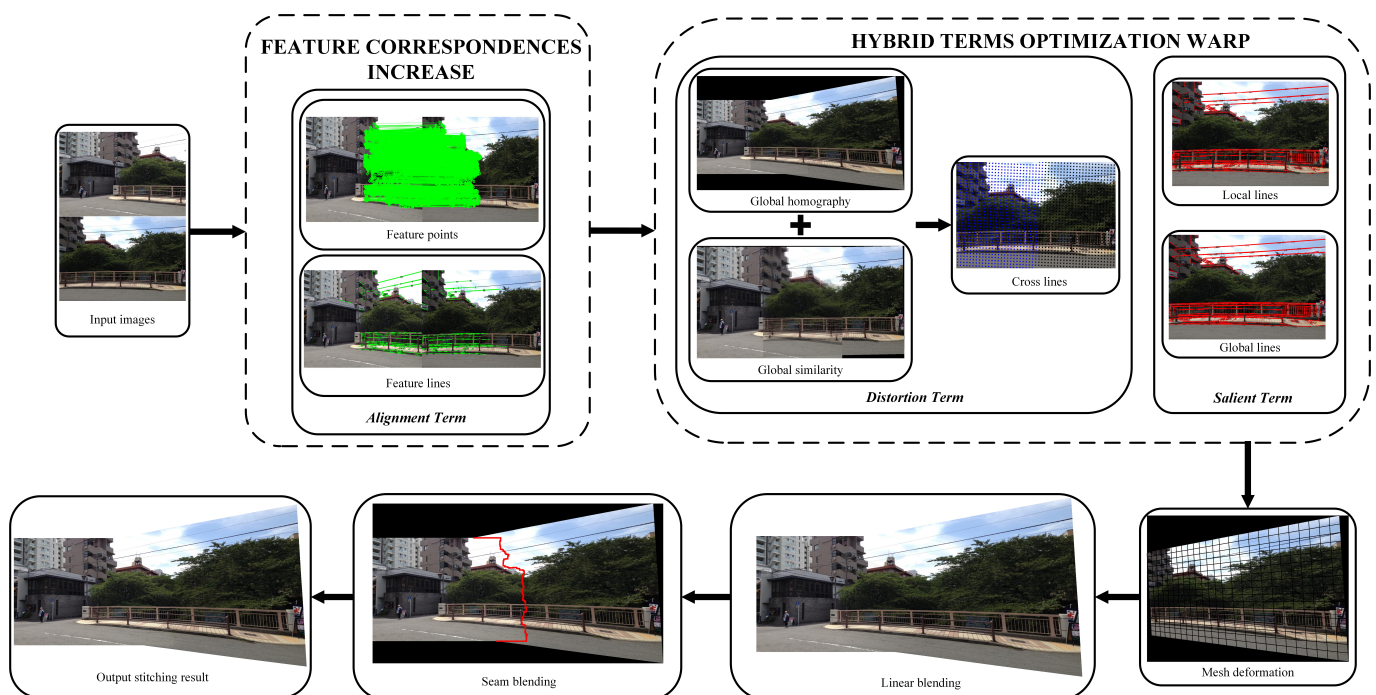


Figure 3. The overall process architecture of our method. The feature correspondences increase module mainly increases the number of feature points to ensure accurate alignment. The hybrid terms optimization warp is based on the above stage to obtain the optimal mesh deformation, which further balances the alignment and distortion.

The contributions of our approach are as follows:

- A novel method to increase feature correspondences is proposed, which can be added to any image stitching model. It solves the problem of insufficient feature correspon-

dence in low-textured or repetitively-textured regions, and it can effectively eliminate misalignment and artifacts.

- A novel hybrid transformation combining global homography transformation and global similarity transformation is proposed to serve as the initial homography for structure preservation, which can flexibly fine-tune the structure of image stitching results compared to conventional global homography.
- Various optimization terms are used to locally adjust the above hybrid warping model, which can effectively mitigate projection and perspective distortion. Our flexible and robust method can effectively balance alignment and distortion, especially on images with low-textured areas in the overlapping region. For images with large parallax or significant foreground-background relationship, seam-cutting blending instead of linear blending is used to eliminate inevitable misalignment or artifacts.

2. Related Work

2.1. Feature Extraction and Matching

Feature points refer to the points where the gray values of the image change drastically or the points with large curvature on the edges of the image, such as corners and spots. In the early days, common corner detection algorithms were Harris corner [27] and FAST [28]. Due to the feature limitations of corner points, Lowe et al. proposed SIFT [15], which is invariant to scale, rotation, and brightness. It is robust and fast, and it has been the most popular feature detector until now. SURF [29] improves the speed compared with SIFT, which is more suitable for practical applications. BRIEF [30] uses the Hamming Distance to achieve matching using the XOR operation between bits. ORB [26] combines the advantages of FAST and BRIEF. Compared with SIFT/SURF, it can detect denser matching points in areas with flat textures, and it has the best comprehensive performance among the current traditional feature detectors.

After feature detection, the Brute Force (BF) [31] algorithm is often used for preliminary matching, but there are often some false matches. In recent years, RANSAC [16] has been the most popular method for filtering out outliers. However, it is affected by the threshold of inliers; especially for locally aligned warping models, inliers are often mistakenly eliminated by RANSAC, resulting in an insufficient number of points in flat textured areas, and the rest cannot be guaranteed to be inliers, resulting in wrong stitching results. Some researchers have taken corresponding countermeasures to the problem of RANSAC. The triangular facet transformation (TFT) [32] uses bundle adjustment (BA) [33] to optimize for false matches that may be due to noise. LPC [21] uses the CN [20] point-line relationship to expand the matching point set, but the two-step RANSAC may lead to the loss of feature correspondences in the weakly-textured region, thus affecting the stitching results. Bian et al. proposed GMS [25], which assumes that adjacent pixels have similar motion and distinguishes between correct and incorrect matches by computing statistical scores within a small area around the matching point. It can be combined with ORB to increase the number of features by increasing the matching point threshold to help the matching quality, which makes it possible to solve challenging image stitching problems. Therefore, compared to RANSAC, GMS is more robust to address feature matching errors and insufficient features in low-texture regions.

2.2. Spatially Varying Warping and Seam Cutting

Traditional image stitching uses a global homography [2] to warp the image in the ideal case where the scene depth has little change in overlapping regions. In recent years, researchers have used multi-homography spatially varying warping to deal with parallax. The dual-homography warping (DHW) [34] divides the scene into a background plane and a foreground plane, and it aligns the background and foreground with two homography matrices, respectively. Smoothly varying affine (SVA) [7] improves local transformation and alignment with multiple affine transformations. APAP [8] divides the image into uniform grid cells, and it performs local homography alignment for each grid, which can better

cope with parallax scenes. However, grid-based spatially varying warping often produces severe distortion in non-overlapping regions. The adaptive as-natural-as-possible (AANAP) warp [17] combines local homography transformation with global similarity transformation to effectively reduce distortion in non-overlapping regions. The shape-preserving half projective (SPHP) warp [35] uses subregional warping to smoothly transition overlapping and non-overlapping regions to reduce distortion.

Due to the difference in parallax and image capture time, artifacts inevitably appear in the overlapping areas of the stitching results. Gao et al. proposed a seam-driven [9] approach to deal with parallax and eliminate artifacts. Parallax-tolerant Image Stitching [10] is inspired by content-preserving warps (CPW) [36] and optimizes the use of seams to improve the stitching performance of large parallax scenes. Seam-guided local alignment (SEAGULL) [11] adds curved and straight structural protection constraints. Perception-based seam cutting for image stitching [37] introduces human perception and saliency detection to obtain the optimal seam, making the stitching result look more natural. Xue et al. proposed a stable seam measurement [38], which is used to estimate a stable seam cutting for mitigating parallax via a hybrid actor–critic architecture.

2.3. Structure Preservation and Distortion Mitigation

Lines are common features in natural images. Line structures bend unnaturally after image warping. The dual-feature warp (DFW) [18] introduces the line segment feature, and the constraint term constrains both points and lines to achieve better alignment. The global similarity prior (GSP) [39] constrains globally similar angles with line alignment, while local and global constraints are combined to reduce structural distortion. The quasi-homography warp (QHW) [40] uses intersecting lines to preserve the nature of homography warping. The single-perspective warp (SPW) [22] utilizes the QHW characteristic; it introduces a variety of protection terms and uses the least square method to obtain the optimal solution of the objective function to obtain the optimal homography. LPC [21] introduces the concept of global lines on the basis of SPW, and it merges short line segments into long ones to reduce the bending of line structures in natural images. Similarity transformation has been proved in [34] to reduce distortion in the overlapping region, and the reason being that similarity transformation are more rectangular and look more natural than homography. Analogously, the work in [41] introduces line and regular boundary preservation constraints to make the boundary of the stitching result as close as possible to a rectangle, while reducing unwanted distortions.

The remainder of this paper is organized as follows. Section 3 presents the proposed image stitching algorithm. The detailed experimental procedure and results are presented in Section 4. Finally, the conclusions are summarized in Section 5.

3. The Proposed Method

The homography transformation is the most common method to achieve image stitching. For the single image warp, global homography warping can handle the cases in the ideal conditions. However, under low texture and large parallax conditions, it is still a big challenge to align accurately even with local homography. It is very dependent on the quality and quantity of features correspondences in the pre-processing stage. The traditional SIFT+RANSAC pattern has been unable to satisfy the above requirements. Meanwhile, the homography transformation inevitably introduces distortion in the non-overlapping region. How to balance alignment and distortion has gradually become the key to single image warping. In this section, we first describe our feature correspondences increase module, and then we propose our hybrid terms optimization warp to address the balance of alignment and distortion.

3.1. Feature Correspondences Increase

Inspired by GMS, we merge the feature correspondences increase module into the image stitching process to solve the problem of insufficient feature correspondences in low-

textured areas in the overlapping region. Compared with the one-step RANSAC, we try to get more correct matches in low-textured or repetitively-textured regions to satisfy the local warping model’s demand for sufficient feature points in overlapping regions to achieve more accurate alignment. Here, we use ORB with high robustness and dense feature points as the detector, BF for preliminary matching, and GMS refinement to distinguish between correct and incorrect matches.

GMS considers that there are several features that match the matching relationship in the correctly matched neighborhoods, while there are almost no features in the incorrectly matched neighborhoods. According to this feature, GMS counts the number of features that match the matching relationship to distinguish between correct and incorrect matches. Let $\{N, M\}$ be the number of feature points of the input image pair $\{I_a, I_b\}$ after ORB+BF, respectively. $X = \{x_1, x_2, \dots, x_n\}$ is all the feature matching neighborhoods from image I_a to I_b . X can be classified as true or false by measuring the local support of each matching pair through GMS. $X_i \subseteq X$ is the subset of all neighborhoods. S_i is the neighborhood support, which is expressed as follows:

$$S_i = |X_i| - 1, \tag{1}$$

where -1 removes the original match from the sum.

GMS assumes that if the motion is smooth in a region, the correct matches have the same spatial locations across multiple neighborhood pairs, while false matches have relative spatial locations across multiple neighborhood pairs different. Equation (1) can be rewritten as follows:

$$S_i = \sum_{k=1}^K |X_{a^k b^k}| - 1, \tag{2}$$

where K is the number of small neighborhoods predicted to move along with feature matching. $\{a^k, b^k\}$ is the predicted region pair, $X_{a^k b^k}$ belonging to X is the feature matching subset.

Let T^{ab} and F^{ab} be the same and different positions of the region $\{a, b\}$, respectively. f_a^b is the nearest neighbor feature in region b to one of the features in region a . $p_t = p(f_a^b | T^{ab})$ and $p_f = p(f_a^b | F^{ab})$. Assuming that each region pair $\{a, b\}$ has $\{n, m\}$ feature points, then:

$$\begin{cases} p_t = t + (1 - t)\beta m / M \\ p_f = \beta(1 - t)(m / M) \end{cases}, \tag{3}$$

where β is the adjustment factor. Thus, S_i be the number of matches in the neighborhood of x_i , the distribution of it follows the binomial distribution in Equation (2):

$$S_i \sim \begin{cases} B(Kn, p_t), & \text{if } x_i \text{ is true} \\ B(Kn, p_f), & \text{if } x_i \text{ is false} \end{cases}. \tag{4}$$

where K is the number of disjoint regions which match i predicts move together.

To calculate S_i more efficiently, the image is divided into 20×20 grids as in [25]. The score S_{ij} for the cell pair $\{i, j\}$ is calculated as follows:

$$S_{ij} = \sum_{k=1}^{K=9} |X_{i^k j^k}|, \tag{5}$$

where $|X_{i^k j^k}|$ is the number of feature matches in the nine-square grid. The score threshold τ is used to distinguish whether the feature matching is correct. If $S_{ij} > \tau$, the matching x_i at grid $\{i, j\}$ is true; otherwise, x_i is false.

In this way, we obtain a denser set of feature correspondences compared with the traditional SIFT+RANSAC pattern. It will be used as input for the subsequent warping.

3.2. Hybrid Terms Optimization Warp

3.2.1. Mathematical Preparation

Let $\mathbf{X}_1 = [x_1, y_1, 1]^T$ and $\mathbf{X}_2 = [x_2, y_2, 1]^T$ denote the feature points obtained by the feature correspondences increase in the input images I_a and I_b , respectively. The alignment of the overlapping region of the two images is a linear transformation from I_a to I_b in homogeneous coordinates, which is defined below:

$$\mathbf{X}_2 = H\mathbf{X}_1, \tag{6}$$

where H is defined as a 3×3 matrix. If it is a global homography transformation, then we have the following:

$$H_{gh} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix}. \tag{7}$$

Similarly, if it is a global similarity transformation, we have the following:

$$H_{gs} = \begin{bmatrix} s \cos \theta & s \sin \theta & t_x \\ s \sin \theta & s \cos \theta & t_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{8}$$

where s and θ are the scale and angle of the transformation, respectively. t_x and t_y are the translation distances along the x and y directions, respectively.

In inhomogeneous coordinates:

$$x'_2 = \frac{h_1x_1 + h_2y_1 + h_3}{h_7x_1 + h_8y_1 + h_9}, \quad y'_2 = \frac{h_4x_1 + h_5y_1 + h_6}{h_7x_1 + h_8y_1 + h_9}. \tag{9}$$

Next, we would like to solve the homography H_{gh} and H_{gs} . Even though these inhomogeneous equations involve the coordinates non-linearly, the coefficients of H appear linearly. Given N pairs of feature correspondences, we can form the following linear system of equations:

$$A\mathbf{h} = \mathbf{0}, \tag{10}$$

where $A \in \mathbb{R}^{2N \times 9}$. It is obtained by transforming Equation (9). Equation (10) can be solved using homogeneous linear least square methods like Singular Value Decomposition (SVD). From SVD, the right singular vector that corresponds to the smallest singular value is the solution \mathbf{h} , and then we reshape \mathbf{h} into the matrix H_{gh} . For similarity transformation, we filter out the similarity matrix corresponding to the rotation angle with the smallest angle according to the feature points as the global similarity H_{gs} .

Global homography has difficulty producing an accurate alignment when the scenes are not coplanar. In contrast, APAP uses spatially varying warping H_* , which expands the homography H into multiple local homographies applied to each grid cell, and it uses moving DLT for image alignment. This allows for better alignment in the overlapping region. Similar to the method in [42], assuming that the feature point p in the image I_a can be represented by a linear combination of a vector composed of four vertices of the mesh $V = [v_1, v_2, v_3, v_4]^T$. It can be expressed as: $p = WV$, and the bilinear interpolation weight of the four mesh vertices $W = [w_1, w_2, w_3, w_4]^T$. Then, these mesh vertices are formed into the transformed mesh vertices $\hat{V} = [\hat{v}_1, \hat{v}_2, \hat{v}_3, \hat{v}_4]^T$ after mesh deformation. Therefore, the deformed vertex coordinate \hat{p} is calculated as the bilinear interpolation as follows:

$$\hat{p} = w_1v_1 + w_2v_2 + w_3v_3 + w_4v_4. \tag{11}$$

As a result, we convert point correspondence constraints to mesh vertex correspondence constraints.

After the above preparation, the energy function can be defined as:

$$E = E_a + E_d + E_s, \quad (12)$$

where E_a addresses the alignment issue by feature correspondences increase model to increase the number of correct matching (see Section 3.2.2), E_d addresses the distortion issue by cross lines with a novel hybrid initial warp prior (see Section 3.2.3), and E_s addresses the salient content-preserving issue by protecting both local and global lines from being bent in the non-overlapping region (see Section 3.2.4). The mesh warping result is obtained by minimizing E (see Section 3.2.5). As all constraint terms are quadratic, they can be solved and minimized by any sparse linear solver.

3.2.2. Alignment Term

The point alignment item E_p is used to align the two images $\{I_a, I_b\}$, which is defined as follows:

$$E_p = \sum_i \|W_i \hat{V}_i - p'_i\|^2 = \|\hat{p} - p'\|^2, \quad (13)$$

where p' indicates the matching points in the image I_b . Line alignment is also considered as an alignment supplement. Given a set of line correspondences $\{l_{aj}, l_{bj}\}$, where $l_{aj} \in I_a$ is represented by the line segment with the endpoints P_{ak} , and $P_{ak} = W_{pk} \hat{V}_{pk}$. $l_{bj} \in I_b$ is represented by the line equation $a_j x + b_j y + c_j = 0$ [18], then the line alignment term is represented as below:

$$E_l = \sum_{j,k} \|l_{bj}^T P_{ak} \lambda_j\|^2, \quad (14)$$

where $\lambda_j = \frac{1}{\sqrt{a_j^2 + b_j^2}}$. With the scalar λ_j , the weight between two endpoints is balanced in terms of geometric meaning.

In summary, the alignment term is represented as:

$$E_a = \lambda_p E_p + \lambda_l E_l. \quad (15)$$

3.2.3. Distortion Term

Both global homography and local homography can cause distortion in the non-overlapping region of the image. This distortion is negligible for content close to the common main plane, but more severe for content further away from the main plane. In order to control the distortion of the target image I_a , a series of horizontal and vertical lines are constructed, called cross lines. They are regarded as the intrinsic structure of the image and are used to reflect the overall warping degree of the target image I_a .

In [40], $s(x, y, k)$ is the slope of the line in I_b corresponding to the line passing (x, y) with slope k in I_a . l_u is the warped horizontal line and l_v is the vertical line that is closest to the boundary between overlapping and non-overlapping areas. If a homography warp is given, there is a unique set of parallel lines corresponding to it. The slopes of the parallel lines before and after warping are as follows:

$$k_1 = -\frac{h_7}{h_8}, \quad s(x, y, k_1) = \frac{h_4 h_8 - h_5 h_7}{h_1 h_8 - h_2 h_7}. \quad (16)$$

Therefore, l_v can be set to be the split line with slope k_1 that is closest to the boundary of the overlapping region and non-overlapping region, and l_u can be set to be orthogonal to l_v before and after warping, as shown below:

$$k_1 \cdot k_2 = -1, \quad s(x, y, k_1) \cdot s(x, y, k_2) = -1, \quad (17)$$

Although cross lines can effectively alleviate the distortion of the global homography and balance the projection and perspective distortion in the non-overlapping region, it

can produce a severely stretched single-perspective stitching result when the scene depth or parallax is large (see Figure 4b). Furthermore, it relies on the initial global homography. If the global homography transformation is misaligned in the overlapping region and severely distorted in the non-overlapping region, the cross lines can only alleviate less distortions and cannot flexibly deal with the distortion and misalignment problems without a fundamental solution that the distortion in the non-overlapping region caused by homography. In [17], the global similarity transformation can effectively mitigate the distortion in the non-overlapping region by iteratively finding the minimum rotation angle of the transformation. Therefore, we mix the global homography transformation with the global similarity transformation, and we adjust the rotation angle of the cross lines by adding weights, so as to reduce the distortion flexibly (as shown in Figure 4d).

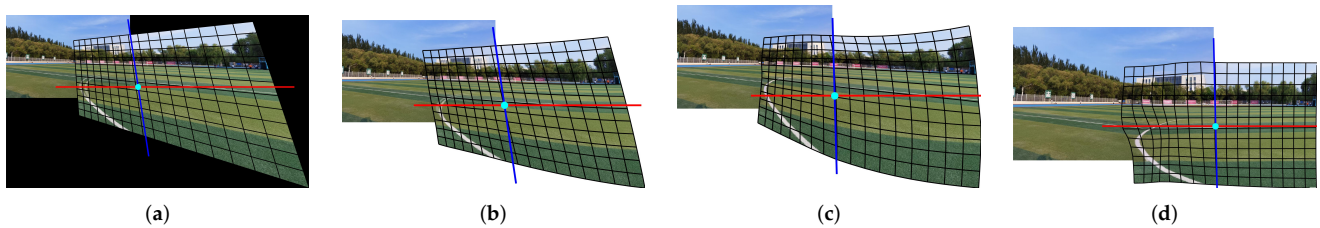


Figure 4. The cross lines distortion mitigation results. (a) Warped mesh result of homography. (b) Warped mesh result of SPW. (c) Warped mesh result of LPC. (d) Warped mesh result of ours. The blue line is closest to the border of the overlapping region and the non-overlapping region, and the red line is perpendicular to it. The smaller the slope of the cross lines, the less distortion.

In this paper, we change the initial global homography to a hybrid global homography transformation and a global similarity transformation as shown in Figure 5. This is defined as follows:

$$H_{init} = H_{gh} + H_{hy}, \quad (18)$$

where H_{init} , H_{gh} , H_{hy} are the initial transformation, the global homography transformation, and the hybrid adjustment terms, respectively. They are all 3×3 matrices. The purpose is to accurately align in the overlapping region, reduce distortion in the non-overlapping region, and flexibly cope with large parallax and large scene depth. The hybrid adjustment term is defined as follows:

$$H_{hy} = W_{hy}(H_{gs} - H_{gh}), \quad (19)$$

where H_{gs} is the global similarity transformation, and W_{hy} is the weighting factor in the range of $[0, 1)$ for adjusting H_{gs} and H_{gh} . If W_{hy} is 0, H_{init} is the global homography transformation H_{gh} , at this time the slope of the cross line in Equation (15) is the largest. Since the overall warping change is the largest, the distortion is the most serious. However, the alignment is better; if W_{hy} is 1, H_{init} is the global similarity transformation H_{gs} , but the slope of the cross lines does not exist. The closer the value of W_{hy} is to 0, the higher the slope; the higher the degree of warp, the better the alignment, but the more serious the distortion. On the contrary, the closer the value of W_{hy} to 1, the smaller the slope, the smaller the transformation angle, the less distortion, but the more serious the misalignment. Therefore, it is necessary to choose the appropriate value of W_{hy} according to different types of images to be stitched. The stitching results at different W_{hy} are shown in Figure 4.

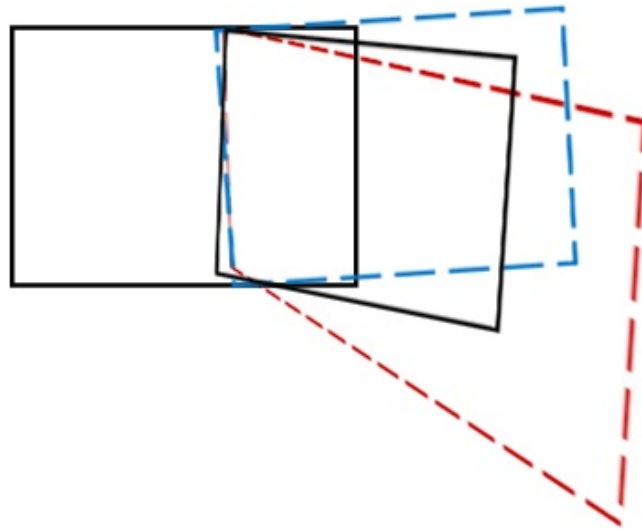


Figure 5. Hybrid global transformation. For the target image, the red dashed frame represents the global homography transformation and the blue dashed frame represents the global similarity transformation. The solid black frame in the middle represents the hybrid global transformation.

By the method discussed above, we divide the distortion term E_d into a perspective distortion term E_{ps} that preserves the perspective given by the new hybrid homography warp in Equation (18) and a projection distortion term E_{pj} that mitigates the projective distortion.

Given the set of cross line correspondences $\{l_{ai}^u, l_{bi}^u\}$ and $\{l_{aj}^v, l_{bj}^v\}$ that are parallel to l_u and l_v , the points of cross lines are recorded as $\{p_k^{ui}\}$ and $\{p_k^{vi}\}$, then

$$\begin{aligned}
 E_{ps} = & \sum_{i=1} \sum_{k=1} |\langle W_{p_{k+1}^{ui}} \hat{V}_{p_{k+1}^{ui}} - W_{p_k^{ui}} \hat{V}_{p_k^{ui}}, \bar{\mathbf{n}}_i^u \rangle|^2 \\
 & + \sum_{j=1} \sum_{k=1} |\langle W_{p_{k+1}^{vj}} \hat{V}_{p_{k+1}^{vj}} - W_{p_k^{vj}} \hat{V}_{p_k^{vj}}, \bar{\mathbf{n}}_j^v \rangle|^2 \\
 & + \sum_{j=1} \sum_{k=1} \|W_{p_k^{vj}} \hat{V}_{p_k^{vj}} + W_{p_{k+2}^{vj}} \hat{V}_{p_{k+2}^{vj}} - 2W_{p_{k+1}^{vj}} \hat{V}_{p_{k+1}^{vj}}\|^2,
 \end{aligned} \tag{20}$$

where $\bar{\mathbf{n}}_i^u$ and $\bar{\mathbf{n}}_j^v$ are the normal vectors of l_{bi}^u and l_{bj}^v . In Equation (19), the first two terms preserve the slopes of l_{bi}^u and l_{bj}^v , and the last term preserve the ratios of lengths on l_{aj}^v .

The projection distortion term for the non-overlapping region of image I_a is defined as follows:

$$E_{pj} = \sum_{i=1} \sum_{k=1} \|W_{p_k^{ui}} \hat{V}_{p_k^{ui}} + W_{p_{k+2}^{ui}} \hat{V}_{p_{k+2}^{ui}} - 2W_{p_{k+1}^{ui}} \hat{V}_{p_{k+1}^{ui}}\|^2. \tag{21}$$

In fact, Equation (20) linearizes the scale on l_{ai}^u in the non-overlapping region of I_a .

In summary, the distortion term is represented as:

$$E_d = \lambda_{ps} E_{ps} + \lambda_{pj} E_{pj}. \tag{22}$$

3.2.4. Salient Term

The salient term is used to preserve the line structure in the non-overlapping region of I_a . Given the set of salient lines $\{I_{ak}^s\}$, each $I_{ak}^s \in I_a$ is denoted by the set of endpoints p_j^k , then:

$$E_{ls} = \sum_{k=1} \sum_{j=1} \|\langle W_{p_{j+1}^k} \hat{V}_{p_{j+1}^k} - W_{p_j^k} \hat{V}_{p_j^k}, \bar{\mathbf{n}}_k \rangle\|^2, \tag{23}$$

where $\bar{\mathbf{n}}_k$ is the normal vector of $\{I_{bk}^s\}$ that is calculated from the new homography H_{init} in Equation (17).

Due to the limitations of line segment detection, the detected line segment is generally not very long, and it may not be the actual line structure affected by the viewing direction or other factors. When some local lines are actually collinear, if they are not globally optimized, there will be significant global line bending after warping. Inspired by [20], we merge some apparently collinear local line segments into global lines, while adding the global line term E_{gs} to the salient term.

Given the set of global lines $\{l_{ag}^s\}$, each $l_{ag}^s \in I_a$ is denoted by the set of endpoints p_k^g , then we have the following:

$$E_{gs} = \sum_{g=1} \sum_{k=1} \|\langle W_{p_{k+1}^g} \hat{V}_{p_{k+1}^g} - W_{p_k^g} \hat{V}_{p_k^g}, \vec{n}_g \rangle\|^2, \tag{24}$$

where \vec{n}_g is the normal vector of $\{l_{bg}^s\}$. Figure 6 also shows the effect of adding the global line term on the stitching results.

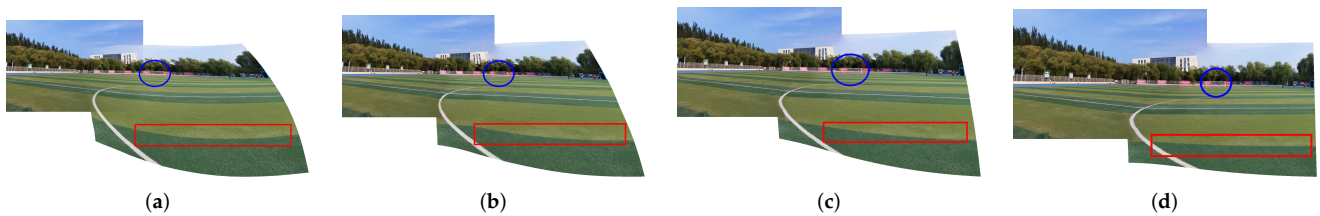


Figure 6. Global lines and hybrid transformation with different W_{hy} . (a) No global lines and $W_{hy} = 0$. (b) With global lines and $W_{hy} = 0$. (c) With global lines and $W_{hy} = 0.4$. (d) With global lines and $W_{hy} = 0.9$. The blue circles show the distortion mitigation effect with different W_{hy} , and the red rectangles show the global lines for structure preservation.

Similarly, local lines term and global line term can be represented as follows:

$$E_s = \lambda_{ls} E_{ls} + \lambda_{gs} E_{gs}. \tag{25}$$

3.2.5. Total Energy Function

The above terms can be written as a total energy function:

$$E = \lambda_p E_p + \lambda_l E_l + \lambda_{ps} E_{ps} + \lambda_{pj} E_{pj} + \lambda_{ls} E_{ls} + \lambda_{gs} E_{gs}. \tag{26}$$

The above function is quadratic, and this turns into an optimization problem which can be solved and minimized by any sparse linear solver.

To sum up, we define alignment terms, distortion terms, and salient terms based on a novel hybrid initial warp prior to address alignment and distortion issues. Among them, the point alignment term in the alignment term is based on our proposed feature correspondences increase model. The stitching process of the two images is shown in Algorithm 1.

Algorithm 1 Stitching two images.**Input:** a target image I_a and a reference image I_b .**Output:** a stitched image.

- 1: Match point and line features between I_a and I_b to obtain $\{p_a, p_b\}$ and $\{l_a, l_b\}$ via feature correspondences increase.
- 2: Calculate a global homography H_{gh} and a global similarity matrix H_{gs} via dual-feature.
- 3: Calculate a hybrid warp H_{init} by Equation (18) and Equation (19).
- 4: Calculate $\{l_{ai}^u, l_{bi}^u\}$ and $\{l_{aj}^v, l_{bj}^v\}$ from H_{init} by Equation (16) and Equation (17).
- 5: Detecting salient line segments in the non-overlapping region to obtain $\{l_{ak}^s\}$.
- 6: Merge collinear local line segments into global lines $\{l_{ag}^s\}$.
- 7: Uniformly sample $\{l_{ai}^u\}, \{l_{ai}^v\}, \{l_{ak}^s\}, \{l_{ag}^s\}$.
- 8: Solve \hat{V} via minimizing the total energy term in Equation (26).
- 9: Warp I_a via the bilinear interpolation on the basis of \hat{V} .
- 10: Stitching the warped I_a with I_b via linear blending or seam-cutting blending.

4. Experiments**4.1. Experimental Setup**

The proposed method has been tested on selected datasets, including some public datasets with 10 pairs of images in [8,10,17,20,34,39], and our own collected datasets with 5 pairs of images. Some input image pairs as examples are shown in Figure 7 for quantitative evaluation and qualitative comparison. All the images are natural RGB ones. In addition to regular natural images, we focus on images with low-textured or repetitively-textured regions in the overlapping region, which will better demonstrate the effectiveness of the proposed method.



Figure 7. Datasets for experimental demonstration. (a–e) are our own images. (f) is from [10].

We compare our approach with some state-of-the-art methods, including homography, APAP, AANAP, SPHP, SPW and LPC. The parameters of the existing methods are set

as suggested by the original papers. In order to facilitate the comparison of alignment effects in the overlapping region and structure preservation effects in the non-overlapping region, while avoiding seam-cutting blending to mask the original misalignment, all the methods use linear blending. In addition, we also show the visual enhancement brought by seam-cutting blending in [37].

In the experiments, we first extract and match ORB feature points, then we remove mismatches with GMS, while using LSD [24] to detect line segments and match them by [21]. The number of ORB feature points is 30,000. Since the image stitching task does not involve rotation and scaling, the parameters of rotation and scaling in GMS are both 0. For the hybrid terms optimization parameters, the deformation mesh size is set to 40×40 , and the inputs λ_p , λ_l , λ_{ps} , λ_{pj} , λ_{ls} , and λ_{gs} are set to 1, 5, 50, 100, 50, and 100, respectively. The initial homography fine-tuning weight W_{hy} is set to 0.4. It is worth noting that W_{hy} and λ_{ps} are inversely proportional. If W_{hy} is too large and the transformation is approximately similar to the global one, the perspective relationship will change significantly. Therefore, the perspective term should be reduced to ensure alignment of the overlapping regions. In this way, the non-overlapping region is close to the rectangle with the least distortion, and the distortion term and the salient term can be appropriately reduced, and vice versa. Since the salient term and the distortion term are competing in [22], the ration of the inputs λ_{pj} and λ_{gs} are 1:1.

The proposed model is mainly implemented in MATLAB (some portions of the code are written in C++), and all the experiments are carried out on a PC with AMD Ryzen 7 4800 H 2.90 GHz CPU and 16 GB RAM.

4.2. Quantitative Evaluation

We quantitatively evaluate the stitching quality of different single-perspective methods using SSIM [43] in the range of [0, 1]. The larger the value of SSIM, the more similar the structure after warping. The overlapping regions of each pair of images in the dataset are extracted to compute the corresponding SSIM after alignment.

Besides, we quantitatively evaluate the alignment accuracy of the proposed method by root mean squared error (RMSE) on feature correspondences:

$$RMSE(f) = \sqrt{\frac{1}{N} \sum_{i=1}^N \|f(p_i) - p'_i\|^2}, \quad (27)$$

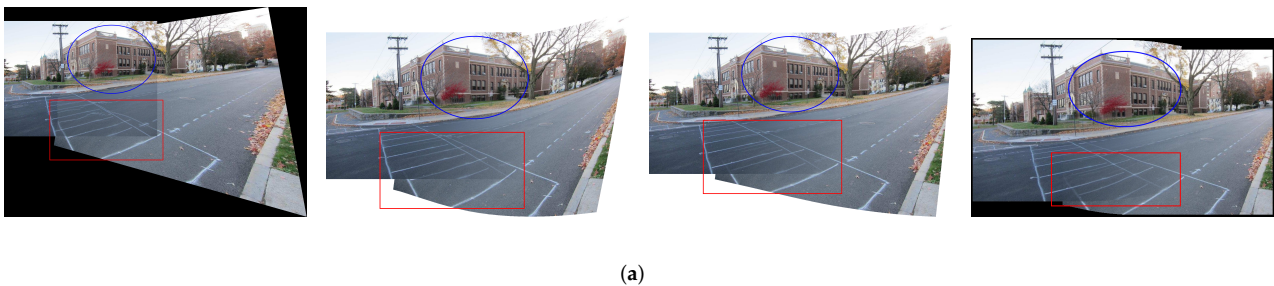
where N is the number of feature correspondences and f represents a planar warp. Table 1 shows the number of matches, SSIM and RMSE using different methods on different datasets. Among them, the first four rows are dense texture images with small parallax, the fifth to eighth rows are images with large parallax and no obvious low-textured areas, and the last seven rows are images with obvious low-textured areas in the overlapping region. The traditional SIFT+RANSAC model is used as a baseline model for comparison. Compared to it, our feature correspondences increase model significantly increases the number of matches generally. The more feature correspondences there are, the greater the image information entropy is, and the better the image detail performance is, so as to obtain a better stitching result. In particular, it allows sufficient matches for grid-based stitching algorithms, especially for low-textured images. Additionally, the values of SSIM and RMSE also show that our method outperforms other methods, especially that our method achieves the highest scores on images with significantly weakly textured or repetitively textured regions. On other types of images, although some scores are low, the scores of our method are comparable with other methods.

Table 1. Comparison on matches, SSIM and RMSE.

Databases	SIFT+RANSAC	Global Homography		APAP		SPW		LPC		OURS		
	Matches	RMSE	SSIM	RMSE	SSIM	RMSE	SSIM	RMSE	SSIM	Matches	SSIM	RMSE
<i>fence</i> [17]	597	0.6489	1.7674	0.7011	1.7773	0.6952	1.4592	0.6807	1.3368	5911	0.7184	1.5014
<i>Pottery</i> [20]	360	0.3971	2.3419	0.5904	2.3419	0.4648	2.0559	0.5144	1.8928	3772	0.6142	1.7384
<i>railtracks</i> [8]	651	0.4474	2.5763	0.6354	2.5767	0.6361	2.217	0.5871	1.5155	6176	0.6738	1.6670
<i>DHW-temple</i> [34]	322	0.5517	2.6633	0.6799	2.6437	0.6057	2.2674	0.4999	2.3199	5474	0.6935	1.6570
<i>MemorialHall</i> [39]	64	0.5370	2.5169	0.5590	2.5169	0.5018	2.2710	0.5217	1.7471	1603	0.5417	1.8711
<i>017</i> [10]	330	0.5871	3.0506	0.6151	3.3019	0.6254	2.3501	0.6139	2.6153	5064	0.6273	2.4313
<i>cup</i> [20]	159	0.4747	2.6436	0.5509	2.1149	0.4778	2.4807	0.4778	2.6069	3041	0.5730	2.0057
<i>office</i> [20]	181	0.5415	3.5423	0.6582	3.5423	0.6150	3.0166	0.6124	2.8839	2991	0.6904	1.9996
<i>intersection</i> [17]	426	0.3612	3.4626	0.4809	3.5073	0.4152	2.5591	0.4285	2.8868	4314	0.5520	1.9684
<i>tower</i> [19]	652	0.5734	3.2967	0.7601	3.2989	0.7702	2.2089	0.8259	1.6391	6409	0.8501	1.6145
<i>runway</i>	208	0.5180	3.0860	0.5892	3.0865	0.5510	2.7476	0.5584	2.0154	3846	0.6632	1.4540
<i>car park</i>	293	0.4044	2.6544	0.4573	2.6544	0.4313	2.4622	0.2328	3.7501	3565	0.4978	2.3074
<i>football field</i>	237	0.4881	3.1110	0.5828	3.1110	0.4872	2.5211	0.4251	1.9414	3641	0.5879	1.6641
<i>sidewalk</i>	245	0.7261	2.3933	0.7681	2.3933	0.7270	1.9704	0.4740	2.9211	3130	0.7953	1.9711
<i>jump runway</i>	117	0.7299	1.8474	0.7457	1.8470	0.7535	1.6250	0.6628	1.9676	2347	0.7499	1.7163

4.3. Qualitative Comparison

First of all, we replace the traditional SIFT/SURF + RANSAC pattern in each method with the proposed feature correspondences increase model. Figure 8 compares the stitching results of some methods before and after applying feature correspondences increase on the dataset *intersection* [17]. As in the original three methods shown in Figure 8a, the inliers are mostly concentrated in the house area as indicated by the blue circle, so the alignment around the house area is better. Since the feature points are only concentrated in the dense-textured region, the feature correspondences of low-textured region is insufficient, which leads to serious misalignment, as shown in the red rectangle. APAP is a grid-based local alignment method, and some pixels will destroy the continuity and consistency of the original texture. However, SPW, LPC and SPHP do not destroy the original texture of the image due to the restriction of alignment terms. Figure 8b shows the results after applying our proposed method. Due to the increase of feature correspondences, the alignment of the three methods is more accurate in low-textured regions. The local alignment property of APAP itself causes it to suffer structural damage, but with less distortion in the non-overlapping region. It also illustrates that grid-based local alignment relies on a sufficient and efficient distribution of the number of feature points. In the case of increased feature correspondences, SPW and LPC ignore the distortion of non-overlapping regions due to over-focusing on alignment. It results in a less smooth transition between overlapping and non-overlapping regions. LPC is better than SPW, but it still has artifacts. SPHP has more accurate alignment compared to using SIFT+RANSAC. However, due to its excellent distortion preserving properties, alignment is not as accurate as SPW and LPC. The above experiments indicate that our proposed feature correspondences increase model can be added to any feature point-based method to replace the original SIFT/SURF+RANSAC pattern to improve alignment accuracy. However, accurate alignment is often accompanied by inevitable distortions in the non-overlapping region. The next issue we want to focus on is how to balance alignment and distortion.

**Figure 8.** Cont.

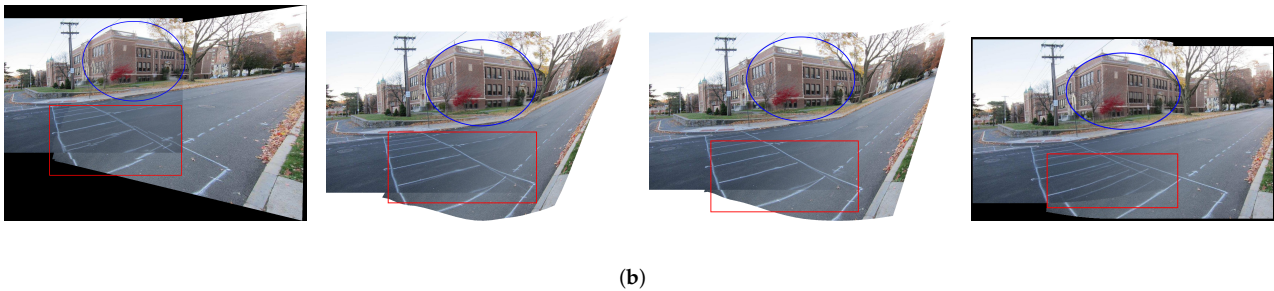


Figure 8. Comparison of the stitching results for increasing the number of features. (a) The original methods of APAP, SPW, LPC and SPHP, respectively. (b) After applying our method. The red boxes represent the low texture region, and the blue circles represent the dense texture region.

Figure 9 shows the stitching results of different methods on the dataset *runway*. The first column shows the stitching results. The runway areas in the red rectangles represent low-textured regions, and the magnified versions are shown in the second column. Similarly, the buildings in the blue rectangles represent the densely-textured regions, and their magnified versions are shown in the third column. Global homography does not yield satisfactory results due to the presence of parallax. There is obvious misalignment, either in the building or the runway. The results of APAP based on local alignment are slightly better than those of global alignment, but the feature points are concentrated in the densely-textured region, while they are insufficient in the low-textured regions. This leads to more accurate alignment in the building and serious misalignment in the runway. Further, global homography and APAP have the same projection and perspective distortion in the non-overlapping regions. The third and fourth rows demonstrate two methods of warping two images. While they effectively mitigate the distortion in the non-overlapping regions, insufficient inliers cause misalignment in low-textured regions, as shown in the red rectangles. The fifth row shows the results of TFA. Although it uses a two-step RANSAC outlier removal method to ensure the matching points as correct as possible, the characteristic of SIFT+RANSAC pattern leads to a serious lack of correct matches in the low-textured regions, which in turn leads to misalignment as shown in the red rectangles. Meanwhile, the triangulation method destroys the original structure, causing the line structure to be bent in the overlapping regions. In the non-overlapping region, its distortion preservation is better than global homography and APAP. In addition, SPW also shows misalignment in the overlapping regions. Although LPC aligned the densely-textured regions, it still caused serious misalignment of the runway areas due to insufficient feature points. Due to its more accurate alignment in the densely-textured regions in the blue rectangles, the distortion in the non-overlapping regions is more severe than SPW. To better demonstrate the effectiveness of our method, we provide three comparisons of results. The eighth row presents the results using only our feature correspondences increase (FCI) model. The ninth row shows the result of using only our hybrid terms optimization (HTO) warp but SIFT+RANSAC as the feature matching stage. The last row is the results of using both FCI and HTO. Using only FCI or HTO may not achieve the results we expect. Alignment using only FCI is significantly more accurate, but there is still slight distortion. Distortion is significantly alleviated by using only HTO but SIFT+RANSAC in the non-overlapping regions, and the target image is more rectangular, but the alignment of overlapping regions is less accurate. When both FCI and HTO are used, our method is aligned in the densely-textured regions. In addition, our alignment is very accurate in the low-textured regions compared to other methods. Furthermore, our method also outperforms SPW and LPC in terms of distortion in the non-overlapping regions. Our method achieves a balance between alignment and distortion to a certain extent.

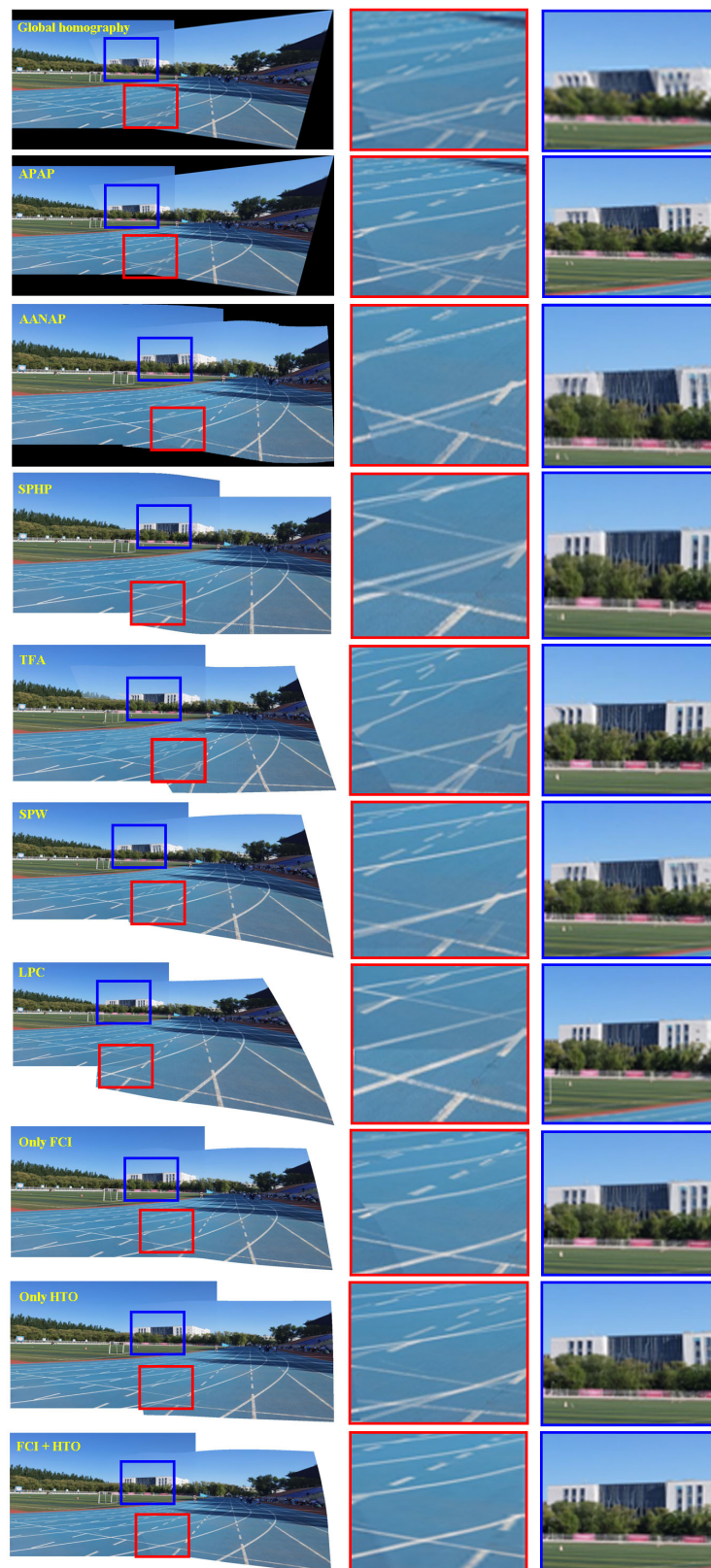


Figure 9. Comparison of stitching quality on *runway*. Row 1: Results using global homography. Row 2: Results using APAP. Row 3: Results of AANAP. Row 4: Results using SPHP. Row 5: Results of TFA. Row 6: Results of SPW. Row 7: Results of LPC. Row 8: Results only using our feature correspondences increase (FCI) model. Row 9: Results only using our hybrid terms optimization (HTO) warp but SIFT+RANSAC. Row 10: Results of our approach by using both FCI and HTO.

4.4. Seam-Cutting Blending

The proposed method has achieved more accurate alignment and it effectively balances alignment and distortion by linear blending. Nonetheless, in some challenging large parallax scenes, our method may produce misalignment if only linear blending is used. Therefore, we introduce seam-cutting blending, which can effectively eliminate the artifacts and misalignment brought by linear blending.

The approach proposed in [10] is a typical seam-driven stitching method. Similar to the other methods we compared, it uses SIFT+RANSAC as feature matching, and it has similar optimization constraints. However, there is no line structure preservation term in its energy function. Figure 10 presents the comparison of [10] and our method on the dataset 025. The red circles highlight the differences between the two methods in the overlapping regions. The results of [10] present significant bending of the line structures at the transparent glass building. In contrast, our method outperforms this approach in terms of line structures preservation since our salient term can effectively protect the line structures from being bent. Besides, our method is also slightly better for distortion in the non-overlapping regions.



(a)



(b)

Figure 10. Comparison of [10] and our method on dataset 025. (a) Stitching result of [10]. (b) Stitching result of our method. The red circles highlight the differences between the two methods in the overlapping regions. Our method protects the line structure better from being bent.

In addition to applying seam-cutting blending to handle large parallax scenes to a certain extent, our method can also deal with continuous structure damage and heavy occlusion issues due to a more accurate alignment. In fact, seam-cutting blending also depends on the pre-alignment. If the pre-alignment is severely misaligned, seam-cutting blending may destroy the original texture structure, resulting in inconsistent and discontin-

uous stitching results due to the inaccurate pre-alignment. Figure 11 shows the difference between LPC and our proposed method using seam-cutting blending on a large parallax example in our own dataset. It can be seen that the linear blending results of our proposed method are better than those of LPC with a different optimal seam. As a result, the seam-cutting blending of LPC destroys the original line structures, resulting in an obvious misalignment. This is shown in the blue circle in Figure 11e. In Figure 11f, our method achieves accurate alignment.

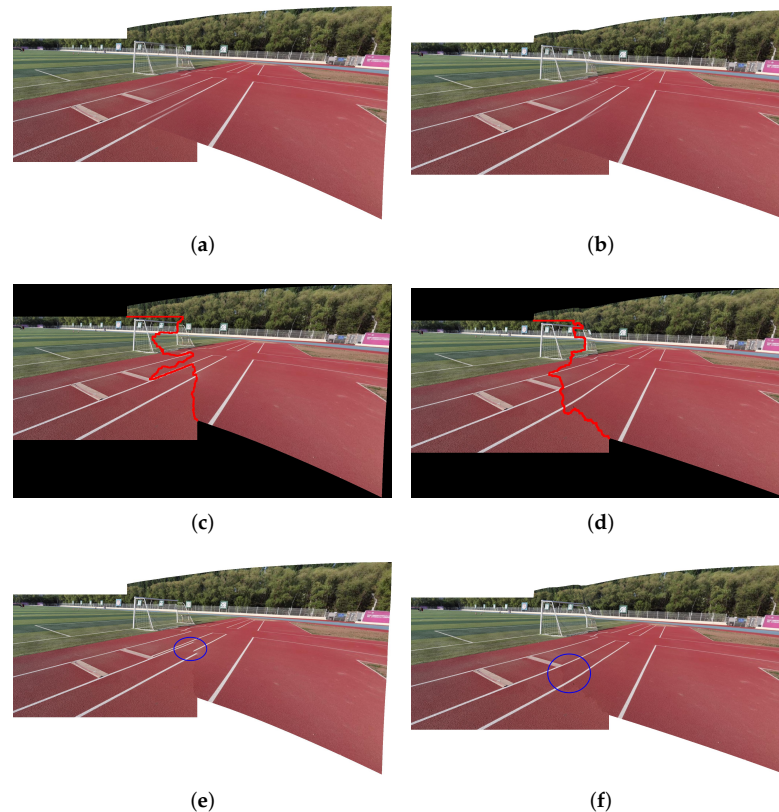


Figure 11. Comparison seam-cutting blending effects of LPC and ours. (a) Stitching result of LPC using linear blending. (b) Stitching result of ours using linear blending. (c) Optimal seam of LPC. (d) Optimal seam of ours. (e) Stitching result of LPC. (f) Stitching result of ours. The difference is shown in the blue circle. It is obvious to see that the seam blending of LPC destroys the original texture structure owing to misalignment seriously. In contrast, our method produces better seam-cutting blending due to more accurate alignment.

Likewise, if the object has a serious occlusion relationship between the foreground and the background, the degree of post-warp alignment will seriously affect the final results. Figure 12 shows an example with a front-to-back occlusion relationship via [19]. It can be seen that the optimal seam of our method effectively avoids the occlusion of the target foreground and background, and achieves accurate alignment (see Figure 12d,f).

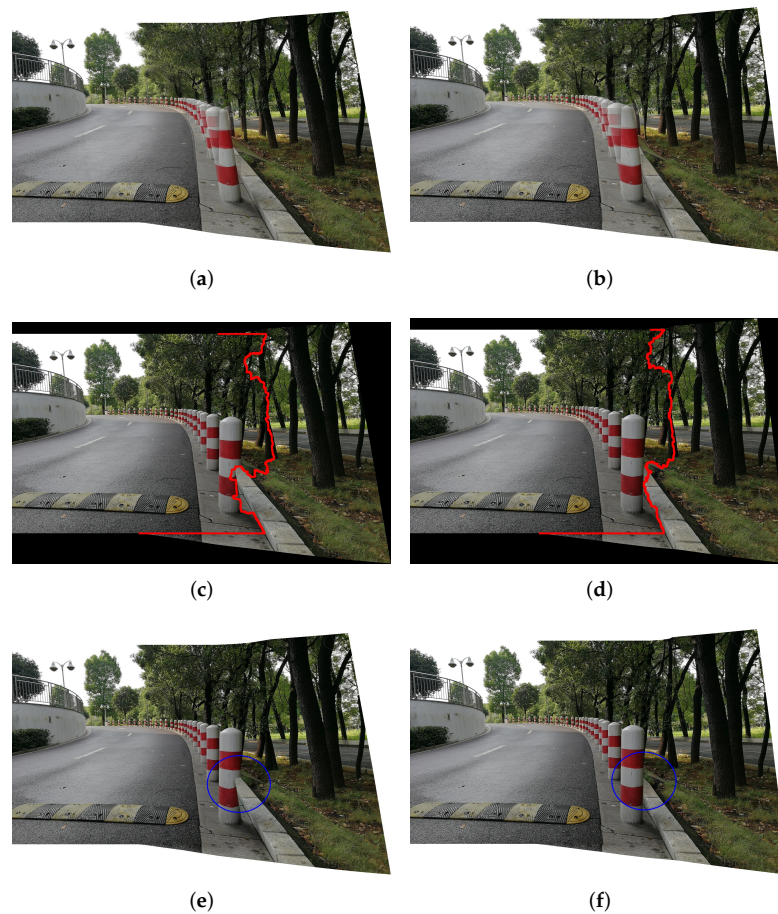


Figure 12. (a–f) Seam-cutting blending results of images with severe occlusions. The red line indicates the optimal seam of LPC and ours. The seam of the LPC destroys the relationship between the foreground and background of the object in the blue circle. However, our method is capable of aligning accurately enough to avoid this problem.

5. Conclusions

In this paper, we propose a novel feature correspondences increase and hybrid terms optimization warp for image stitching. It can deal with the misalignment issue caused by insufficient feature correspondences of low-textured areas in the overlapping region and the distortion issue in non-overlapping regions, with better balancing alignment and distortion. The proposed method with seam-cutting blending can solve the issue of large parallax and occlusion. Both quantitative evaluation and qualitative comparison experiments show that the proposed method is more accurate in alignment and less distortion in the non-overlapping region compared with other methods, especially on images with low-textured areas in the overlapping region.

Author Contributions: Software, writing—original draft preparation, writing—review and editing, Y.C.; formal analysis, Y.W.; project administration, Y.W.; investigation, W.H.; validation, W.P.; supervision, Y.W. and W.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No. 62072212), the Development Project of Jilin Province of China (No. 20220508125RC).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Szeliski, R. Image Alignment and Stitching: A Tutorial. *Found. Trends Comput. Graph. Vis.* **2007**, *2*, 1–104. [[CrossRef](#)]
2. Brown, M.A.; Lowe, D.G. Automatic Panoramic Image Stitching using Invariant Features. *Int. J. Comput. Vis.* **2006**, *74*, 59–73. [[CrossRef](#)]
3. Caesar, H.; Bankiti, V.; Lang, A.H.; Vora, S.; Liong, V.E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; Beijbom, O. nuScenes: A Multimodal Dataset for Autonomous Driving. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 11618–11628.
4. Luan, Y.; Han, C.; Wang, B. An Unsupervised Video Stabilization Algorithm Based on Key Point Detection. *Entropy* **2022**, *24*, 1326. [[CrossRef](#)]
5. Madhusudana, P.C.; Soundararajan, R. Subjective and Objective Quality Assessment of Stitched Images for Virtual Reality. *IEEE Trans. Image Process.* **2019**, *28*, 5620–5635. [[CrossRef](#)] [[PubMed](#)]
6. Chang, C.H.; Chen, C.J.; Chuang, Y.Y. Spatially-Varying Image Warps for Scene Alignment. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; pp. 64–69.
7. Lin, W.Y.; Liu, S.; Matsushita, Y.; Ng, T.T.; Cheong, L.F. Smoothly varying affine stitching. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 345–352.
8. Zaragoza, J.; Chin, T.J.; Tran, Q.H.; Brown, M.S.; Suter, D. As-Projective-As-Possible Image Stitching with Moving DLT. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 2339–2346.
9. Gao, J.; Li, Y.; Chin, T.J.; Brown, M.S. Seam-Driven Image Stitching. In *Proceedings of the Eurographics 2013—Short Papers*; Otaduy, M.A., Sorkine, O., Eds.; The Eurographics Association: Vienna, Austria, 2013. [[CrossRef](#)]
10. Zhang, F.; Liu, F. Parallax-Tolerant Image Stitching. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3262–3269.
11. Lin, K.; Jiang, N.; Cheong, L.F.; Do, M.N.; Lu, J. SEAGULL: Seam-Guided Local Alignment for Parallax-Tolerant Image Stitching. In Proceedings of the ECCV, Amsterdam, The Netherlands, 11–14 October 2016.
12. Nie, L.; Lin, C.; Liao, K.; Liu, S.; Zhao, Y. Unsupervised Deep Image Stitching: Reconstructing Stitched Features to Images. *IEEE Trans. Image Process.* **2021**, *30*, 6184–6197. [[CrossRef](#)] [[PubMed](#)]
13. Zhang, L.; Huang, H. Image Stitching with Manifold Optimization. *IEEE Trans. Multimed.* **2022**. . 2022.3161839. [[CrossRef](#)]
14. Nie, L.; Lin, C.; Liao, K.; Liu, S.; Zhao, Y. Depth-Aware Multi-Grid Deep Homography Estimation With Contextual Correlation. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 4460–4472. [[CrossRef](#)]
15. LoweDavid, G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.
16. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [[CrossRef](#)]
17. Lin, C.C.; Pankanti, S.; Ramamurthy, K.N.; Aravkin, A.Y. Adaptive as-natural-as-possible image stitching. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1155–1163.
18. Li, S.; Yuan, L.; Sun, J.; Quan, L. Dual-Feature Warping-Based Motion Model Estimation. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 4283–4291.
19. Li, J.; Wang, Z.; Lai, S.; Zhai, Y.; Zhang, M. Parallax-Tolerant Image Stitching Based on Robust Elastic Warping. *IEEE Trans. Multimed.* **2018**, *20*, 1672–1687. [[CrossRef](#)]
20. Jia, Q.; Li, Z.; Fan, X.; Zhao, H.; Teng, S.; Ye, X.; Latecki, L.J. Leveraging Line-point Consistency to Preserve Structures for Wide Parallax Image Stitching. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 12181–12190.
21. Jia, Q.; Gao, X.; Fan, X.; Luo, Z.; Li, H.; Chen, Z. Novel Coplanar Line-Points Invariants for Robust Line Matching Across Views. In Proceedings of the ECCV, Amsterdam, The Netherlands, 11–14 October 2016.
22. Liao, T.; Li, N. Single-Perspective Warps in Natural Image Stitching. *IEEE Trans. Image Process.* **2020**, *29*, 724–735. [[CrossRef](#)] [[PubMed](#)]
23. Joo, K.; Kim, N.; Oh, T.H.; Kweon, I.S. Line meets as-projective-as-possible image stitching with moving DLT. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2015; pp. 1175–1179.
24. von Gioi, R.G.; Jakubowicz, J.; Morel, J.M.; Randall, G. LSD: A Line Segment Detector. *Image Process. Line* **2012**, *2*, 35–55. [[CrossRef](#)]
25. Bian, J.; Lin, W.Y.; Matsushita, Y.; Yeung, S.K.; Nguyen, T.D.; Cheng, M.M. GMS: Grid-Based Motion Statistics for Fast, Ultra-Robust Feature Correspondence. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 June 2017; pp. 2828–2837.
26. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G.R. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Tokyo, Japan, 25–27 May 2011; pp. 2564–2571.
27. Harris, C.G.; Stephens, M.J. A Combined Corner and Edge Detector. In Proceedings of the Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988.
28. Rosten, E.; Drummond, T. Machine Learning for High-Speed Corner Detection. In Proceedings of the ECCV, Graz, Austria, 7–13 May 2006.
29. Bay, H.; Tuytelaars, T.; Gool, L.V. SURF: Speeded Up Robust Features. In Proceedings of the ECCV, Graz, Austria, 7–13 May 2006.

30. Calonder, M.; Lepetit, V.; Strecha, C.; Fua, P.V. BRIEF: Binary Robust Independent Elementary Features. In Proceedings of the ECCV, Crete, Greece, 5–11 September 2010.
31. Lin, W.Y.; Cheng, M.M.; Lu, J.; Yang, H.; Do, M.N.; Torr, P.H.S. Bilateral Functions for Global Motion Modeling. In Proceedings of the ECCV, Zurich, Switzerland, 6–12 September 2014.
32. Li, J.; Deng, B.; Tang, R.; Wang, Z.; Yan, Y. Local-Adaptive Image Alignment Based on Triangular Facet Approximation. *IEEE Trans. Image Process.* **2020**, *29*, 2356–2369. [[CrossRef](#)] [[PubMed](#)]
33. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle Adjustment—A Modern Synthesis. In Proceedings of the Workshop on Vision Algorithms, Corfu, Greece, 21–22 September 1999.
34. Gao, J.; Kim, S.J.; Brown, M.S. Constructing image panoramas using dual-homography warping. In Proceedings of the CVPR, Colorado Springs, CO, USA, 20–25 June 2011; pp. 49–56.
35. Chang, C.H.; Sato, Y.; Chuang, Y.Y. Shape-Preserving Half-Projective Warps for Image Stitching. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3254–3261.
36. Liu, F.; Gleicher, M.; Jin, H.; Agarwala, A. Content-preserving warps for 3D video stabilization. *ACM Trans. Graph.* **2009**, *28*, 44. [[CrossRef](#)]
37. Li, N.; Liao, T.; Wang, C. Perception-based seam cutting for image stitching. *Signal Image Video Process.* **2018**, *12*, 967–974. [[CrossRef](#)]
38. Xue, W.; Xie, W.; Zhang, Y.; Chen, S. Stable Linear Structures and Seam Measurements for Parallax Image Stitching. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 253–261. [[CrossRef](#)]
39. Chen, Y.S.; Chuang, Y.Y. Natural Image Stitching with the Global Similarity Prior. In Proceedings of the ECCV, Amsterdam, The Netherlands, 11–14 October 2016.
40. Li, N.; Xu, Y.; Wang, C. Quasi-Homography Warps in Image Stitching. *IEEE Trans. Multimed.* **2018**, *20*, 1365–1375. [[CrossRef](#)]
41. Zhang, Y.; Lai, Y.K.; Zhang, F.L. Content-Preserving Image Stitching With Piecewise Rectangular Boundary Constraints. *IEEE Trans. Vis. Comput. Graph.* **2021**, *27*, 3198–3212. [[CrossRef](#)] [[PubMed](#)]
42. Liu, S.; Yuan, L.; Tan, P.; Sun, J. Bundled camera paths for video stabilization. *ACM Trans. Graph. (TOG)* **2013**, *32*, 1–10. [[CrossRef](#)]
43. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.