Letter

Image Stitching Based on Nonrigid Warping for Urban Scene

Lixia Deng [ID]**, Xiuxiao Yuan ***, **Cailong Deng** [ID]**, Jun Chen** [ID] **and Yang Cai**

School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China;
denglixia@whu.edu.cn (L.D.); dcl@whu.edu.cn (C.D.); jchen_cug@whu.edu.cn (J.C.); Y.Tsai@whu.edu.cn (Y.C.)
* Correspondence: yuanxx@whu.edu.cn

Abstract: Image stitching based on a global alignment model is widely used in computer vision. However, the resulting stitched image may look blurry or ghosted due to parallax. To solve this problem, we propose a parallax-tolerant image stitching method based on nonrigid warping in this paper. Given a group of putative feature correspondences between overlapping images, we first use a semiparametric function fitting, which introduces a motion coherence constraint to remove outliers. Then, the input images are warped according to a nonrigid warp model based on Gaussian radial basis functions. The nonrigid warping is a kind of elastic deformation that is flexible and smooth enough to eliminate moderate parallax errors. This leads to high-precision alignment in the overlapped region. For the nonoverlapping region, we use a rigid similarity model to reduce distortion. Through effective transition, the nonrigid warping of the overlapped region and the rigid warping of the nonoverlapping region can be used jointly. Our method can obtain more accurate local alignment while maintaining the overall shape of the image. Experimental results on several challenging data sets for urban scene show that the proposed approach is better than state-of-the-art approaches in both qualitative and quantitative indicators.

Keywords: image alignment; image stitching; nonrigid warping; parallax-tolerant; urban scene

1. Introduction

Urban scene images acquired by optical sensors have a wide range of applications in urban informatization, such as environmental monitoring, road planning, street-view map production, and 3D urban reconstruction [1–4]. Due to the limitations of the camera's viewing angle and shooting distance, the area covered by a single image is small. Therefore, it is necessary to use image stitching technology to expand the coverage of the image and obtain more information from the target area.

Image stitching is a process of merging a group of images into a larger image with a wider field-of-view of the scene. It can usually be solved by aligning images based on their common features. The sparsely scattered, distinctive, and well-localized key points provided by the sparse feature matchers have been widely used in image correlation. Although the sparse feature matchers lack the corresponding density provided by the dense matching method [5], it can provide the advantages of wide baseline, fast speed, and unlimited data types [6]. Most of the image alignment algorithms aim to find a two-dimensional global warp model between two overlapped images, such as similarity, affine, and homography. A fine global warp minimizes the total registration error instead of exactly aligning all the features; therefore, it is robust but not sufficient to adapt to all scenes. In computer vision, homography model is widely used to describe the projection relationship within an image pair. However, it only works when the scene is planar or the camera undergoes pure rotation [7]. When the parallax is too large to be ignored, unexpected "ghosts" will appear in the stitched images aligned

by a single global warp model. For street view images in urban scenes that contain rich objects and complex depth changes, it is necessary to consider image stitching methods that can handle parallax.

In order to improve alignment accuracy in the presence of parallax, scholars have conducted extensive work in the field of computer vision. In summary, the existing methods are based on the following three ideas. One of the ideas is to search for seam lines to bypass the misalignment in the overlapped region [8–11]. Seam-based methods usually have a high computational cost and are more suitable for the situation in which there are obvious foreground objects and background in the images. Another idea is to adopt multiple transformation models [12–14]. The third idea is to use surface fitting to deal with the parallax on the two-dimensional image [15–17]. Actually, it is believed for the last two ideas that different regions of the image should utilize different warp models because, in the image obtained by pinhole imaging, objects closer to the shooting center will have greater parallax. Therefore, they all tried to find an image alignment model that changes with space.

In addition, we noticed that the quality of feature matching directly affects the stitching quality. Most image alignment approaches employed the Random Sample Consensus (RANSAC) algorithm [18] to remove outliers of the matched features, and a global transformation usually serves as the minimal solver of RANSAC. There are contradictions between global RANSAC and spatially varying alignment: (1) If the threshold is too small, inliers might be rejected because they do not conform to the global transformation used by the RANSAC method, which is not conducive to local alignment; (2) If the threshold is too large, outliers might be preserved and lead to a poor stitching result. Good feature correspondences can refine the alignment, and a good alignment can verify the existing correspondences. Therefore, we need a more flexible feature match refinement method which can preserve spatially varying projection of the feature points.

In this paper, we propose a parallax-tolerant image stitching method that follows the idea of spatially varying alignment. Our goal is to find a good feature correspondence and, at the same time, determine a fine warp model to reduce registration errors. First, we establish a new feature mapping relationship base on semiparametric fitting. The semiparametric function includes a smoothness constraint based on the Motion Coherence Theory (MCT) [19], which provides greater flexibility for finding good feature maps through rough feature correspondence. Features that do not conform to the fitting functions are regarded as outliers. This idea is inspired by Lin et al. [20,21]; they used a complex smooth function to fit the feature correspondence and deal with the piecewise noises that RANSAC cannot handle. We design a smoothing function that is more suitable for two-dimensional plane warping and stitching. The advantage is that the alignment model can be directly derived from the feature correspondence. Then, we obtain a nonrigid warping based on the Gaussian radial basis function (GRBF) to eliminate misalignments in the overlapped region. Compared with TPS, GRBF is more suitable for local deformation [22]. Homography is a warping from one two-dimensional plane to another two-dimensional plane, while the nonrigid warp is more like performing a three-dimensional surface fitting first and then projecting onto a two-dimensional plane. Therefore, we can eliminate parallax errors which may be left by a single homography warping. Finally, we gradually change the nonrigid warp to the global homography warp to reduce unnecessary distortion in the nonoverlapping region. Meanwhile, a grid-based interpolation calculation method is used to improve efficiency. Experiments on several challenging image sets prove that our method can effectively reduce the projection errors, and can be well combined with other global methods.

2. Related Works

2.1. Feature Matching

In feature-based image stitching, feature matching is an important foundation. Lowe and David [23] proposed a sparse feature descriptor known as scale-invariant feature transform (SIFT), which is invariant to image translation, rotation, and scale. It is also robust to addition of noise, affine distortion, and changes in illumination [24]. Numerous studies have shown that the SIFT

is the most widely used feature descriptor in image stitching and the performance has been demonstrated [25]. After preliminary matching based on feature descriptors, RANSAC was usually used to eliminate outliers based on the geometric relationship between images.

Among the researches we have investigated, Zhang et al. [26] introduced an outlier rejection method based on local homography to remove incorrect feature matchings; this method can only be used under the framework of the APAP method and is not universal. Guo et al. [27] assumed that the scene contains two planes. First, they found matches on one of the planes using the RANSAC method; then, they found matches on the other plane from the remaining points at the adjacent frame. Similar to DHW, this method is only suitable for specific scenes. Chen et al. introduced a nonrigid matching algorithm based on vector field consensus (VFC) [28] to the mosaic system for generating accurate feature matching [29]. In most methods other than global warping, the model fitted by RANSAC and the model used for alignment are relatively independent. Therefore, the retained feature points cannot help to optimize the stitching field.

2.2. Parallax-Tolerant Image Stitching

Gao et al. [8] estimated multiple warps from multiple sets of features, then used the quality of the seam line to evaluate the alignment performance of different warps and selected the best one. Zhang et al. [9] estimated reasonable seam lines by considering geometric alignment and image content, and optimized local alignment with reference to the Content Preserving Deformation (CPW) method [30]. K. Lin et al. [10] followed the mosaic line guidance, introduced contour detection and straight line detection, and used curve and straight line structures to maintain constraints in the deformation. The line matching method has obvious advantages, but the high computational complexity limits its application range. Herrmann et al. [31] made full use of object detection [32] and combined the multiple registration algorithm [11] to construct an object-centric image mosaic framework. Multiple potential planes generated by multiple registration can effectively solve the occlusion of foreground objects on the background, but it also makes the search of seam lines more complicated. In general, seam-based methods usually require high computational cost because they involve foreground and background recognition, multiple registration, seam evaluation and search, and manual interaction. They are more suitable for the situation in which there are obvious foreground objects and background in the images.

Since a global homography transformation will cause inaccurate image registration, further misalignment, and ghosting, Gao et al. proposed a dual-homography warping (DHW) model [12], which divided the scene into a background plane and a foreground plane, and used two homography matrices to align them. Since their premise was that the scene consists of two main planes, it performed well in certain specific scenes, but it could not handle more complex scenes. Lin et al. [13] proposed a smoothly varying affine (SVA) model to deal with parallax. Zaragoza et al. [14] extended this idea to a smoothly varying homography model and proposed the as-projective-as-possible (APAP) warp. The image was divided into grids, and then they used moving-DLT to calculate local adaptive homography for each grid. APAP achieves more accurate alignment in overlapping areas than DHW and has better extrapolation quality in nonoverlapping areas. Zhang et al. [26] proposed a multiviewpoint panoramic stitching method based on APAP, a local homography verification method was used to roughly align the images, and various prior constraints were used to improve the alignment through an iterative optimization scheme. Based on APAP, many works use similar methods to design more prior constraints; the combination of different constraints leads to local optimization instead of global optimization, and the computational cost is higher. Li et al. [16] proposed the elastic local alignment (ELA) and aligned images based on the Thin Plate Spline (TPS) model. The TPS function simulates the distortion of a plane based on the principle of minimum surface bending energy. It is a commonly used deformation function in biology and medical images. This function has a global nature—that is, all anchor points will have an impact on the desired point. Chen et al. [17] proposed a drone image stitching method that uses compactly supported radial basis function (CSRBF) instead of

TPS to reduce local registration errors. This inspired us to think about the application of different RBFs in image stitching.

The spatially varying warping models can handle moderate parallax and provide satisfactory stitching performance, but it usually causes projective distortion outside the overlapped region. Therefore, many scholars have made further improvements. C.H. Chang et al. proposed the shape-preserving half-projective (SPHP) warps [33] from the perspective of shape correction. They made the warp gradually change from local warp to global similarity, and added similarity constraints to the entire image. Lin and Pankanti proposed an adaptive as-natural-as-possible (AANAP) warp [34], which combines a linear homography warp and a global similarity warp with minimum rotation angle, thereby creating a natural-looking mosaic. Nan Li et al. proposed a quasi-homography warp [35], which squeezes the mesh of the corresponding homography warp, but does not change its shape. These methods tried to achieve a certain balance between projection distortion and perspective distortion in nonoverlapping region, so that images can be stitched with better visual effects.

Video stitching also involves the processing of parallax, especially for videos captured by mobile cameras (e.g., smartphones or UAVs). Many researches generally perform image stitching and stabilization simultaneously to solve the ghosting and blur in the stitched video. Guo et al. [27] proposed a video stitching method acquired by two mobile handheld cameras. The intertransformation between different cameras was estimated to obtain the spatial alignment, and the intratransformation within each video was estimated to maintain the temporal smoothness. They use APAP warping method for spatial alignment to deal with parallax. Nie et al. [36] introduced a background recognition method. The backgrounds of input videos were first identified, and a seam-based strategy was used to obtain the final stitched video.

3. The Proposed Method

The overall workflow of our proposed image stitching is illustrated in Figure 1. First, we use the SIFT method to obtain matched feature points, and then use semiparametric fitting with motion coherence constraints to eliminate mismatches. Next, a nonrigid warp model is used to align the overlapping region. In order to maintain the shape of the nonoverlapping region, the nonrigid warping is gradually transformed into a global warping. Finally, a simple linear fusion method is used to blend the stitched images. In this section, we will give a detailed description of the feature match refinement, the nonrigid warping, and its combination with other global models.
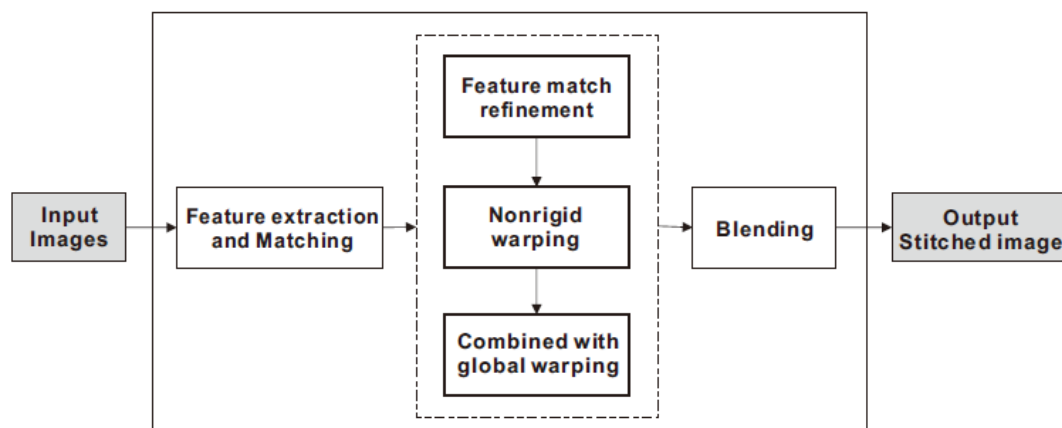

Figure 1. The overview flowchart of our proposed image stitching approach.

3.1. Feature Match Refinement Based on Semiparametric Function Fitting

Establishing a warping model between images is the basis of stitching. When we use the method based on sparse feature points, the problem becomes to establish feature mapping relationship based on feature correspondences. Given a set of N putative matched features $S = \{(p_i, q_i)\}_{i=1}^{N}$ from image I_p and I_q, $p_i = (x_i, y_i)$ and $q_i = (u_i, v_i)$ are two-dimensional vectors that denote the image coordinates of feature points. Our goal is to fit an appropriate function to map the coordinates from the first image to the second image, and the mapping f from R^2 to R^2 can be constructed as two mappings from R^2 to R separately—that is, $f = (f_x, f_y)$, under the constraints $f_x(p_i) = u_i$ and $f_y(p_i) = v_i$ for $i = 1, \ldots, N$. Since parametric functions such as rigid affine or homography cannot reflect the local spatial changes of feature points, we naturally think of a smoother semiparametric function. In this article, we use a semiparametric function composed of parametric affine and nonparametric terms. Taking the x dimension, for example, the mapping f_x with the input domain $p = (x, y)$ can be expressed as

$$f_x(p) = \alpha_1 x + \alpha_2 y + \alpha_3 + \phi_x(p) \tag{1}$$

where $\phi_x(p)$ is a smooth function with motion coherence constraint [19,37] as follows:

$$\Psi_x = \int_{\mathbb{R}^2} \frac{|\bar{\phi}_x(\omega)|^2}{\bar{g}(\omega)} d\omega \tag{2}$$

$\bar{\phi}_x(.)$ denotes the Fourier transform of function $\phi_x(.)$, while $\bar{g}(.)$ is the Fourier transform of a Gaussian function $g(r, \sigma) = e^{-|r|^2/\sigma^2}$ with spatial distribution σ, and $|\cdot|$ denotes the Euclidean distance calculation.

In order to find the smoothest mapping, we appropriately relax the registration constraint and introduce the motion coherence constraint Ψ_x as a regular term. This is expressed as the energy function

$$E_x = \sum_{i=1}^{N} |u_i - (\alpha_1 x_i + \alpha_2 y_i + \alpha_3 + \phi_x(p_i))|^2 + \lambda \Psi_x \tag{3}$$

where λ is a constant represents the weight given to the regularization term. It is difficult to directly minimize E_x due to the existence of continuous functions $\phi_x(p)$ and Ψ_x. Fortunately, Andriy Myronenko et al. [37] and Wenyan Lin et al. [38] have deduced the discrete forms:

$$\phi_x(p) = \sum_{j=1}^{N} w_x(j) g(p - p_j, \sigma) = \sum_{j=1}^{N} w_x(j) e^{-|p-p_j|^2/\sigma^2} \tag{4}$$

where $g(p - p_j, \sigma) = e^{-|p-p_j|^2/\sigma^2}$ is the Gaussian radial basis function and $\{w_x(j)\}_{j=1}^{N}$ are unknown variables;

$$\Psi_x = w_x^T G w_x \tag{5}$$

where $G_{N \times N}$ is a square symmetric matrix with elements $G(i, j) = g(p_i - p_j, \sigma)$; it can also be called the Gaussian radial basis kernel. $w_x = [w_x(1), \ldots, w_x(N)]^T$ is a $N \times 1$ vector, used as the weights of the radial basis functions.

Substituting Equations (4) and (5) into Equation (3) yields

$$\arg \min_{\{f_x(p)\}} \sum_{i=1}^{N} |u_i - (\alpha_1 x_i + \alpha_2 y_i + \alpha_3 + \phi_x(p_i))|^2 + \lambda \Psi_x$$

$$= \arg \min_{\left\{\alpha_1, \alpha_2, \alpha_3, \{w_x(j)\}_{j=1}^{N}\right\}} \sum_{i=1}^{N} \left| u_i - \left(\alpha_1 x_i + \alpha_2 y_i + \alpha_3 + \sum_{j=1}^{N} w_x(j) g(p_i - p_j, \sigma) \right) \right|^2 + \lambda w_x^T G w_x \tag{6}$$

where the energy is dependent on $N + 3$ variables $\alpha_1, \alpha_2, \alpha_3$, and $\{w_x(j)\}_{j=1}^N$. Minimizing the overall energy function in Equation (6) leads to the parametrized f_x as follows:

$$f_x(\boldsymbol{p}) = \alpha_1 x + \alpha_2 y + \alpha_3 + \sum_{j=1}^N w_x(j) g\left(\boldsymbol{p} - \boldsymbol{p}_j, \sigma\right) \tag{7}$$

The mapping from x_i, y_i to v_i has the similar form as

$$f_y(\boldsymbol{p}) = \beta_1 x + \beta_2 y + \beta_3 + \sum_{j=1}^N w_y(j) g\left(\boldsymbol{p} - \boldsymbol{p}_j, \sigma\right) \tag{8}$$

and the energy function is

$$E_y = \sum_{i=1}^N \left| v_i - \left(\beta_1 x_i + \beta_2 y_i + \beta_3 + \sum_{j=1}^N w_y(j) g\left(\boldsymbol{p} - \boldsymbol{p}_j, \sigma\right)\right) \right|^2 + \lambda \boldsymbol{w}_y^T \boldsymbol{G} \boldsymbol{w}_y \tag{9}$$

Since $\boldsymbol{G}$ is a positive definite matrix, the overall energy minimization problem in Equations (6) and (9) can be solved using the following linear system [39]

$$\begin{bmatrix} \boldsymbol{G} + \lambda \boldsymbol{I} & \boldsymbol{P} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_x & \boldsymbol{w}_y \\ \boldsymbol{a} & \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{u} & \boldsymbol{v} \end{bmatrix} \tag{10}$$

where $\boldsymbol{P}$ is a $N \times 3$ matrix with the ith row $[x_i, y_i, 1]$, $\boldsymbol{u} = [u_1, \ldots, u_N]^T$ and $\boldsymbol{v} = [v_1, \ldots, v_N]^T$ represent the coordinates of target points in I_q. $\boldsymbol{w}_x = [w_x(1), \ldots, w_x(N)]^T$ and $\boldsymbol{a} = [\alpha_1, \alpha_2, \alpha_3]^T$, and $\boldsymbol{w}_y = [w_y(1), \ldots, w_y(N)]^T$ and $\boldsymbol{b} = [\beta_1, \beta_2, \beta_3]^T$ represent two sets of pending variables.

A brief proof is as follows: Consider the energy function E_x in Equation (6). The matrix form of its derivative with respect to $\boldsymbol{w}_x$ should be zero.

$$\frac{\partial E_x}{\partial \boldsymbol{w}_x} = 2\boldsymbol{G}\left((\boldsymbol{G}\boldsymbol{w}_x + \boldsymbol{P}\boldsymbol{a}) - \boldsymbol{u}\right) + 2\lambda \boldsymbol{G}\boldsymbol{w}_x = 0 \tag{11}$$

Multiplying Equation (11) by $1/2\boldsymbol{G}^{-1}$, we obtain

$$(\boldsymbol{G} + \lambda \boldsymbol{I})\,\boldsymbol{w}_x + \boldsymbol{P}\boldsymbol{a} = \boldsymbol{u} \tag{12}$$

Through the above analysis, we successfully treat image warping as a general matching problem with smoothness constraint. Each feature point has its own associated mapping parameters, rather than all points sharing the same set of parameters. f_x and f_y can be regarded as a pair of smooth surface fitting functions. We transform the smooth function into the sum of a finite number of radial basis functions, so that the problem of minimizing a convex cost function is transformed into solving a linear system. After each solution, we use the median absolute deviation (MAD) method [40] to remove outliers—that is, the points whose deviation from the fitted function is lager than 1.48 times MAD will be regarded as outliers. The mapping parameters are recalculated using inliers, and this process repeats three times. This is a more flexible feature match refinement method that can reserve more points from a set of initial correspondences based on feature descriptors.

3.2. Nonrigid Warping Based on Gaussian Radial Basis Functions

As mentioned before, the smooth function $\phi(.)$ has a discrete form in Equation (4). It is a linear sum of Gaussian radial basis functions (GRBFs). They are constructed based on the Euclidean distances between the target point to the control points. Taking I_q as the reference image, we transform the coordinates (x, y) of an arbitrary point in image I_p into the coordinate system of the reference image

to become (x', y'). The feature points $\{p_i = (x_i, y_i)\}_{i=1}^{M}, M \leq N$ in image I_p are used as the control points, and their correspondences in image I_q are $\{q_i = (u_i, v_i)\}_{i=1}^{M}$. Our nonrigid warp model has the same form as the feature mapping function, which is a polynomial plus the linear sum of GRBFs in each dimension. The formula is as follows:

$$\begin{cases} x' = f_x(x,y) = \alpha_1 x + \alpha_2 y + \alpha_3 + \sum\limits_{i=1}^{M} w_x(i) e^{-|(x,y) - p_i|^2 / \sigma^2} \\ y' = f_y(x,y) = \beta_1 x + \beta_2 y + \beta_3 + \sum\limits_{i=1}^{M} v_x(i) e^{-|(x,y) - p_i|^2 / \sigma^2} \end{cases} \tag{13}$$

where $\{w_x(1), \ldots, w_x(M), \alpha_1, \alpha_2, \alpha_3\}$, and $\{w_y(1), \ldots, w_y(M), \beta_1, \beta_2, \beta_3\}$ are variables calculated based a set of M features correspondences after match refinement. The linear system is also used in Equation (10).

In order to better apply nonrigid warp to image stitching, it is necessary to overcome the large amount of calculation, because, for our nonrigid warp model, each pixel on the target image has its own deformation parameters, and its Euclidean distances to all control points need to be calculated. Therefore, the pixel-by-pixel calculation will be time-consuming. We use mesh deformation to speed up the calculation: before resampling, we divide the image into a grid mesh of $C_1 \times C_2$ cells, calculate the deformation on the grid nodes first, and then obtain other points' coordinates through linear interpolation. A visualization example in Figure 2 shows how the image is warped based on our nonrigid warp model and mesh deformation. The parallax between pixels is regarded as the "height" above the image plane in the imaginary third dimension. Then, the smooth surface fitted by the nonrigid model is reprojected to the reference image plane. Therefore, our method is suitable for image stitching with smooth varying parallax.
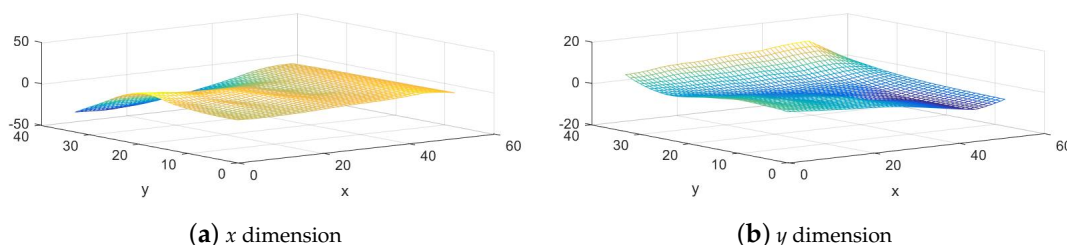

(**a**) *x* dimension   (**b**) *y* dimension

Figure 2. Nonrigid warp using mesh deformation.

3.3. Smooth Transition to Global Warping

In addition to warping, image stitching also involves extrapolating the warp model calculated based on the overlapping region to the nonoverlapping region. Because of the strong intervention of the matching points, the nonrigid warp leads to better alignment in the overlapping region. However, if it is forced to extrapolate to the nonoverlapping region, this part of the image will be excessively distorted. Therefore, we choose a common global warp (such as similarity and homography warps) to maintain the shape of the image in nonoverlapping regions.

According to the given feature correspondences, the least squares method is usually used to minimize the projection error of all feature points to solve a global warp model for image stitching, and it is easy to implement using the Levenberg–Marquardt (LM) algorithm [41].

Similarity warp describes the rotation, translation, and scaling of the image, and its matrix form is as follows:

$$S = \begin{bmatrix} s\cos(\theta) & -s\sin(\theta) & t_x \\ s\sin(\theta) & s\cos(\theta) & t_y \\ 0 & 0 & 1 \end{bmatrix} \tag{14}$$

The corresponding coordinate transformation is:

$$\begin{cases} x' = S_x(x,y) = s\cos(\theta)x - s\sin(\theta)y + t_x \\ y' = S_y(x,y) = s\sin(\theta)x + s\cos(\theta)y + t_y \end{cases} \tag{15}$$

The homography warp, also known as the perspective transformation, has the matrix form as follows:

$$\boldsymbol{H} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & 1 \end{bmatrix} \tag{16}$$

The corresponding coordinate transformation is:

$$\begin{cases} x' = H_x(x,y) = \frac{h_{00}x + h_{01}y + h_{02}}{h_{20}x + h_{21}y + 1} \\ y' = H_y(x,y) = \frac{h_{10}x + h_{11}y + h_{12}}{h_{20}x + h_{21}y + 1} \end{cases} \tag{17}$$

A simple method is to directly set the nonrigid warp of the nonoverlapping region to zero, but it will cause a sudden change in the overlapping edge. In this paper, the nonrigid warp is gradually reduced to achieve a smooth transition. As the point $p(x,y)$ gradually moves away from the overlapping area, the scale parameter ε gradually changes from 1 to 0 (taking nonrigid warp + similarity warp, for example):

$$\begin{cases} x' = \varepsilon f_x(x,y) + (1-\varepsilon)S_x \\ y' = \varepsilon f_y(x,y) + (1-\varepsilon)S_y \end{cases} \tag{18}$$

ε is calculated as:

$$\varepsilon = \begin{cases} 1 & , W_s \leq 0 \\ 1 - W_s/W_b & , 0 < W_s \leq W_b \\ 0 & , W_b < W_s \end{cases} \tag{19}$$

where $W_s = \max(x - x_b, x_a - x, y - y_b, y_a - y)$, $W_b = \eta * \min(x_b - x_a, y_b - y_a)$. $[x_a, x_b]$, $[y_a, y_b]$ are the coordinate ranges in x and y directions of the overlapped region calculated by the global warp, and η is a constant used to control the width of the transition area.

A comparison of using different global warps is illustrated in Figure 3, using image pair "building" from [34]. The main difference is in nonoverlapping region. Homography warp preserves all straight lines, but the region of an object is enlarged or stretched compared to its appearance in the original image. Similarity warp preserves the original shape of the object, since they purely involve translation, rotation, and uniformly scaling, but the perspectives of an object in two images may be inconsistent with each other [35]. In Figure 3a, the streetlight and door are stretched. In Figure 3b, the streetlight maintains its original shape, but the shape of the top of the building changes from straight to slightly curved. When stitching wide-field-of-view images, homography may cause large projective distortion in the nonoverlapping region, which is inconsistent with human cognitive habits. So we prefer to use a similarity warp to maintain the shape of nonoverlapping region. This is consistent with the conclusion of the SPHP [33] method, which also used similarity transformation in the nonoverlapping region.
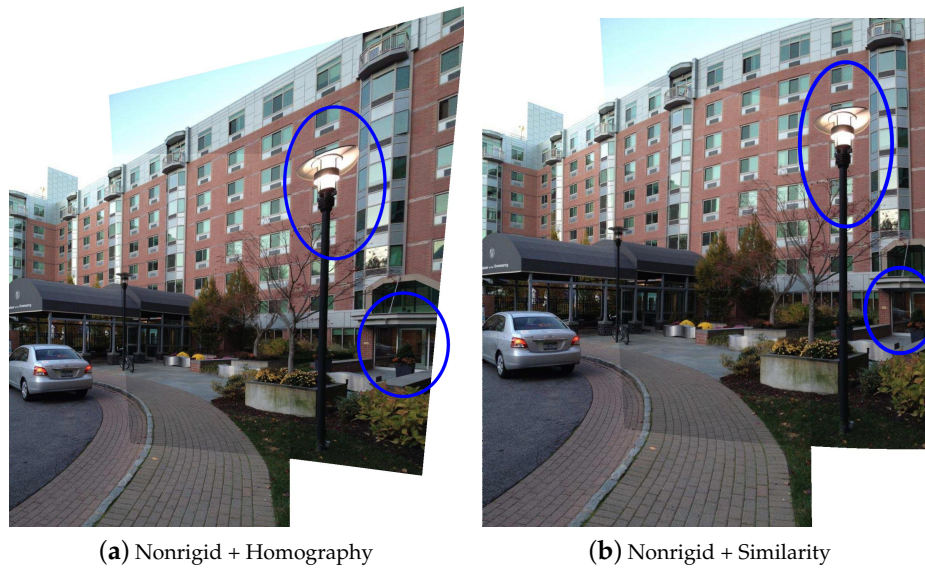
(**a**) Nonrigid + Homography   (**b**) Nonrigid + Similarity

Figure 3. Combination with different global models. The visual effect of homography warp is not as good as similarity warp, because the streetlight and door in nonoverlapping region are severely deformed.

4. Experiments and Discussion

We compare our nonrigid warp against the global homography warp and other two spatially varying warps for image stitching—that is, as-projective-as-possible (APAP) warp [14] and elastic local alignment (ELA) [16]. The SIFT [23] method is used to provide initial feature correspondences. In order to cogently evaluate these methods, we simply blend the aligned images by pixel intensity averaging so that any misalignment remain obvious. The image data sets includes several urban scene images from other related works.

The experiments are performed on a laptop with Intel i7 CPU@2.70GHz and Matlab Code.

4.1. Parameter Settings

The semiparametric function and the nonrigid warp model involve two free parameters λ and σ. λ represents the trade-off between the feature registration and the smoothness constrain, and σ reflects the strength of the interaction between the feature points. A larger σ will lead to a flatter warping, and the same is to λ. In order to take into account both the image size and the distribution of feature points, we set $\sigma = 100 \times (ho + wo)/ptnum$, where ho and wo represent the height and width of the overlapped region and $ptnum$ represents the number of feature points. λ is set to be $\pi/3$ correspondingly.

The constant η used in smooth transition is an empirical value. It acts as a scale parameter of the width of overlapping region wo. The larger the η, the larger the width of the transition area. Let the width of a single image be wi. After experimenting with multiple sets of images, we suggest setting η by making $W_b = 1.5 \times (wi - wo)$. Figure 4 shows a set of stitched results of image pair "roundabout" [34], where W_b are 0.25, 0.5, 0.75, 1.0, 1.25, and 1.5 times of $(wi - wo)$, respectively, and the corresponding η are 0.0398, 0.0796, 0.1195, 0.1593, 0.1991, and 0.2389.

In the mesh deformation stage, the larger the grid cell, the shorter the calculation time, but the precision of model fitting will also decrease. Therefore, it is necessary to find a suitable grid size to balance efficiency and stitching quality. We selected five pairs of images to count the deformation time corresponding to different grid sizes. The results are shown in Figure 5, where the result of 1×1 pixel is the time it takes for pixel-by-pixel deformation. It can be seen from the chart that when the grid size is 5×5 pixels, the warping time has been greatly shortened, but the acceleration effect of larger grid

sizes such as 20×20 pixels and 25×25 pixels is no longer obvious. In our experiments, the size of the grid cell is set to 10×10 pixels.
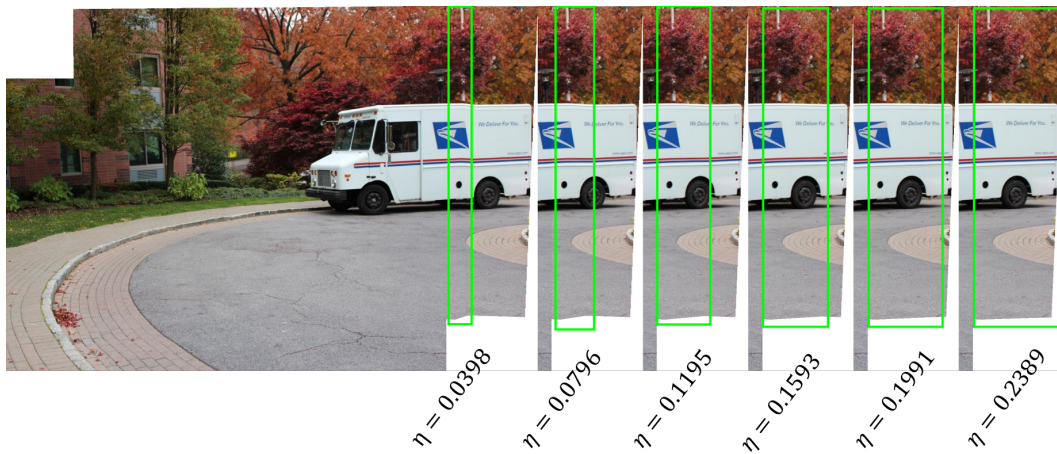

Figure 4. Different transition effects with different η. The green boxes represent the approximate transition region. A small η will cause an unnatural transition, so we try to make the transition area cover the nonoverlapping region on the right side.

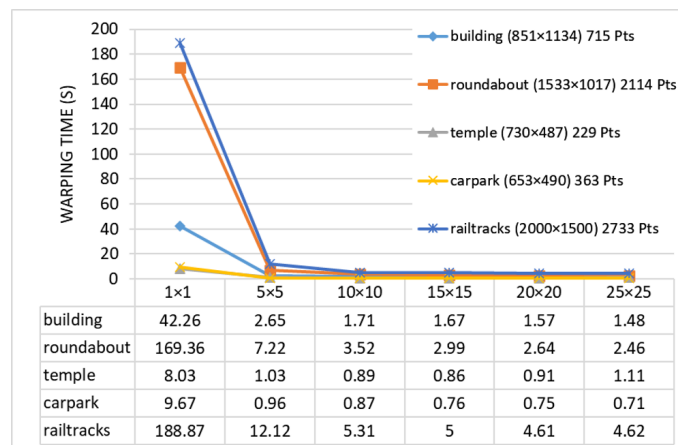|	1×1	5×5	10×10	15×15	20×20	25×25
building	42.26	2.65	1.71	1.67	1.57	1.48
roundabout	169.36	7.22	3.52	2.99	2.64	2.46
temple	8.03	1.03	0.89	0.86	0.91	1.11
carpark	9.67	0.96	0.87	0.76	0.75	0.71
railtracks	188.87	12.12	5.31	5	4.61	4.62

Figure 5. Warping time statistics for different grid sizes. The selected images have different image size and different number of feature points. For each grid size, we count the warping time 10 times and take the average time.

4.2. Qualitative Comparisons

First of all, our feature match refinement is compared with the traditional RANSAC method. Homography is selected as the global model to be fitted for RANSAC. The maximum number of iterations in the experiments is set to 500; this determines the computational efficiency of RANSAC algorithm and has been proved to be a reliable empirical value in the APAP [14] and ELA [16] methods. Another threshold minDistance is used to determine the feature points that are fit well by global model, which directly determines the number and distribution of matched features. With a smaller minDistance, some inliers may be eliminated because they do not conform to the fitted global model, as shown in Figure 6a. With a larger minDistance, some outliers will be retained, which are not conducive to our nonrigid warping, as shown in Figure 6c,d. Figure 6e,f show that our semiparametric fitting method can retain more matched features, thereby producing better stitching results.
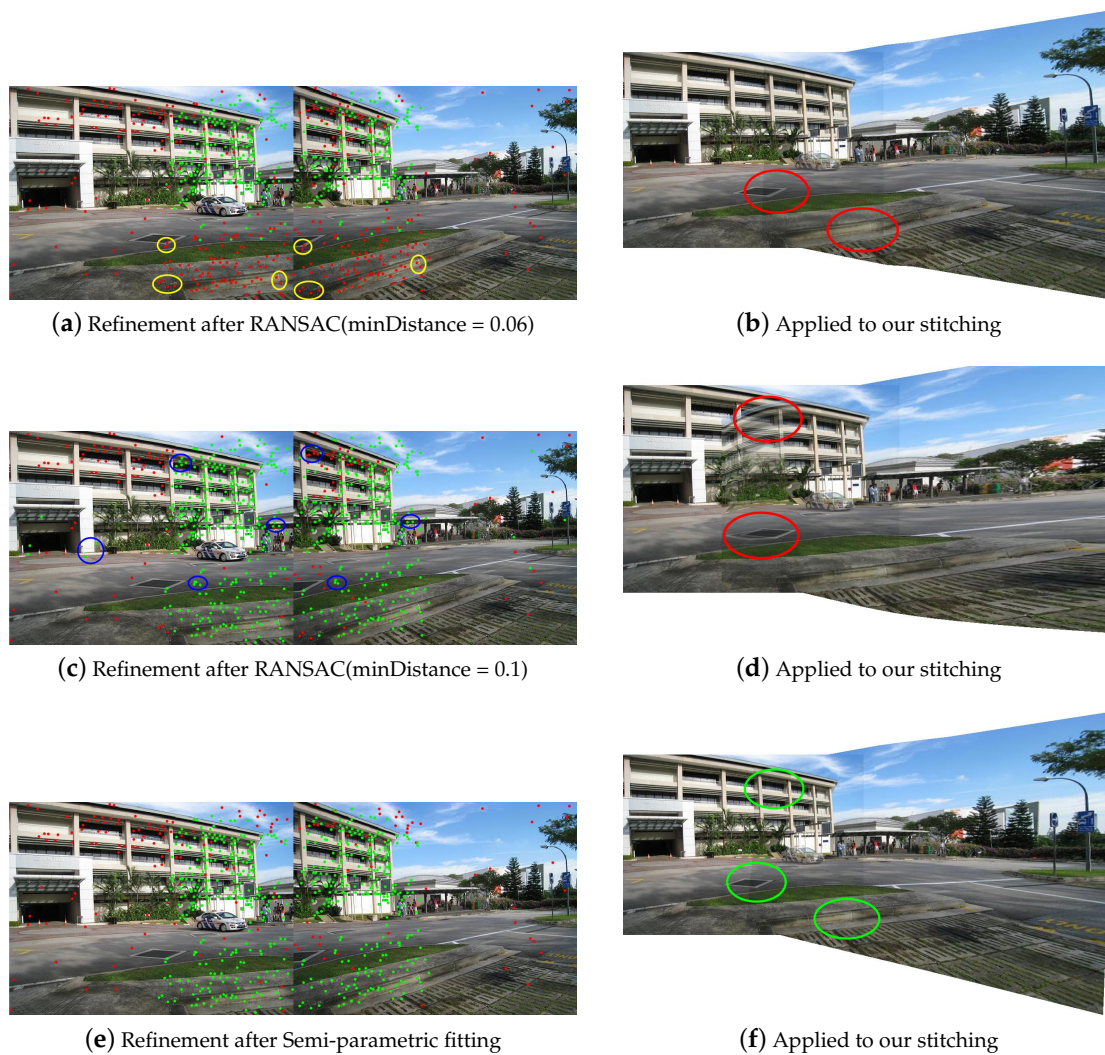
(**a**) Refinement after RANSAC(minDistance = 0.06)

(**b**) Applied to our stitching

(**c**) Refinement after RANSAC(minDistance = 0.1)

(**d**) Applied to our stitching

(**e**) Refinement after Semi-parametric fitting

(**f**) Applied to our stitching

Figure 6. Feature match refinement. (**a**,**c**) are the outlier removal results of global RANSAC. The green dots and the red dots indicate the retained and removed features, respectively. Some of the removed inliers are marked in yellow, while the remaining outliers are marked in blue. (**b**,**d**) are the stitching results of our nonrigid warp based on the global RANSAC results. (**e**) is the feature refinement result of our semiparametric fitting. (**f**) is the stitching result of our method.

Then, we selected three pairs of urban scene images to visually demonstrate the stitching quality of various methods. They are "temple" from AANAP [34], "carpark" from DHW [12], and "railtrack" from APAP [14]. Figures 7–9 show the results of different methods, three representative regions of each resulting image are highlighted. The first row shows the results of the global homography, which serves as the baseline for comparison, with obvious misalignment in all highlighted areas. The second row shows the results of APAP, the third row shows the results of ELA, and the fourth row shows the result of our nonrigid warp combined with a global similarity warp.

Figure 7 is a case of image pairs with low overlap. In the result of APAP, the nonoverlapping region of the image is severely stretched, and ghost still exists in the regions marked in green and blue. In the result of ELA, the ground suffered some distortion around the overlapping borders. In all marked regions, our nonrigid warp achieves a more accurate local alignment. Compared to the ELA method, our transition to nonoverlapping region is smoother. In Figure 8, the scene can be clearly divided into the background and the ground. In the green marked regions, APAP aligns the manhole cover on the ground, but fails to align the steps in the background, and ELA's results are exactly the opposite. Our nonrigid warp is more flexible to align the background and ground

simultaneously. Figure 9 shows a challenging data set used in APAP which is rich of complex curved features. Our nonrigid warp also has a good performance and can successfully align railways and rod-shaped objects. The marked regions show better alignments than APAP and ELA methods.

In general, our nonrigid warp performs better in the selected data sets. The nonrigid warping leads to more accurate local alignment. The combination with similarity warps leads to better visual effects in non-overlapping region.

(**a**) Baseline

(**b**) Baseline details

(**c**) APAP

(**d**) APAP details

(**e**) ELA

(**f**) ELA details

(**g**) NR + S

(**h**) NR + S details

Figure 7. Comparison of stitching quality on "temple". (**a**,**b**) are results of the baseline with obvious local misalignments. (**c**,**d**) are results of as-projective-as-possible (APAP). (**e**,**f**) are resluts of elastic local alignment (ELA). (**g**,**h**) are resluts of nonrigid + similarity warping.
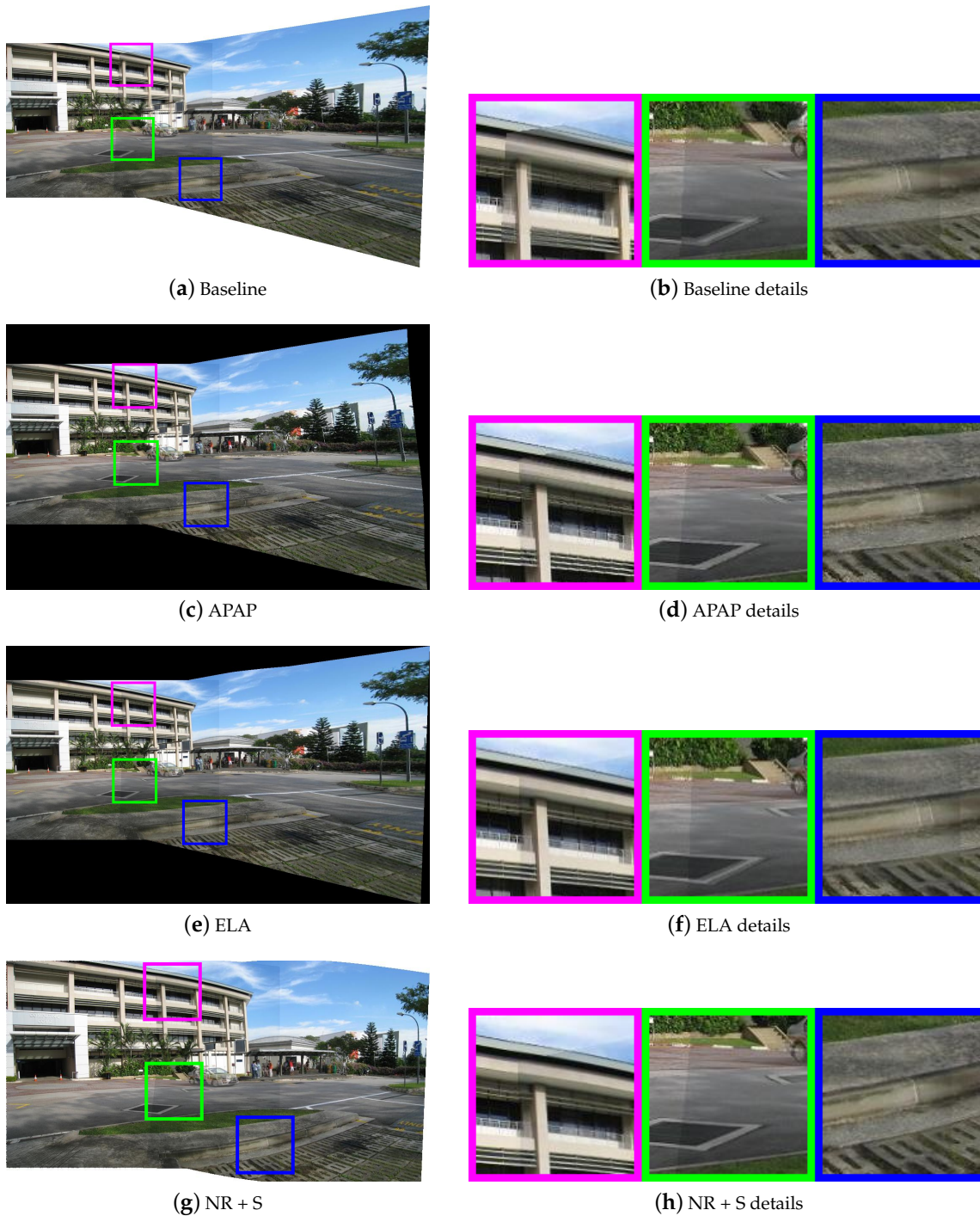
(**a**) Baseline

(**b**) Baseline details

(**c**) APAP

(**d**) APAP details

(**e**) ELA

(**f**) ELA details
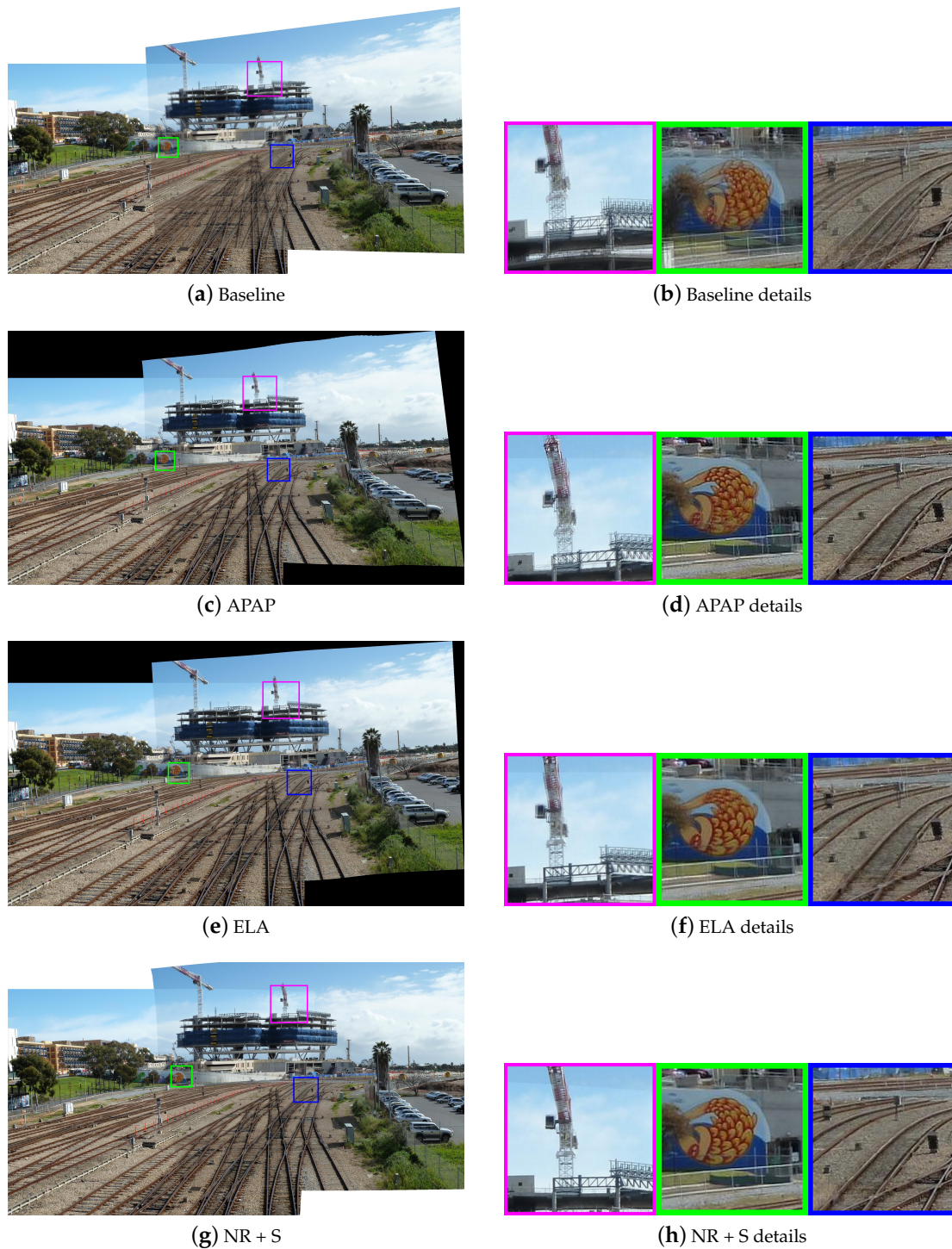
(**g**) NR + S

(**h**) NR + S details

Figure 8. Comparison of stitching quality on "carpark". (**a**,**b**) are results of the baseline with obvious local misalignments. (**c**,**d**) are results of APAP. (**e**,**f**) are results of ELA. (**g**,**h**) are results of nonrigid + similarity warping.

(**a**) Baseline

(**b**) Baseline details

(**c**) APAP

(**d**) APAP details

(**e**) ELA

(**f**) ELA details

(**g**) NR + S

(**h**) NR + S details

Figure 9. Comparison of stitching quality on "railtrack". (**a**,**b**) are results of the baseline with obvious local misalignments. (**c**,**d**) are results of APAP. (**e**,**f**) are results of ELA. (**g**,**h**) are results of nonrigid + similarity warping.

4.3. Quantitative Comparisons

To quantify the alignment accuracy of our nonrigid warp $f = \{f_x, f_y\} : \mathbb{R}^2 \mapsto \mathbb{R}^2$, we compute the root mean squared error (RMSE) of f on a set of corresponding feature points $\{(x,y)_i, (x',y')_i\}_{i=1}^{M}$, where $\text{RMSE}(f) = \sqrt{\frac{1}{M} \sum_{i=1}^{M} \left(\left| f_x(x_i,y_i) - x_i' \right|^2 + \left| f_y(x_i,y_i) - y_i' \right|^2 \right)}$.

In addition to the matched feature points used for the calculation of the warp model, we also manually selected 20–30 groups of uniformly distributed checkpoints, and also counted their root mean square errors. For each pair of images, we repeat the statistics 20 times, and then use the average of the results. Table 1 shows the REMSs of feature points and checkpoints, corresponding to homography warp, APAP, ELA and our nonrigid warp respectively. Compared with other methods, nonrigid warp can obtain smaller REMSs (shown in bold), which means our method has higher alignment accuracy.

Table 1. Comparison of average root mean squared error RMSE for the proposed method and other methods.

Image Pair		Number	Baseline	APAP	ELA	Nonrigid
temple [34]	matches	-	3.98	2.51	0.88	**0.35**
	checkpoints	25	3.34	2.18	1.33	**1.24**
carpark [12]	matches	-	4.71	2.10	1.70	**1.084**
	checkpoints	24	6.06	1.85	4.98	**0.89**
railtracks [14]	matches	-	14.54	4.70	4.10	**1.28**
	checkpoints	21	21.54	1.80	2.29	**1.81**
building [34]	matches	-	3.66	4.33	2.81	**1.79**
	checkpoints	23	3.2	1.59	2.15	**1.76**

4.4. Limitations

The proposed nonrigid warping fits the parallax in a similar way to smooth surface fitting. Occlusions in the images will cause discontinuous changes in depth differences and make the occluded parts lack matched features. Therefore, if there are severe occlusions in the images, our method will be powerless. Figure 10 shows a failure case in which ghosting effects appear around the foreground objects. Similar to other spatially varying alignment methods, straight lines will be curved in order to achieve precise local alignment. For applications that need to preserve straight lines, adding line features may be helpful. Another limitation of our method is effectiveness. In the implementation of the experiment, the ELA method is faster than APAP, and our method is somewhere in between. The main time cost is in the step of feature match refinement. In the case of sufficient and well-matched features, an effective speed-up method is to skip the semiparametric fitting and directly calculate the nonrigid warp model.
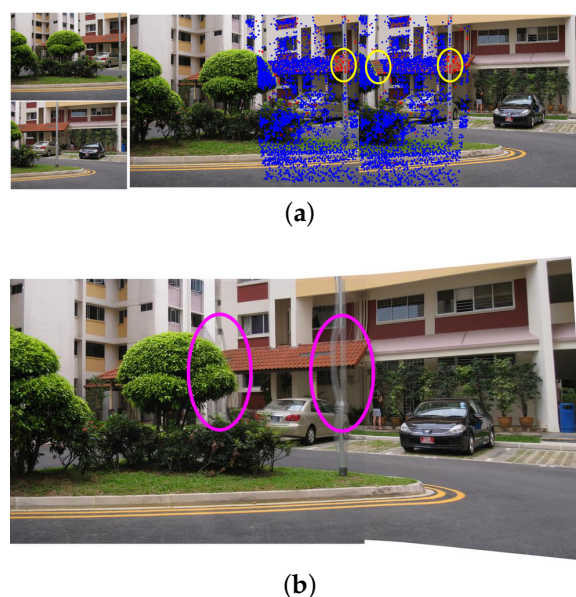

(a)

(b)

Figure 10. A failure case using our method. (**a**) Occlusions will cause the lack of feature points. (**b**) The artifacts are circled in the stitched image.

5. Conclusions

In this paper, we propose an effective image stitching method based on nonrigid warping. First, the semiparametric functions fitting is used to refine the features matched by descriptions. This new feature mapping relationship provides more feature points and helps to eliminate the influence of parallax. Second, a nonrigid warp model based on the Gaussian radial basis functions is derived from the semiparametric functions, and a uniform grid is set on the image plane to speed up the calculation of the warping. As a kind of surface fitting model, the proposed nonrigid warp can adapt to the spatial change of the projection relationship. This results in a more precise alignment in the overlapping region of the image. Finally, the nonrigid warp is effectively combined with a global transformation to improve local alignment while reducing distortion in nonoverlapping regions. The stitching quality of our method is evaluated through several comparative experiments. Our method has good performance in both visual effects and accuracy. In terms of quality, there is less blur and ghost in our stitching results. In terms of quantity, the projection error of feature points using our nonrigid warping is smaller than that of feature points using ELA and APAP methods. In future work, we will try to improve efficiency of our method, and the proposed nonrigid warp model will be applied to aerial image stitching to eliminate the parallax caused by terrain fluctuations.

Author Contributions: X.Y. was the supervisor of the whole system and experiment; L.D. designed and performed the experiments and drafted the manuscript; C.D. and J.C. conducted the collection of experimental data and gave advice on the manuscript modification; Y.C. provided assistance in terms of the code. All authors have read and agreed to the published version of the manuscript.

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

References

1. Li, G.; Ji, X.; Zhang, M. Image Stitching in Smog Weather based on MSR and SURF. *Int. J. Perform. Eng.* **2018**, *14*. [CrossRef]
2. Yuan, S.; Yang, K.; Li, X.; Cai, H. Automatic Seamline Determination for Urban Image Mosaicking Based on Road Probability Map from the D-LinkNet Neural Network. *Sensors* **2020**, *20*, 1832. [CrossRef]
3. Li, L.; Yao, J.; Xie, R.; Xia, M.; Zhang, W. A unified framework for street-view panorama stitching. *Sensors* **2017**, *17*, 1. [CrossRef]
4. Zhang, W.; Li, M.; Guo, B.; Li, D.; Guo, G. Rapid texture optimization of three-dimensional urban model based on oblique images. *Sensors* **2017**, *17*, 911. [CrossRef]
5. Shao, Z.; Yang, N.; Xiao, X.; Zhang, L.; Peng, Z. A multi-view dense point cloud generation algorithm based on low-altitude remote sensing images. *Remote Sens.* **2016**, *8*, 381. [CrossRef]
6. Shao, Z.; Li, C.; Li, D.; Altan, O.; Zhang, L.; Ding, L. An Accurate Matching Method for Projecting Vector Data into Surveillance Video to Monitor and Protect Cultivated Land. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 448. [CrossRef]
7. Szeliski, R. Image alignment and stitching: A tutorial. *Found. Trends® Comput. Graph. Vis.* **2006**, *2*, 1–104. [CrossRef]
8. Gao, J.; Li, Y.; Chin, T.J.; Brown, M.S. *Seam-Driven Image Stitching*; Eurographics (Short Papers). Available online: http://dx.doi.org/10.2312/conf/EG2013/short/045-048 (5 November 2020).
9. Zhang, F.; Liu, F. Parallax-tolerant image stitching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3262–3269.
10. Lin, K.; Jiang, N.; Cheong, L.F.; Do, M.; Lu, J. Seagull: Seam-guided local alignment for parallax-tolerant image stitching. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 370–385.

11. Herrmann, C.; Wang, C.; Strong Bowen, R.; Keyder, E.; Krainin, M.; Liu, C.; Zabih, R. Robust image stitching with multiple registrations. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 53–67.

12. Gao, J.; Kim, S.J.; Brown, M.S. Constructing image panoramas using dual-homography warping. In Proceedings of the CVPR 2011, Providence, RI, USA, 20–25 June 2011; pp. 49–56.

13. Lin, W.Y.; Liu, S.; Matsushita, Y.; Ng, T.T.; Cheong, L.F. Smoothly varying affine stitching. In Proceedings of the CVPR 2011, Providence, RI, USA, 20–25 June 2011; pp. 345–352.

14. Zaragoza, J.; Chin, T.J.; Brown, M.S.; Suter, D. As-projective-as-possible image stitching with moving DLT. In Proceedings of the IEEE conference on computer vision and pattern recognition, Portland, OR, USA, 23–28 June 2013; pp. 2339–2346.

15. Jia, J.; Tang, C.K. Image stitching using structure deformation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 617–631. [CrossRef]

16. 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]

17. Chen, J.; Wan, Q.; Luo, L.; Wang, Y.; Luo, D. Drone Image Stitching Based on Compactly Supported Radial Basis Function. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 4634–4643. [CrossRef]

18. 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]

19. Yuille, A.L.; Grzywacz, N.M. The motion coherence theory. In Proceedings of the ICCV 1988, Tampa, FL, USA, 5–8 December 1988.

20. Lin, W.Y.D.; Cheng, M.M.; Lu, J.; Yang, H.; Do, M.N.; Torr, P. Bilateral functions for global motion modeling. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 341–356.

21. Lin, W.Y.; Wang, F.; Cheng, M.M.; Yeung, S.K.; Torr, P.H.; Do, M.N.; Lu, J. CODE: Coherence based decision boundaries for feature correspondence. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 34–47. [CrossRef]

22. Chen, T.L.; Geman, S. Image warping using radial basis functions. *J. Appl. Stat.* **2014**, *41*, 242–258. [CrossRef]

23. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]

24. Shao, Z.; Chen, M.; Liu, C. Feature matching for illumination variation images. *J. Electron. Imaging* **2015**, *24*, 033011. [CrossRef]

25. Wei, L.; Zhong, Z.; Lang, C.; Yi, Z. A survey on image and video stitching. *Virtual Real. Intell. Hardw.* **2019**, *1*, 55–83. [CrossRef]

26. Zhang, G.; He, Y.; Chen, W.; Jia, J.; Bao, H. Multi-viewpoint panorama construction with wide-baseline images. *IEEE Trans. Image Process.* **2016**, *25*, 3099–3111. [CrossRef]

27. Guo, H.; Liu, S.; He, T.; Zhu, S.; Zeng, B.; Gabbouj, M. Joint video stitching and stabilization from moving cameras. *IEEE Trans. Image Process.* **2016**, *25*, 5491–5503. [CrossRef]

28. Ma, J.; Zhao, J.; Tian, J.; Yuille, A.L.; Tu, Z. Robust point matching via vector field consensus. *IEEE Trans. Image Process.* **2014**, *23*, 1706–1721. [CrossRef]

29. Chen, J.; Xu, Q.; Luo, L.; Wang, Y.; Wang, S. A robust method for automatic panoramic UAV image mosaic. *Sensors* **2019**, *19*, 1898. [CrossRef]

30. Liu, F.; Gleicher, M.; Jin, H.; Agarwala, A. Content-preserving warps for 3D video stabilization. *ACM Trans. Graph. (TOG)* **2009**, *28*, 1–9.

31. Herrmann, C.; Wang, C.; Strong Bowen, R.; Keyder, E.; Zabih, R. Object-centered image stitching. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 821–835.

32. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 91–99.

33. Chang, C.H.; Sato, Y.; Chuang, Y.Y. Shape-preserving half-projective warps for image stitching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3254–3261.

34. Lin, C.C.; Pankanti, S.U.; Natesan Ramamurthy, K.; Aravkin, A.Y. Adaptive as-natural-as-possible image stitching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1155–1163.

35. Li, N.; Xu, Y.; Wang, C. Quasi-homography warps in image stitching. *IEEE Trans. Multimed.* **2018**, *20*, 1365–1375. [CrossRef]

36. Nie, Y.; Su, T.; Zhang, Z.; Sun, H.; Li, G. Dynamic video stitching via shakiness removing. *IEEE Trans. Image Process.* **2017**, *27*, 164–178. [CrossRef]

37. Myronenko, A.; Song, X.; Carreira-Perpinán, M.A. Non-rigid point set registration: Coherent point drift. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 3–6 December 2007; pp. 1009–1016.

38. Lin, W.Y.; Cheng, M.M.; Zheng, S.; Lu, J.; Crook, N. Robust non-parametric data fitting for correspondence modeling. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; pp. 2376–2383.

39. Rohr, K.; Stiehl, H.S.; Sprengel, R.; Beil, W.; Buzug, T.M.; Weese, J.; Kuhn, M. Point-based elastic registration of medical image data using approximating thin-plate splines. In Proceedings of the International Conference on Visualization in Biomedical Computing, Hamburg, Germany, 22–25 September 1996; pp. 297–306.

40. Leys, C.; Ley, C.; Klein, O.; Bernard, P.; Licata, L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* **2013**, *49*, 764–766. [CrossRef]

41. Lourakis, M.I. Sparse non-linear least squares optimization for geometric vision. In Proceedings of the European Conference on Computer Vision, Crete, Greece, 5–11 September 2010; pp. 43–56.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.