Preprints.org

Article

SC-AOF: A Sliding Camera and Asymmetric Optical Flow-Based Blending Method for Image Stitching

Jiayi Chang [*] , Qin Li , Yanju Liang , Liguo Zhou [*]

Article

SC-AOF: A Sliding Camera and Asymmetric Optical Flow-Based Blending Method for Image Stitching

Jiayi Chang [1,2,3,*], Qin Li [2], Yanju Liang [3] and Liguo Zhou [4,*]

[1] Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100049, China;

[2] University of Chinese Academy of Sciences, Beijing 101408,China; liqing@ime.ac.cn

[3] Wuxi Iot Innovation Center Co., Ltd, Wuxi 214135,China; liangyanju@wiot.tech

[4] Technical University of Munich, Munich 80333, Germany

[*] Correspondence: changjy7628@126.com (J.C.); liguo.zhou@tum.de (L.Z.)

Abstract: Parallax processing and structure preservation has long been important and challenging tasks in image stitching. In this paper, an image stitching method based on sliding camera to eliminate perspective deformation and asymmetric optical flow to solve parallax is proposed. By maintaining the viewpoint of two input images in the mosaic non-overlapping area and creating a virtual camera by interpolation in the overlapping area, the viewpoint is gradually transformed from one to another image view-point so as to complete the smooth transition of the two image viewpoints and reduce perspective deformation. Two coarsely aligned warped images are generated with the help of a global projection plane. After that, the optical flow propagation and gradient descent method are used to quickly calculate the bidirectional asymmetric optical flow between the two warped images, and the optical flow-based method is used to further align the two warped images to reduce parallax. In the image blending, the softmax function and registration error are used to adjust the width of the blending area, further eliminating ghosting and reduce parallax. Finally, the comparative experiments show that our method can effectively eliminate perspective deformation, reduce misalignment, and save running time.

Keywords: image stitching; sliding cameras; asymmetric optical flow; image blending

1. Introduction

Image stitching is a technology that can align and blend multiple images to generate a high-resolution, wide-field-of-view and artifact-free mosaic. It has broad and promising applications in many fields such as virtual reality, remote sensing mapping, and urban modeling. The calculation of the global homography, as an important step in image stitching [1,2], directly determines the image alignment accuracy and the final user experience. However, global homography only works for planar scenes or rotation-only camera motions. For non-planar scenes or when the optical centers of cameras do not coincide, homography tends to cause misalignment, resulting in blurring and ghosting in the mosaic. It can also cause perspective deformation, making the final mosaic blurred and severely stretched at the edges. Many solutions have proposed to solve the problems of parallax and perspective deformation in image stitching, so as to improve the quality of stitched images. But most state-of-art mesh-based [3–5]and multi-plane [6–8] method are time-consuming and vulnerable to false matches.

In this work, an innovative image stitching method combining sliding camera (SC) and asymmetric optical flow (AOF), referred to as the SC-AOF method, is proposed to reduce both perspective deformation and alignment error. In the non-overlapping area of the mosaic, the SC-AOF method manages to keep the viewpoint of the mosaic same as or one rotation around camera Z axis from those of input images. In the overlapping area of the mosaic, the viewpoint is changed from one input image viewpoint to another, which can effectively solve the perspective deformation at the edge. A global projection plane is estimated to project input images onto the mosaic; After that, an asymmetric optical flow method is employed to further align the images. In the blending, softmax

function and alignment error are used to dynamically adjust the width of the blending area to further eliminate ghosting and improve the mosaic quality. This paper makes the following contributions:

• The SC-AOF method innovatively uses an approach based on sliding camera to reduce perspective deformation. Combined with either a global projection model or a local projection model, this method can effectively reduce the perspective deformation.
• An optical flow-based image alignment and blending method is adopted to further mitigate misalignment and improve stitching quality of mosaic generated by a global projection model.
• Each step in the SC-AOF method can be combined with other methods to improve the stitching quality of the other methods. '

This article is organized as follows. Section 2 presents the related works. Section 3 firstly introduces the overall method of this article, then an edge stretching reduction method based on sliding camera and a local misalignment reduction method based on asymmetric optical flow are elaborated in detail. Section 4 presents our qualitative and quantitative experimental results compared with other methods. Finally, Section 5 concludes our method.

2. Related Works

For the local alignment, APAP(as-projective-as-possible) [8,9] use weighted DLT (direct linear transform) method to estimate the location-dependent homography and then eliminate misalignment. But if some key-points match incorrectly, the image areas near these key-points may have incorrect homography, resulting in serious alignment errors and distortion. REW (robust elastic warping) [10,11] uses the TPS (thin-plate spline) interpolation method to convert discrete matched feature points into a deformation field, which is used to warp the image and achieve accurate local alignment. TFT (triangular facet approximation) [6] uses the Delaunay triangulation method and the matched feature points to triangulate the mosaic canvas, and the warping inside each triangle is determined by the homography calculated based on the three triangle vertices, so the false matches will lead to serious misalignment. The warping-residual-based image stitching method [7] first estimates multiple homography matrices, and calculates warping residuals of each matched feature point using the multiple homography matrices. The homography of each region is estimated using moving DLT with the difference of the warping residuals as weight, which make the method can handle larger parallax then APAP, but is less robust to the incorrectly estimated homography. NIS (natural image stitching) [12] method estimates a pixel-to-pixel transformation based on feature matches and the depth map to achieve accurate local alignment. In [13], by increasing feature correspondences and optimizing hybrid terms, sufficient correct feature correspondences are obtained in the low-texture areas to eliminate misalignment. The two methods require additional calculations to enhance robustness, but also are susceptible to the uneven distribution and false matches of feature points.

For perspective deformation, SPHP (shape preserving half projective) [14,15] spatially combines perspective transformation and similarity transformation to reduce deformation. Perspective transformation can better align pixels in overlapping areas, and similarity transformation preserves the viewpoint of the original image in non-overlapping areas. AANAP (Adaptive as-natural-as-possible) [16] derives the appropriate similarity transform directly based on matched feature points, and use weights to gradually transit from perspective transform to similarity transform. The transitions from the homography of overlapping area to the similarity matrix of non-overlapping area adopted by SPHP and AANAP are artificial and unnatural, and can generate some "strange" homography matrices, causing significant distortion in the overlapping area. GSP (global similarity prior) [17,18] adds a global similarity priori to constrains the warping of each image so that it resembles a similarity transformation as a whole and avoid large perspective distortion. In SPW (single-projective warp) [19], the quasi-homography warp [20] is adopted to mitigate projective distortion and preserve single perspective. SPSO (Structure Preservation and Seam Optimization) [4] uses a hybrid warping model based on multi-homography and mesh-based warp to obtain precise alignment of areas at different depths while preserving local and global image structures. GES-GSP (geometric structure preserving-global similarity prior) [21] employs deep learning-based edge

detection to extract various types of large-scale edges, and further introduces large-scale geometric structure preserving to GSP to preserve the curves in images and protect them from distortion. These methods all require constructing and solving a super large linear equation to acquire the corresponding coordinates after mesh warping.

Based on the above analysis, generating a natural mosaic quickly and robustly remains a challenging task.

3. Methodology

The flow chart of the SC-AOF algorithm is illustrated in Figure 1. The detail about each of its steps are described below.
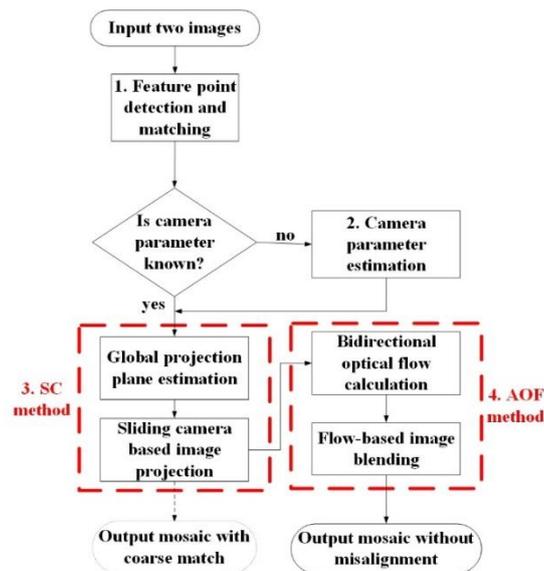

Figure 1. Flow chat of SC-AOF method. After the detection and matching of feature points, the camera parameters are obtained in advance or estimated. Then the two warped images are calculated using SC method, and the mosaic that is coarsely aligned can be obtained. Finally, the AOF method is used to further align the two warped images to generate a blended mosaic with higher alignment accuracy.

Step 1: Feature point detection and matching. SIFT (Scale-invariant feature transform) and SURF (speed-up robust feature) methods are generally used to detect and describe key-points from two input images. Using KNN (k-nearest neighbor) method, a group of matched points is extracted from the key-points and used for camera parameter estimation in step 2 and global projection plane calculation in step 3.

Step 2: Camera parameter estimation. The intrinsic and extrinsic camera parameters are the basis of the SC method, and can be obtained in advance or estimated. When camera parameters are known, we can skip step 1 and directly start from step 3. When camera parameters are unknown, they can be estimated by minimizing the epipolar and planar errors, as described in Section 3.3.

Step 3: Sliding camera based image projection. In this step, we estimate the global projection plane first, then adjust the camera projection matrix and generate a virtual camera in the overlapping area by interpolation, and obtain the warped images by global planar projection, as detailed in section 3.1. Misalignment can be found in the two warped images obtained in current step. Therefore, we need to use AOF method in step 4 to further improve the alignment accuracy.

Step 4: Flow-based image blending. In this step, we first calculate the bidirectional asymmetric optical flow between the two warped images; then further align and blend the two warped images to generate mosaic using the optical flow (see section 3.2 for more details).

3.1. SC: Viewpoint Preservation Based on Sliding Camera

The sliding camera (SC) method is proposed for the first time to solve perspective deformation, and is the first step in the SC-AOF method. For this reason, this section will first introduce the stitching process of this method, and then details how to calculate the global projection plane and the sliding projection matrix required by this method.

3.1.1. SC Stitching Process

In order to ensure that the mosaic can maintain the perspective of the two input images, the SC method is used. That is, in the non-overlapping area, the viewpoints of the two input images are preserved. In the overlapping area, the viewpoint of camera is gradually transformed from I_1 to I_2.

As shown in Figure 2, the image I_1 and I_2 are back-projected onto the projection surface n, so that the corresponding non-overlapping areas Ω_1, Ω_2 and overlapping area Ω_o are obtained. Assume that the pixels in the mosaic I are $u^1, u^2, \ldots, u^8$, which correspond to the sampling points $S_1, S_2, \ldots, S_8$ in the projection surface n. When the sampling points are within the projection area Ω_1 of image I_1, the mosaic is generated from the viewpoint of I_1. S_1, S_2, S_3 are the intersection points of the backprojection lines of u^1, u^2, u^3 in I_1 and the projection surface n. Therefore, $u^i = P_1 S_i (i = 1,2,3)$, where P_1 is the projection matrix of I_1. When the sampling points are within the projection area Ω_2 of image I_2, the mosaic is generated from the camera viewpoint of I_2. Similarly, we obtain S_i and $u^i = P_2 S_i$, where i=6,7,8. In the overlapping area Ω_o of I_1 and I_2, the SC method is used to generate a virtual camera, whose viewpoint gradually transitions from the viewpoint of I_1 to that of I_2. S_4 and S_5 are the intersection points of the back-projection lines of u^4, u^5 in the visual camera and projection plane n, respectively. The virtual camera's image is generated from images I_1 and I_2 using perspective transformation. For examples, pixel u^4 of the virtual camera corresponds to pixel u_1^4 in I_1 and pixel u_2^4 in I_2, and are generated by blending the latter two pixels.
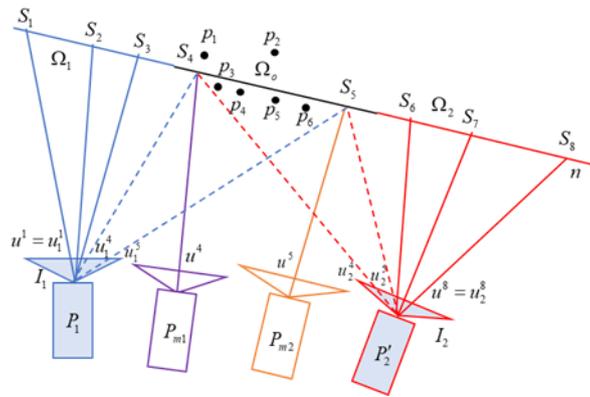

Figure 2. Image stitching based on sliding cameras. n is the projection surface, which is fitted by scene points $p_1, p_2, \ldots, p_6$. Stitched image I can be generated by projection of sampling points $S_1, S_2, \ldots, S_8$. The points S_1, S_2, S_3 in the area Ω_1 are generated by back-projection of pixels in I_1. Similarly, the points S_6, S_7, S_8 in the area Ω_2 are generated by back-projection of pixels in I_2. The points S_4, S_5 in the area Ω_o are generated by back-projection of pixels in virtual cameras. The pixel values of S_4, S_5 correspond to the fused pixel values of projection in I_1 and I_2.

Global projection surface calculation. In order to match the corresponding pixels u_1^4 of I_1 and u_2^4 of I_2, the projection surface n needs to be as close as possible to the real scene point, we can use the moving plane method [7–9] or the triangulation method [6] to obtain a more accurate scene surface. Since the SC-AOF method will use the optical flow to further align the images, for the stitching speed and stability, only the global plane is calculated as the projection surface. Section 3.1.2 will calculate the optimal global projection surface using the matched points.

Sliding camera generation. Generally, since the pixel coordinates of I_1 and I_2 are not uniform, in the mosaic I, when $I(\tilde{u}) = I_1(P_1 S)$ in the non-overlapping area of I_1, $I(\tilde{u}) = I_2(P_2 S)$ is false in the

non-overlapping area of I_2, where S is the sampling point on the projection surface. It is necessary to adjust the projection matrix of I_2 to P_2', so that $I(\tilde{u}) = I_2(P_2'S)$. As the red camera is shown in Figure 2. Section 3.1.3 will deduce the adjustment method of the camera projection matrix, and interpolate in the overlapping area to generate a sliding camera, and obtain the warped images of I_1 and I_2.

3.1.2. Global Projection Surface Calculation

The projection matrices of cameras C_1 and C_2 corresponding to images I_1 and I_2 are:

$$P_1 = K_1[I|0] \quad P_2 = K_2R[I|-t] \tag{1}$$

where K_1 and K_2 are the intrinsic parameter matrices of C_1 and C_2 respectively; R is the inter-camera rotation matrix; and t is location of the optical center of C_2 in the coordinate system C_1.

The relationship between the projection u_1 in I_1 and the projection u_2 in I_2 of 3D point p on plane n is:

$$\tilde{u}_2 \sim K_2R(I_{3\times3} + tn^T)K_1^{-1}\tilde{u}_1 = H\tilde{u}_1 \tag{2}$$

where $\tilde{u}_1$ and $\tilde{u}_2$ are the homogeneous coordinates of u_1 and u_2 respectively; The intersection point p satisfies $n^Tp + 1 = 0$. $\sim$ means that $\tilde{u}_2$ is parallel to $H\tilde{u}_1$.

If camera parameters K_1, K_2, R and t are known, then we can deduce the following (3) from (2)

$$n^T\tilde{y}_1 = -\frac{(R^T\tilde{y}_2 \times t)^T(R^T\tilde{y}_2 \times \tilde{y}_1)}{(R^T\tilde{y}_2 \times t)^T(R^T\tilde{y}_2 \times t)} = b \tag{3}$$

where $\tilde{y}_1 = K_1^{-1}\tilde{u}_1$ and $\tilde{y}_2 = K_2^{-1}\tilde{u}_2$ are the normalized coordinates of $\tilde{u}_1$ and $\tilde{u}_2$.

We use (3) of all matched points to construct a overdetermined equation and obtain the fitted global plane n by solving this equation. Since the optical flow-based stitching method will be used to further align the images, the RANSAC method is not used here to calculate the plane with the most inliers. Instead, the global plane that fits all feature points as closely as possible is selected, misalignment caused by global plane projection will be better sloved during optical flow blending.

3.1.3. Projection Matrix Adjustment and Sliding Camera Generation

To preserve the viewpoint in the non-blending area of I_2, it is only required to satisfy $I(\tilde{u}) = I_2(N\tilde{u}) = I_2(\tilde{u}_2)$. N is a similarity transformation between I_2 and I, and can be obtained by fitting the matched feature points:

$$N_* = \min_S \sum_{j=1}^{n} ||N\tilde{u}_1^j - \tilde{u}_2^j||^2 \tag{4}$$

where $\tilde{u}_1^j$ and $\tilde{u}_2^j$ are the homogeneous coordinates of pixels in I_1 and I_2 respectively.

Therefore, in the non-overlapping area of I_2, $\tilde{u} = N_*^{-1}\tilde{u}_2 = N_*^{-1}K_2R_2(S-t)$, where S is the corresponding 3D point of $\tilde{u}_2$ on plane n. So we get the projection matrix $P_2' = N_*^{-1}K_2R_2$.

By RQ decomposition, the internal parameter matrix K_2' and rotation R_2' are extracted from P_2':

$$\begin{aligned} N_*^{-1}K_2R_2 &= K_2(K_2^{-1}N_*^{-1}K_2)R_2 \\ &= K_2(K_*R_*)R_2 = K_2'R_2' \end{aligned} \tag{5}$$

where K_2' and R_2' are upper triangular matrix and rotation matrix respectively; and the third line of both matrices is $(0 \quad 0 \quad 1)$.

Compared with P_2, P_2' has a different intrinsic parameter matrix, and its rotation matrix only differs by one rotation around Z axis, and its optical center t is not changed.

From I_1 to I_2, the projection matrix transitions from P_1 to P_2'. K_m, R_m and t_m in the projection matrix $P_m = K_m[R_m|-t_m]$ of the virtual camera can be calculated by weighting the corresponding intrinsic parameter matrices, rotation matrices and translation vectors of I_1 and I_2:

$$t_m = (1-m)0_{3\times3} + mt \tag{6}$$

$$K_m = (1 - m)K_1 + mK_2' \tag{7}$$

$$q_m = \frac{sin((1-m)\theta)}{sin(\theta)}q_1 + \frac{sin(m\theta)}{sin(\theta)}q_2 \tag{8}$$

where q_1, q_2, q_m represent the quaternions corresponding to $I_{3\times3}$, R_2' and R_m, θ is the angle between q_1 and q_2, and m is the weighting coefficient.

The weighting coefficient m can be calculated by the method in AANAP [16]:

$$m = \langle \overrightarrow{k_m P_*}, \overrightarrow{k_m K_M} \rangle / |\overrightarrow{k_m K_M}|^2 \tag{9}$$

The meaning of each variable is shown in Figure 3. In the overlapping area, if u corresponds to sliding camera (K_m, R_m, t_m), then the relation between u and u_i in $I_i(i = 1,2)$ can be expressed as:

$$\tilde{u} = K_m R_m (I + t_m n^T/d)K_1^{-1}\tilde{u}_1 = H_m^1 \tilde{u}_1 \tag{10}$$

$$\tilde{u} = H_m^1 H^{-1}\tilde{u}_2 = H_m^2 \tilde{u}_2 \tag{11}$$

(10) and (11)are also applicable to non-overlapping area. Projecting I_1 and I_2 through H_m^1 and H_m^2 onto the mosaic respectively to get warped images I_1' and I_2'. Obviously:

$$I_i'(\tilde{u}) = I_i((H_m^i)^{-1}\tilde{u})(i = 1,2) \tag{12}$$

Figure 4 shows the experiment result on two school images used in [10]. Due to the parallax between I_1 and I_2, blending I_1' and I_2' will cause ghosting. Therefore, the next section will use an optical flow based blending method (AOF) to further align the images.
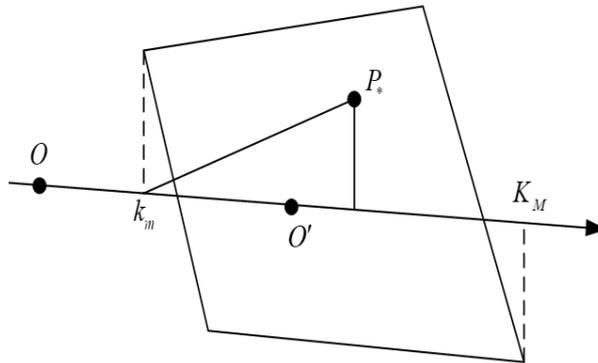

Figure 3. The diagram of gradient weight. A quadrilateral is the boundary of the overlapping area of I_1 and mapped image of I_2 using H^{-1}. where O is the center of I_1 and O' is the warped point of the center point of I_2 using H^{-1}. k_m and K_M are the projection points closest to O and O' on the line OO' of the quadrilateral vertices, respectively. P_* indicates the pixel coordinates within the overlapping area that need to calculate weighted parameter m.

(**a**)   (**b**)
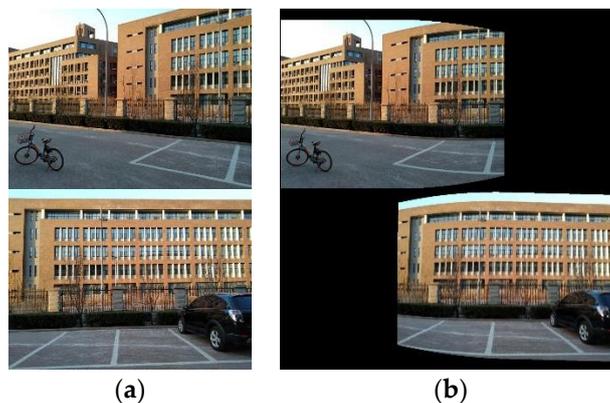
(c)

Figure 4. Image stitching based on sliding cameras and global projection plane. (a) shows the two original images to be stitched together I_1 and I_2; (b) shows the mapped images I_1' and I_2'; (c) shows the average blending images of I_1' and I_2'. That is, in the overlapping area, the blended value is $(I_1' + I_2')/2$.

3.2. SC: Viewpoint Preservation Based on Sliding Camera

The mosaic generated by the SC method will inevitably has misalignment in most cases. So the optical-flow-based method is further employed to achieve more accurate alignment. This section firstly introduces the image alignment process based on asymmetric optical flow (AOF), and then details the calculation method of AOF.

3.2.1. Image Fusion Process of AOF

I_1 and I_2 are projected onto the custom projection surface to obtain warped images I_1' and I_2', which are blended to generate the mosaic I . As the 3D points of the scene are not always on the projection plane, ghosting artifacts can be seen in the mosaic, as shown in Figure 4 in the above section. Direct multi-band image blending will lead to artifacts and blurring. As shown in Figure 5, point P is projected to two points p_1 and p_2 in the mosaic, resulting in duplication of content. To solve the ghosting problem in the mosaic, the optical flow-based blending method in [22] is adopted.

Suppose $F_{2 \rightarrow 1}(p_2)$ represents the optical flow value of p_2 in I_2' and $F_{1 \rightarrow 2}(p_1)$ represents the optical flow value of p_1 in I_1'. If the blending weight of pixel $\tilde{p}$ in the overlapping area is λ (from the non-overlapping area of I_1' to the non-overlapping area of I_2', λ gradually transitions from 0 to 1, as shown in Figure 5, then after blending, the pixel value of image I at $\tilde{p}$ is:
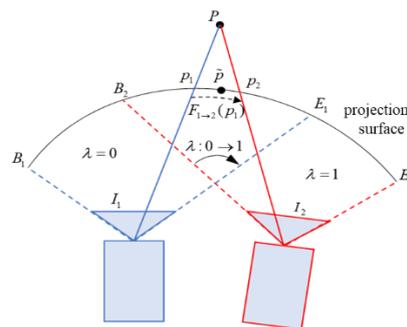

Figure 5. Image blending based on optical flow. B_1E_2 is the projection surface of the mosaic. In the overlapping areas (denoted by B_2E_1) of I_1 and I_2, we need to blend I_1' and I_2'. The 3D point P is outside the projection surface. when P is projected onto the projection surface, ghosting points p_1 and p_2 points appear. Through the weighted blending of asymmetric optical flow, the p_1 and p_2 are merged into point $\tilde{p}$, which solves the ghosting problem of stitching.

$$I(\tilde{p}) = (1 - \lambda)I_1'(\tilde{p}_1) + \lambda I_2'(\tilde{p}_2) \tag{13}$$

Where $\tilde{p}_1 = \tilde{p} + \lambda F_{2\rightarrow1}(\tilde{p})$ represents the corresponding value of $\tilde{p}$ in I_1', $\tilde{p}_2 = \tilde{p} + (1 - \lambda)F_{1\rightarrow2}(\tilde{p})$ represents the corresponding value of $\tilde{p}$ in I_2'. That is, for any pixel $\tilde{p}$ in the overlapping area of the mosaic, its final pixel value can be obtained by a weighted combination of its corresponding values in the two warped images using optical flow.

In order to get a better blending effect, following the method presented by Meng and Liu [23], a softmax function is used to make the mosaic transition quickly from I_1' to I_2', narrowing the blending area. And if the optical flow value of a warped image is larger, the salience is higher, and the blending weight of the warped image should be increased accordingly. Therefore, the following blending weight β can be used as the blending weight:

$$\beta = \frac{exp(\alpha_s\lambda(1 + \alpha_m M_2))}{exp(\alpha_s(1 - \lambda)(1 + \alpha_m M_1)) + exp(\alpha_s\lambda(1 + \alpha_m M_2))} \tag{14}$$

where $M_1 = ||F_{2\rightarrow1}(\tilde{p})||$ and $M_2 = ||F_{1\rightarrow2}(\tilde{p})||$ represents the optical flow magnitude; α_s is the shape coefficient of the softmax function; and α_m denotes the enhancement coefficient of the optical flow. The larger α_s and α_m are, the closer β is to 0 or 1, and the smaller the image transition area.

Also, similar to multi-band blending, a wider blending area is used in smooth and color-consistent areas, and a narrower blending area is used in color-inconsistent area. And the pixel consistency is measure using D_c:

$$D_c = ||I_1(\tilde{p}_1) - I_2(\tilde{p}_2)|| \tag{15}$$

The final blending parameter α is obtained：

$$\lambda_d = \tanh(c_d D_c) \tag{16}$$

$$\alpha = (1 - \lambda_d)\lambda + \lambda_d\beta \tag{17}$$

β corresponds to a fast transition from I_1' to I_2', λ corresponds to a linear transition from I_1' to I_2'. When the color differs smally, he transition from I_1' to I_2' is linear, and when the color difference is large, we tend to have a fast transition from I_1' to I_2'.

Then the pixel value of the mosaic is:

$$I(\tilde{p}) = (1 - \alpha)I_1(\tilde{p}_1) + \alpha I_2(\tilde{p}_2) \tag{18}$$

The curve in the left panel of Figure 6 shows the curve of β changing with λ under different optical flow intensities. β can be used to achieve quick transition of the mosaic from I_1' to I_2', narrowing the transition area. In case of large optical flow, the blending weight of the corresponding image can be increased to reduce the transition area. The curve in the right panel of Figure 6 shows the impact of λ_d on the curve of α as a function of λ. When λ_d is small, a wider fusion area tends to be used; otherwise, a narrower fusion area is used, which is similar to the blending of different frequency bands in multi-band blending method.
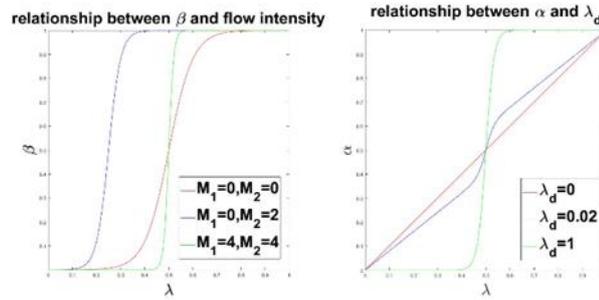
Figure 6. Blending parameter curves. The figure on the left shows the β curves at different optical flow intensities. The right figure shows the α curve at different λ_d values.

3.2.2. Calculation of Asymmetric Optical Flow

The general pipeline of the optical flow calculation is to construct an image pyramid, calculate the optical flow of each layer from coarse to fine, and use the estimated current-layer optical flow divided by the scaling factor as the initial optical flow of the finer layer until the optical flow of the finest layer is obtained [23–25]. Different methods are proposed to achieve better solutions that satisfy brightness constancy assumptions, solve large displacements and appearance variation [27,28], address edge blur and improve temporal consistency [29–31]. Recently, some deep learning methods are proposed. For example, RATF(recurrent all-pairs field transforms for optical flow)[32] extracts per-pixel features, builds multi-scale 4D correlation volumes for all pairs of pixels, and iteratively updates a flow field through a recurrent unit. FlowFormer(optical flow Transformer) [33] based on transformer neural network architecture designs a novel encoder which effectively aggregates cost information of correlation volume into compact latent cost tokens, and a recurrent cost decoder which recurrently decodes cost features to iteratively refine the estimated optical flows.

In order to improve the optical flow calculation speed, we use the method based on optical flow propagation and gradient descent adopted in Facebook surround360 [34] to calculate the optical flow. When calculating the optical flow of each layer, first calculate the optical flow of each pixel from top to bottom and from left to right. From the optical flow values of the current-layer left and top pixels and upper-layer same-position pixel, the value with minimum error represented by (19) is selected as the initial value of the current pixel. Then a gradient descent method is performed to update the optical flow value of the current pixel, and then spread to the right and bottom pixels, as a candidate for the initial optical flow of the right and bottom pixels. After completing the forward optical flow propagation from top to bottom and from left to right, perform a reverse optical flow propagation and gradient descent from bottom to top and from right to left to obtain the final optical flow value.

When calculating the optical flow value $F(u)$ of pixel u, the error function $E(F(u))$ used is:

$$E(F(u)) = E_I + \alpha_S E_S + \alpha_T E_T \tag{19}$$

$$E_I(F(u)) = ||\nabla I_1(u) - \nabla I_2(u + F(u))|| \tag{20}$$

$$E_S(F(u)) = ||F(u) - G(u; \sigma) * F(u)|| \tag{21}$$

$$E_T(F(u)) = ||F(u) * D(1/W, 1/H)|| \tag{22}$$

where, E_I denotes the optical flow alignment error of the edge image (which is Gaussian filtered to improve the robustness), E_S denotes the consistency error of the optical flow, $G(u; \sigma) * F(u)$ denotes the Gaussian-filtered optical flow of pixel u; E_T denotes the magnitude error after normalization of optical flow, with excessively large optical flow being penalized; W and H are the width and height of current-layer image, respectively; and $D(1/W, 1/H)$ denotes the diagonal matrix with diagonal elements $1/W$ and $1/H$.

3.3. *Estimation of Image Intrinsic and Extrinsic Parameters*

The SC-AOF method requires a known camera parameters of images I_1 and I_2. When only the intrinsic parameters K_1 and K_2 of an image are known, the essential matrix $[t]_\times R$ between two images can be obtained by feature point matching, and the rotation matrix R and translation vector t between images can be obtained by decomposing the essential matrix. When both intrinsic and extrinsic parameters are unknown, the intrinsic parameters can be estimated by calibration [35,36] firstly, and then the extrinsic parameters of the image can be estimated accordingly. In these cases, both intrinsic and extrinsic parameters of image I_1 and I_2 can be estimated robustly.

When none of the above methods is feasible, it is necessary to calculate the fundamental matrix from the matched feature points, and restore the camera internal and external parameters.

When camera has zero skew, known principal point and aspect ratio, then each intrinsic parameter matrix has only 1 degree of freedom (focal length of camera). The total degree freedom of camera parameters is 7(t has 2 degrees of freedom because the scale cannot be recovered, R has three degrees of freedom, and each camera has 1 degree of freedom), which is equal to the fundamental matrix F's degree of freedom. The internal and external parameters of the image can be recovered using a self-calibration method[37]. But even if these constraints are met, The solved camera parameters by [37] suffers from large errors when the scene is approximately planar and the matching error is large. Therefore, here we use the method of optimizing the objective function in [6] to solve the internal and external parameters of the camera.

In order to get an accurate fundamental matrix, firstly, the feature points need to be distributed more evenly in the image. As shown in Figure 7, a uniform and sparse distribution of feature points can both reduce the computation time and obtain more robust intrinsic and extrinsic camera parameters and global projection plane, which will lead to improved stitching results.
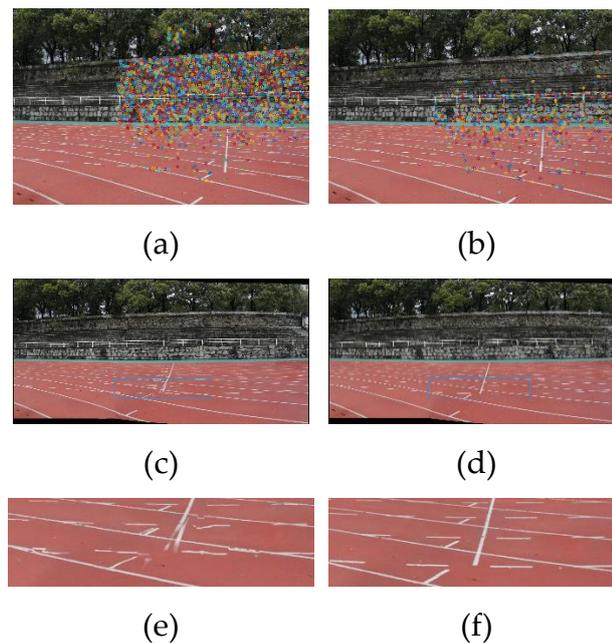

(a)   (b)

(c)   (d)

(e)   (f)

Figure 7. The impact of feature point distribution on stitching results. The feature points in (a) are concentrated in the grandstand. The corresponding mosaic (c) is misaligned in the playground area. The feature points in (b) are evenly distributed within a 2*2 grid. Although the total number of feature points is smaller, but the mosaic (d) has better quality. (e) and (f) show the detail of mosaics.

Secondly, it is necessary to filter the matched feature points to exclude the influence of outliers. Use the similarity transformation to normalize the matched feature points. After normalization, the mean value of the feature points is 0, and the average distance to the origin is $\sqrt{2}$.

Thirdly, multiple homographies are estimated to exclude outlier points. Let F_{cond} and n denote all matched feature points and the total number of matched feature points. In F_{cond}, The RANSAC method with threshold $\eta = 0.01$ is applied to compute homography H^1 and its inlier set F_{inlier}^1, and

the matches of isolated feature points which have no neighboring points within 50 pixel distance is removed from F_{inlier}^1. A new candidate set is generated by removing F_{inlier}^1 from F_{cond}. Repeat the above steps to calculate m homography matrices H^m and corresponding inlier set F_{inlier}^m until $||F_{inlier}^m||<20$ or $|F_{inlier}^m|| < 0.05n$. The final inlier set is $F_{inlier} = \cup_{i=1}^{m} F_{inlier}^i$. If $m = 1$, then there is only one valid plane. In this case, apply the RANSAC method with threshold $\eta = 0.1$ to recalculate homography H^1 and the corresponding inlier set F_{inlier}^1.

After excluding the outliers, for any matched points $\{x_1, x_2\}$ in the inlier set F_{inlier}, the cost function is:

$$E(x_1, x_2) = (1 - \lambda)h(r_e, \sigma_e) + \lambda h(r_p, \sigma_p) \qquad (23)$$

where the λ balanced epipolar constraint and the infinite homography constraint, and generally take $\lambda = 0.01$. h is a robust kernel function, which can mitigate the effect of mis-matched points on the optimization of camera internal and external parameters. r_e and r_p denote the projection errors of the epipolar constraint and of the infinite homography constraint, respectively:

$$r_e = \frac{1}{\rho} x_2^T K_2^{-T} [t]_\times R K_1^{-1} x_1 = \frac{1}{\rho} x_2^T F x_1 \qquad (24)$$

$$r_p = x_2 - \frac{1}{\omega} K_2 R K_1^{-1} x_1 \qquad (25)$$

where ρ denotes the length of the vector composed of the first two components of $F x_1$. That is, assuming $F x_1 = (a, b, c)^T$, then $\rho = \sqrt{a^2 + b^2}$. w represents the third component value of the vector $K_2 R K_1^{-1} x_1$.

4. Experiment

To verify the effectiveness of the SC-AOF method, The mosaics generated by our method and the existing APAP [4], AANAP [16], SPHP [14], TFT [7], REW [10] and SPW [18] are compared on some typical datasets used by other methods to verify the feasibility and advantages of the SC-AOF method in solving deformation and improving alignment; Next, SC-AOF method is used together with other method to demonstrate its good compatibility.

4.1. Effectiveness Analysis of SC-AOF Method

In this section, various methods of image stitching are compared and analyzed based on three indicators: perspective deformation, local alignment and running speed. The experimental setup is as follows.

- The first two experiments compare typical methods for solving perspective deformation and local alignment respectively, and all the methods in the first two experiments are included in the third experiment to show the superiority of SC-AOF method in all aspects.
- Since the averaging methods generally underperforms linear blending ones, all methods to be compared adopt linear blending to achieve the best performance.
- All methods other than ours use the parameters recommended by their proposers. Our SC-AOF method has the following parameter settings in optical flow-based image blending: $\alpha_s = 10$, $\alpha_m = 100$, and $c_d = 10$

4.1.1. Perspective Deformation Reduction

Figure 8 shows the results of the SC-AOF method versus SPHP, APAP, AANAP and SPW methods for perspective deformation reduction in image stitching. School, building and park square datasets were used in this experiment. We can see from Figure 8 that, compared with the other methods, our SC-AOF method makes the viewpoint of the stitched image change in a more natural manner and effectively eliminates perspective deformation. As explained below, all other methods underperform our SC-AOF method.

The image stitched using the APAP method has its edges stretched to a large extent. This is because it does not process perspective deformation. This method only serves as a reference to verify the effectiveness of perspective deformation reducing algorithms.

The AANAP algorithm can achieve a smooth transition between the two viewpoints, but results in severely "curved edges". And there is even more severe edge stretching for the park square dataset than that of the APAP method. This is because, when the AANAP method extrapolates from homographies, it linearizes the homography in addition to similarity transformation, causing affine deformation in the final transformation.
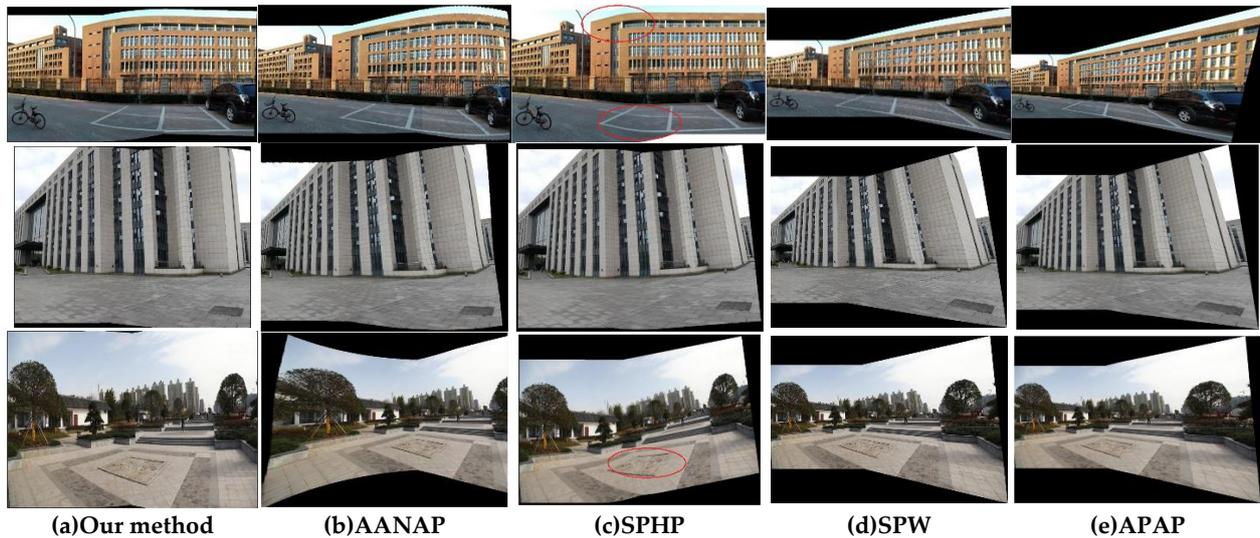

(a)Our method   (b)AANAP   (c)SPHP   (d)SPW   (e)APAP

Figure 8. Comparison of perspective deformation processing. (a), (b), (c), (d) and (e) are the mosaics generated by our method, AANAP, SPHP, SPW and APAP on the datasets , respectively.

Compared with the APAP method, the SPW method makes no significant improvement in perspective deformation, except for the image in the first row. SPW preserves perspective consistency, so multiple-viewpoint method excels in solving perspective deformation compared to single-viewpoint method.

The SPHP algorithm performs well overall. However, it causes severe distortions in some areas (red circled in Figure 8c) due to the rapid change of viewpoints. This is because the SPHP method estimates the similarity transformation and interpolated homographies from global homography. As a result, the similarity transformation can't not reflect the real scene information and the interpolated homographies may deviate from a reasonable image projection.

4.1.2. Local Alignment

Figure 9 shows the results of the SC-AOF method versus APAP, TFT and REW methods for local alignment in image stitching. It can be seen that SC-AOF performs well in all scenes, showing the effectiveness of our method in local alignment.

- The APAP method performs fairly well in most images, though with some alignment errors. This is because the moving DLT method smooths the mosaics to some extent.
- The TFT generated stitched image of excellent quality in planar areas. But when there is a sudden depth change in the scene, there will be serious distortions. This is because large errors appear when calculating plane using three vertices of a triangle in the area with sudden depth changes.
- The REW method has large alignment error in the planar area and aligns the images better than the APAP and TFT method in all other scenes. This is because the fewer feature points in the planar area might be filtered out as mismatched points by the REW method.

The SSIM (structural similarity) is employed to objectively describe the alignment accuracy of different methods. The scores of all methods on the datasets temple, rail tracks, garden, building, school, park-square, wall, and cabinet are listed in Table 1.

Table 1. Comparison on SSIM.

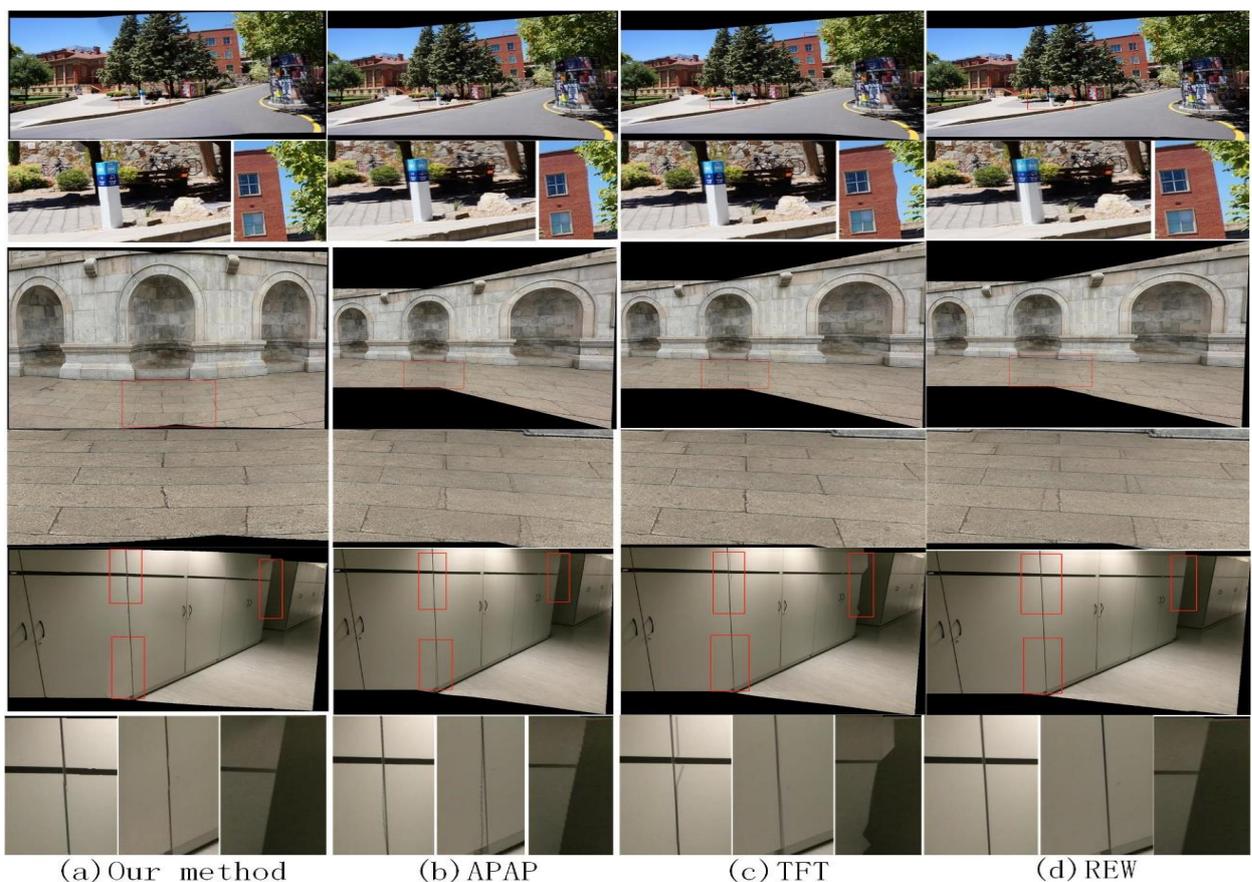|	APAP	AANAP	SPHP	TFT	REW	SPW	Ours
temple	0.90	0.91	0.73	0.95	0.94	0.85	0.96
rail-tracks	0.87	0.90	0.62	0.92	0.92	0.87	0.93
garden	0.90	0.94	0.81	0.95	0.95	0.92	0.93
building	0.93	0.94	0.89	0.74	0.96	0.90	0.96
school	0.89	0.91	0.67	0.90	0.91	0.87	0.93
wall	0.83	0.91	0.68	0.90	0.93	0.81	0.92
park-square	0.95	0.96	0.80	0.97	0.97	0.95	0.97
cabinet	0.97	0.97	0.87	0.97	0.98	0.97	0.96

Figure 9. Comparison of image registration. (a), (b), (c) and (d) show the alignment results of our method, APAP, TFT and REW on the datasets garden, wall, and cabinet, respectively.

- APAP and AANAP have high scores on all image pairs, but the scores are lower than our method and REW, proving that APAP and AANAP blur mosaics to some extent.
- When SPHP is not combined with APAP, only the global homography is used to align the images, resulting in lower scores compared to other methods.
- TFT has higher scores on the datasets except for the building dataset. TFT can improve alignment accuracy but also bring instability.
- SPW combines quasi-homography and content-preserving warping to align images, which adding other constraints while also reducing the accuracy of alignment, resulting in lower scores compared to REW and our method.

• Both REW and our method use global homography matrix to coarsely align the images. Afterwards, in REW and our method, a deformation field and optical flow are applied to further align the images respectively. Therefore, the both methods has higher scores and robustness than other methods.

4.1.3. Stitching Speed Comparison

The running speed is a direct reflection of the efficiency of each stitching method. Table 2 shows the speed of the SC-AOF method versus the APAP, AANAP, SPHP, TFT, REW and SPW methods. The temple, rail tracks, garden, building, school, park-square, wall, and cabinet datasets contributed by [4][10] are used in this experiment. It can be seen that the REW algorithm has the fastest stitching speed. The reason is that it only needs to calculate TPS parameters based on feature point matching and then compute the transformations of grid points quickly. Our SC-AOF method ranks second in terms of stitching speed, and the AANAP algorithm requires the longest running time. Both the APAP and AANAP methods calculate the local homographies based on moving DLT, and the AANAP method needs to calculate the Taylor expansion of anchor points. Combined with the results in Figures 8 and 9, it is verified that the SC-AOF method has the best overall performance, which can maintain a desirable operation efficiency while guaranteeing the final image quality. Our method can have broad applications and promotion significance.

Table 2. Comparison on elapsed time.

	APAP	AANAP	SPHP	TFT	REW	SPW	Ours
temple	8.8	27.6	20.5	2.8	1.1	4.5	3.5
rail-tracks	50.4	161.6	91.0	36.1	35.4	260.5	29.5
garden	57.9	148.0	72.3	41.1	20.6	64.4	32.9
building	14.3	47.0	19.3	53.4	2.8	10.4	6.6
school	8.6	37.6	3.5	6.8	4.4	9.8	10.8
wall	16.4	81.0	12.1	37.5	13.6	15.7	37.2
park-square	51.3	194.4	91.5	53.8	13.5	149.4	30.2
cabinet	6.3	23.8	4.2	1.6	0.8	3.8	3.7

4.2. Compatibility of SC-AOF Method

The SC-AOF method can not only be used independently to generate stitched image with reduced perspective deformation and low alignment error, but also be decomposed (into SC method and image blending method) and combined with other methods to improve the quality of the mosaic.

4.2.1. SC Module Compatibility Analysis

The sliding camera (SC) module in the SC-AOF method can not only be used in the global alignment model, but also be combined with other local alignment models (e.g., APAP and TFT) to solve perspective deformation while maintaining the alignment accuracy. The implementation steps are as follows.

1. Use the global similarity transformation to project I_2 onto the I_1 coordinate system to calculate the size and mesh vertices of the mosaic;

2. Use (6)-(9) to calculate the weights of mesh vertices and the projection matrix, replace the homography H in (2) with the homography matrix in local alignment model, and bring them into (12) to compute the warped images and blend them.

Figure 10 presents the stitched images when using the TFT algorithm alone vs. using the TFT algorithm combined with the SC method. It can be found that the combined method is more effective

in mitigating the edge stretching, and that it generates more natural images. This shows that the SC method can effectively solve perspective deformation suffered by the local alignment method.
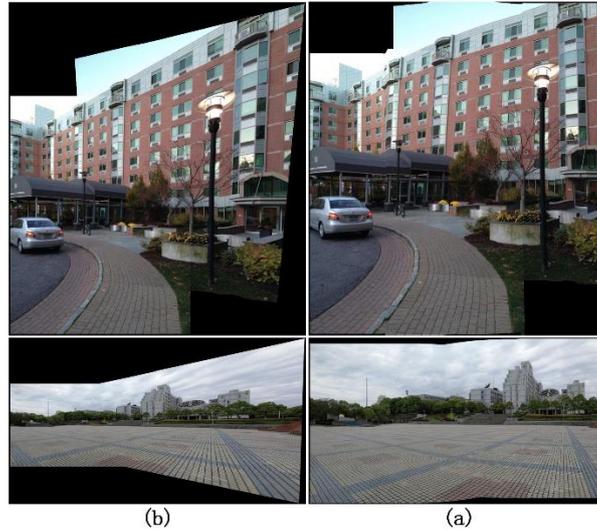

Figure 10. The combination of TFT and moving cameras method. (a) is the mosaics created using TFT, and (b) is the mosaics obtained by adding the moving camera method to TFT.

4.2.2. Blending Module Compatibility Analysis

The asymmetric optical flow-based blending in the SC-AOF method can also be used in other methods to enhance the final stitching effect. The implementation steps are as follows.

1. Generate two projected images using one of the other algorithms and calculate the blending parameters based on the overlapping areas;
2. Set the optical flow value to be 0, replace linear blending parameter λ with α in (17) to blend warped images, preserve the blending band width in the low frequency-area and narrow the blending width in the high-frequency area to get a better image stitching effect.

Figure 11 shows the image stitching effect of the APAP algorithm when using linear blending vs. when using our bending method. It can be seen that the blurring and ghosting in the stitched images are effectively mitigated when using our blending method. This shows that our blending algorithm can blend the aligned images better.
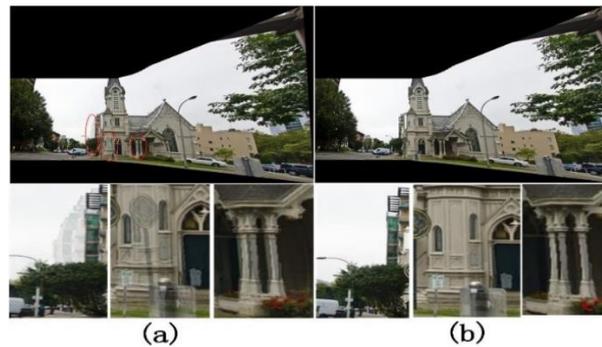
Figure 11. The combination of APAP and our blending method. (a) shows the mosaic and detail view generated by the APAP using linear blending. (b) show the results of APAP combined with our blending method.

5. Conclusions

In this paper, to solve the perspective deformation and misalignment in image stitching using homographies, a SC-AOF method is proposed. In image warping, a new virtual camera and a projection matrix are generated as the observation perspective in the overlapping area by interpolating between two projection matrices; and the overlapping area transitions gradually from one viewpoint to another to achieve viewpoint preserving and smooth transition of the stitched image and thus solve the perspective deformation problem. In image blending, optical flow-based blending algorithm is proposed to further improve alignment accuracy. The width of blending area is automatically adjusted according to softmax function and alignment accuracy. Finally, extensive comparison experiments are conducted to demonstrate the effectiveness of our algorithm in reducing perspective deformation and improving alignment accuracy. In addition, our algorithm had broad applicability, as its component modules can be used with other algorithms to mitigate edge stretching and improve alignment accuracy.

However, the proposed local alignment method would fail if the input images contain large parallax, which will cause severe occlusion to prevent us from obtaining the correct optical flow. The problem of local alignment failure caused by large parallax also exists in other local alignment methods. Exploring more robust optical flow calculation and occlusion processing method to reduce misalignment in large parallax scene is an interesting research direction of the future work.

References

1. Abbadi N K E L, Al Hassani S A, Abdulkhaleq A H. A review over panoramic image stitching techniques[C]//Journal of Physics: Conference Series. IOP Publishing, 2021, 1999(1): 012115.
2. Gómez-Reyes J K, Benítez-Rangel J P, Morales-Hernández L A, et al. Image mosaicing applied on UAVs survey[J]. Applied Sciences, 2022, 12(5): 2729.
3. Xu Q, Chen J, Luo L, et al. UAV image stitching based on mesh-guided deformation and ground constraint[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 4465-4475.
4. Wen S, Wang X, Zhang W, et al. Structure Preservation and Seam Optimization for Parallax-Tolerant Image Stitching[J]. IEEE Access, 2022, 10: 78713-78725.
5. Tang W, Jia F, Wang X. An improved adaptive triangular mesh-based image warping method[J]. Frontiers in Neurorobotics, 2023, 16: 1042429.

6. Li J, Deng B, Tang R, et al. Local-adaptive image alignment based on triangular facet approximation[J]. IEEE Transactions on Image Processing, 2019, 29: 2356-2369.

7. Lee K Y, Sim J Y. Warping residual based image stitching for large parallax[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 8198-8206.

8. Zhu S, Zhang Y, Zhang J, et al. ISGTA: an effective approach for multi-image stitching based on gradual transformation matrix[J]. Signal, Image and Video Processing, 2023, 17(7): 3811-3820.

9. Zaragoza J, Chin T J, Brown M S, et al. As-Projective-As-Possible Image Stitching with Moving DLT[C]// Computer Vision and Pattern Recognition (CVPR). IEEE, 2013.

10. Li J, Wang Z, Lai S, et al. Parallax-tolerant image stitching based on robust elastic warping[J]. IEEE Transactions on multimedia, 2017, 20(7): 1672-1687.

11. Xue F, Zheng D. Elastic Warping with Global Linear Constraints for Parallax Image Stitching[C]//2023 15th International Conference on Advanced Computational Intelligence (ICACI). IEEE, 2023: 1-6.

12. Liao T, Li N. Natural Image Stitching Using Depth Maps[J]. arXiv preprint arXiv:2202.06276, 2022.

13. Cong Y, Wang Y, Hou W, et al. Feature Correspondences Increase and Hybrid Terms Optimization Warp for Image Stitching[J]. Entropy, 2023, 25(1): 106.

14. Chang C H, Sato Y, Chuang Y Y. Shape-preserving half-projective warps for image stitching[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 3254-3261.

15. Chen J, Li Z, Peng C, et al. UAV image stitching based on optimal seam and half-projective warp[J]. Remote Sensing, 2022, 14(5): 1068.

16. Lin C C, Pankanti S U, Ramamurthy K N, et al. Adaptive as-natural-as-possible image stitching[C]//Computer Vision & Pattern Recognition.IEEE, 2015.DOI:10.1109/CVPR.2015.7298719.

17. CHEN Yusheng, CHUANG Yungyu. Natural Image Stitching with the Global Similarity Prior[C]. European Conference on Computer Vision. Springer International Publishing, 2016.

18. Cui J, Liu M, Zhang Z, et al. Robust UAV thermal infrared remote sensing images stitching via overlap-prior-based global similarity prior model[J]. IEEE journal of selected topics in applied earth observations and remote sensing, 2020, 14: 270-282.

19. Liao T, Li N. Single-perspective warps in natural image stitching[J]. IEEE transactions on image processing, 2019, 29: 724-735.

20. Li N, Xu Y, Wang C. Quasi-homography warps in image stitching[J]. IEEE Transactions on Multimedia, 2017, 20(6): 1365-1375.

21. Du P, Ning J, Cui J, et al. Geometric Structure Preserving Warp for Natural Image Stitching[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 3688-3696.

22. Bertel T, Campbell N D F, Richardt C. Megaparallax: Casual 360 panoramas with motion parallax[J]. IEEE transactions on visualization and computer graphics, 2019, 25(5): 1828-1835.

23. Meng M, Liu S. High-quality Panorama Stitching based on Asymmetric Bidirectional Optical Flow[C]//2020 5th International Conference on Computational Intelligence and Applications (ICCIA). IEEE, 2020: 118-122.

24. Hofinger M, Bulò S R, Porzi L, et al. Improving optical flow on a pyramid level[C]//European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 770-786.

25. Shah S T H, Xuezhi X. Traditional and modern strategies for optical flow: an investigation[J]. SN Applied Sciences, 2021, 3(3): 289.

26. Zhai M, Xiang X, Lv N, et al. Optical flow and scene flow estimation: A survey[J]. Pattern Recognition, 2021, 114: 107861.

27. Liu C, Yuen J, Torralba A. Sift flow: Dense correspondence across scenes and its applications[J]. IEEE transactions on pattern analysis and machine intelligence, 2010, 33(5): 978-994.

28. Zhao S, Zhao L, Zhang Z, et al. Global matching with overlapping attention for optical flow estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 17592-17601.

29. Rao S, Wang H. Robust optical flow estimation via edge preserving filtering[J]. Signal Processing: Image Communication, 2021, 96: 116309.

30. Jeong J, Lin J M, Porikli F, et al. Imposing consistency for optical flow estimation[C]//Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition. 2022: 3181-3191.

31. Anderson R, Gallup D, Barron J T, et al. Jump: virtual reality video[J]. ACM Transactions on Graphics (TOG), 2016, 35(6): 1-13.

32. Teed Z, Deng J. Raft: Recurrent all-pairs field transforms for optical flow[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. Springer International Publishing, 2020: 402-419.

33. Huang Z, Shi X, Zhang C, et al. Flowformer: A transformer architecture for optical flow[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 668-685.

34.   https://github.com/facebookarchive/Surround360
35.   Zhang Y J. Camera calibration[M]//3-D Computer Vision: Principles, Algorithms and Applications. Singapore: Springer Nature Singapore, 2023: 37-65.
36.   Zhang Y, Zhao X, Qian D. Learning-Based Framework for Camera Calibration with Distortion Correction and High Precision Feature Detection[J]. arXiv preprint arXiv:2202.00158, 2022.
37.   Fang J, Vasiljevic I, Guizilini V, et al. Self-supervised camera self-calibration from video[C]//2022 International Conference on Robotics and Automation (ICRA). IEEE, 2022: 8468-8475.