

Article

Improvement of Image Stitching Using Binocular Camera Calibration Model

Mengfan Tang ¹ , Qian Zhou ¹, Ming Yang ^{1,*}, Yifan Jiang ¹ and Boyan Zhao ²¹ College of Computer & Information Science, Southwest University, Chongqing 400715, China² Scivatar Technologits Co., Ltd., Chongqing 401220, China

* Correspondence: yangming@swu.edu.cn

Abstract: Image stitching is the process of stitching several images that overlap each other into a single, larger image. The traditional image stitching algorithm searches the feature points of the image, performs alignments, and constructs the projection transformation relationship. The traditional algorithm has a strong dependence on feature points; as such, if feature points are sparse or unevenly distributed in the scene, the stitching will be misaligned or even fail completely. In scenes with obvious parallaxes, the global homography projection transformation relationship cannot be used for image alignment. To address these problems, this paper proposes a method of image stitching based on fixed camera positions and a hierarchical projection method based on depth information. The method does not depend on the number and distribution of feature points, so it avoids the complexity of feature point detection. Additionally, the effect of parallax on stitching is eliminated to a certain extent. Our experiments showed that the proposed method based on the camera calibration model can achieve more robust stitching results when a scene has few feature points, uneven feature point distribution, or significant parallax.

Keywords: image stitching; camera calibration; layered projection; binocular ranging; stereo correction



Citation: Tang, M.; Zhou, Q.; Yang, M.; Jiang, Y.; Zhao, B. Improvement of Image Stitching Using Binocular Camera Calibration Model.

Electronics **2022**, *11*, 2691. <https://doi.org/10.3390/electronics11172691>

Academic Editor: Oscar Deniz Suarez

Received: 13 July 2022

Accepted: 24 August 2022

Published: 27 August 2022

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



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Image stitching technology is widely used in medical, aerial photography [1], assisted driving, surveillance, virtual reality (VR), and other fields [2], but there are still some problems to be solved. Traditional image stitching algorithms have the following shortcomings: firstly, the dependence on the scene feature points is strong, which can easily lead to stitching misalignments or even complete failure, and the robustness is relatively low. Secondly, the stitching effect is different in different scenes, and lighting and parallax have obvious effects. Additionally, it is impossible to use the global homography projection change relationship for image alignment [3–6].

To address the shortcomings of traditional methods, the following methods are proposed in this paper: (1) An image stitching method based on a special plane. This method takes advantage of the fact that the relative positions of the cameras are invariant and places a checker pattern on the special plane for camera calibration. The obtained internal parameters and the external parameters relative to the pattern can then construct an accurate projection relationship between the two cameras about this plane. (2) We further introduce a hierarchical projection method based on depth information. This method uses the internal and external parameters obtained from camera calibration for stereo correction and to project images taken by the binocular cameras into a form with parallel optical axes, a co-planar imaging plane, and identical internal parameters. It is then possible to obtain the horizontal parallax of the corresponding pixel point in the overlapping area of the image by stereo matching and to calculate the depth information of the point according to the focal length and baseline length of the binocular lens. The depth information is used to layer the original image, and each layer is mapped using different relationships.

Finally, the image stitching results can be obtained by superimposing all projections. The experimental results showed that our method is more robust than other algorithms based on feature points when a scene has few feature points, uneven distribution of feature points, or significant parallax.

2. Related Work

2.1. Image Stitching

Image stitching refers to the process of seamlessly stitching several overlapping pictures into a new picture with higher resolution and a wider viewing angle through pixel alignment. In 1996, Richard Szeliski [7] proposed the Levenberg Marquardt (LM) algorithm to improve the quality of stitched panoramic images. In recent years, to solve the most critical parallax problem in image stitching, scholars in the industry have proposed algorithms such as Global Similarity Priority (GSP) [8] and Seam-guided Local Alignment (SEAGULL) [9]. Most of these algorithms are based on the meshing concept of As-Projective-As-Possible Image Stitching (APAP) [10]; on this basis, mechanisms such as line alignment constraint and contour detection were added to improve the stitching performance. However, such algorithms usually have higher requirements for stitching images. In addition, some scholars have combined image stitching with deep learning, giving rise to Learned Invariant Feature Transform (LIFT) [11]. This algorithm is based on a convolutional neural network (CNN) [12] and uses backward propagation for end-to-end training. Its training data adopts the feature points detected by Structure-from-Motion (SFM) [11]. The feature point detection performance of this model comprehensively exceeds that of Scale Invariant Feature Transform (SIFT) [13,14]. However, due to the cumbersome training process, it is still unable to be put into practical application.

2.2. Camera Calibration

To determine the relationship between the coordinates of objects in the real world and their pixel coordinates on a camera imaging plane, a geometric camera imaging model must be established, and in real-world cases, the parameters of the camera must be obtained through experiments and calculations [3], i.e., camera calibration. In 1971, Abdel-Aziz [15] first proposed a camera calibration method based on Direct Linear Transform (DLT) [16] transformation and developed a linear equation as a mathematical model of camera imaging through the corresponding relationship between three-dimensional space points and two-dimensional pixels. However, because the linear equation can only calculate linear relationships and cannot consider the distortion effect of the camera, the parameters obtained by this method are only applicable to some scenes. In 1992, Faugeras and Luong [17] proposed a camera self-calibration method which does not need a fixed reference object; instead, it is only necessary to change the camera viewpoint to shoot multiple images and establish a connection according to the same points within the images. Although this method is not limited by a reference object, the calibration process is complex and its use has not been extensive. In 1999, Zhang Zhengyou [18] proposed a camera calibration method based on a planar pattern which uses a nonlinear model for the calculation to solve the optimal results regarding the camera parameters. This method not only has high precision and low manufacturing cost of the selected reference, but also is suitable for various calibration scenes in daily life.

3. Method

3.1. Establishment of Camera Calibration Model

In this section, we first introduce the image stitching method based on the camera parameters, as image stitching methods based on the camera calibration model depend on the internal and external parameters of the camera, which may be obtained by offline calibration. Using the checker pattern for monocular calibration, we fixed the camera position, took twenty pictures of the pattern at different positions and angles, and then measured and recorded the horizontal distance between adjacent corner points on the

pattern. To construct the world coordinate system, we took the plane of the pattern as the $Z = 0$ plane, the corner point at the top left of the pattern as the origin, and the vertical outward direction of the pattern as the Z -axis. At the same time, the world coordinates of all corner points in the figure were constructed according to the measured real distance; there were then stored in a list. Then, using the SIFI algorithm, we detected all of the corner points of the figure and recorded their pixel coordinates in a separate list. The calibration process is shown in Figure 1.



Figure 1. World coordinate system construction and corner detection.

We took 20 pictures of the pattern shown in Figure 1 at different positions and angles, recorded the world coordinates and pixel coordinates of their corner points, and solved the internal and external parameters using the perspective projection model using Equation (1):

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0_3^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{1}$$

where Z_c is the scale factor, which represents the distance from the corner in the figure to the camera imaging plane, (u, v) is the pixel coordinate of the corner point, (X_w, Y_w, Z_w) is the world coordinate of the corner point, and the right side of (1) is the internal parameter matrix of the camera and the external parameter matrix of the relative pattern, respectively. Since we specified $Z = 0$ as the plane of the pattern, the Z_w value of all corner points was 0; as such, Equation (1) could be simplified follows:

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} [r_1 \quad r_2 \quad t] \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \tag{2}$$

where r_1, r_2 are the first and second column components of the rotation matrix R . Letting homography matrix H be the product of internal parameter matrix and external parameter matrix:

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} [r_1 \quad r_2 \quad t] \tag{3}$$

Since the degree of freedom of homogeneous matrix H was 8, we set $h_{33} = 1$ and then substituted it into Equation (2):

$$\begin{cases} h_{31}uX_w + h_{32}uY_w + u - h_{11}X_w - h_{12}Y_w - h_{13} = 0 \\ h_{31}vX_w + h_{32}vY_w + v - h_{21}X_w - h_{22}Y_w - h_{23} = 0 \end{cases} \tag{4}$$

There are eight unknown parameters in (4); at the very least, the pixel coordinates and world coordinates of the four diagonal points are needed to construct the linear equations and solve them. Using the constraints of the orthogonality of r_1 and r_2 units, the internal and external parameter matrices in each picture were obtained. After obtaining the above parameter matrix, the transformation relationship between world coordinates and pixel

coordinates could be constructed. However, the essence of image stitching is to project a floating image onto a plane where the target image is located. Therefore, the transformation relationship of the pixel coordinates between two images is required. Taking a binocular camera as an example, we let the internal parameter matrices of the left and right cameras be K_l and K_r , the external parameter matrices be E_l and E_r , the world coordinates of point P_w on the pattern be $(X_w, Y_w, Z_w, 1)$, and its projection points on the imaging surfaces of the left and right cameras be $p_l(u_l, v_l, 1)$ and $p_r(u_r, v_r, 1)$, which can be listed as (5):

$$\begin{cases} p_l = K_l E_l P_w \\ p_r = K_r E_r P_w \end{cases} \quad (5)$$

where $E_l = [r_{l1} \ r_{l2} \ t_l]$, $E_r = [r_{r1} \ r_{r2} \ t_r]$. After transforming Equation (5), the result was as shown in Equation (6):

$$p_l = K_l E_l E_r^{-1} K_r^{-1} p_r \quad (6)$$

In Equation (6), E_l and E_r are the external parameters of the camera imaging surface relative to the pattern, and there was a constraint condition of $Z_w = 0$. Therefore, the coordinate transformation relationship could only produce a good splicing effect on the plane where the pattern was located. When the model was used for image registration directly, obvious ghosting, dislocation, and even deformation occurred, as shown in Figures 2 and 3.



Figure 2. Image stitching results.

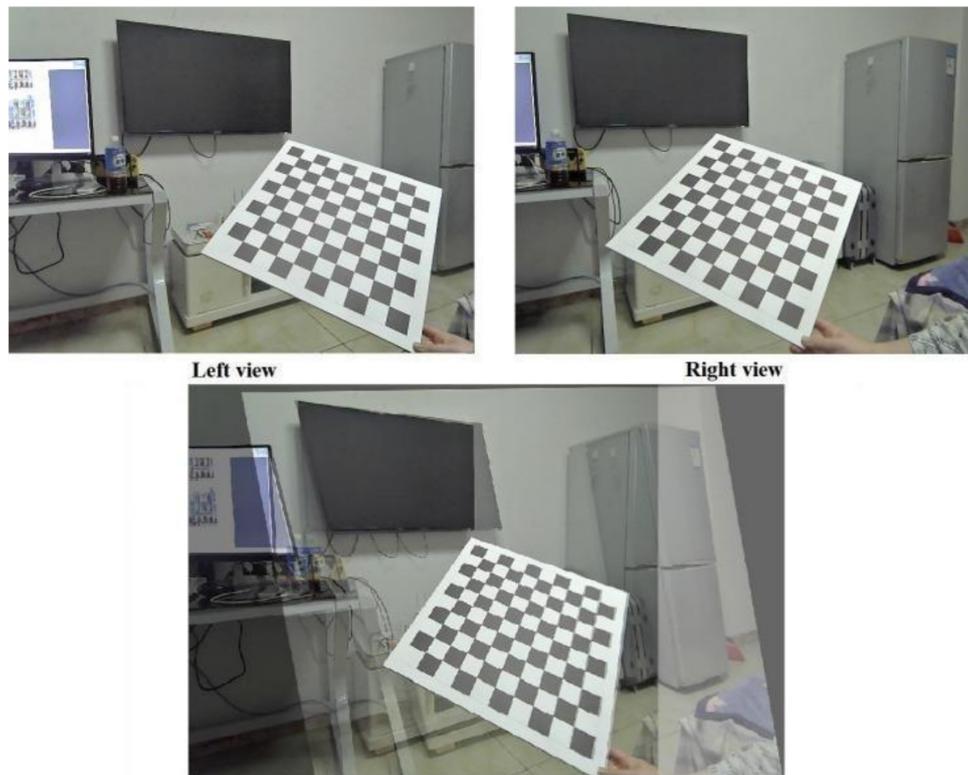


Figure 3. Image stitching results.

3.2. Design of Camera Calibration Method for Special Plane

In the camera calibration scene mentioned in the previous section, because the background object and the pattern were in different planes, the same coordinate transformation relationship could not be applied for alignment, so a more stable projection transformation model was required. The design was as follows: we fixed the binocular camera, made the line of the left and right camera optical center parallel to the wall, and fixed the pattern on the wall or vertical plane. We then took a set of pictures of the pattern in which the left and right camera imaging surface were approximately parallel to each other, as shown in Figure 4.

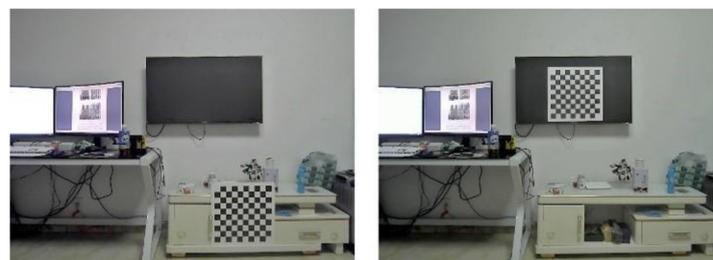


Figure 4. Camera calibration method for the special plane.

As can be seen in Figure 4, the checker pattern was fixed on the vertical plane of the cabinet and TV cabinet respectively. The binocular camera was placed at a position whereby the baseline was parallel to the wall. Therefore, it was considered that a stable projection model had been established. The external parameters were calibrated when the reference object was in this attitude and were used to construct the coordinate transformation relationship for image registration; the effect of this is shown in Figures 5 and 6. Although there were still a number of ghost dislocation phenomena in the stitching re-

sults, the plane behind the pattern achieved a relatively good stitching effect. This was because although the pattern and the object behind it were not on the same planes, the two planes were approximately parallel. It could be determined that the TV plane and the pattern plane had the same rotation matrices relative to the camera imaging plane, and that the translation vector was only slightly different in the t_z component. Therefore, using the external parameters calibrated by the checker pattern to construct the projection transformation relationship can also result in a better image stitching effect for the plane behind it.

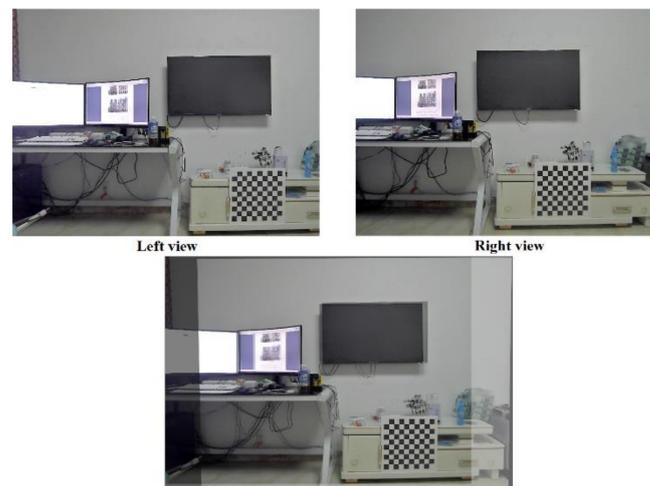


Figure 5. Image stitching results.

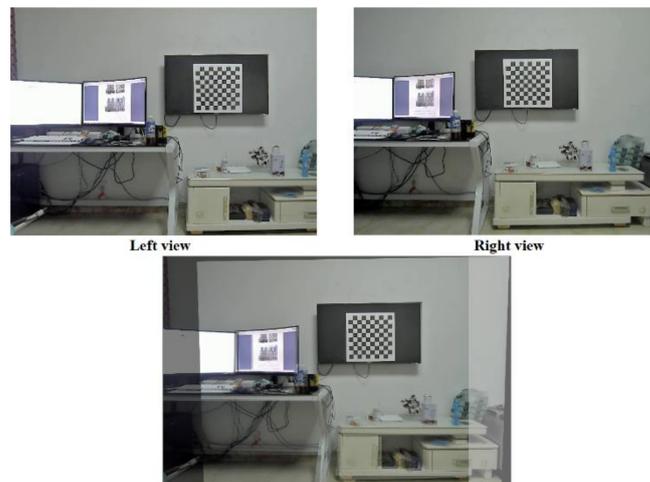


Figure 6. Image stitching results.

3.3. Design of Hierarchical Projection Method for Depth Information

A camera calibration method based on a special plane and a spatially layered image registration model was constructed with the external parameters obtained from the pattern. Although the model achieved a good stitching effect at each distance from the scene, the split image registration model could not be directly put into practical application. Therefore, it was necessary to segment the image according to the distance; to this end, an image layering method based on depth information was proposed. To obtain the depth information in the scene, we first had to calibrate the camera and obtain the relative external parameters between the left and right cameras, including the rotation matrix and translation vector. We then used this parameter to stereo correct the image and obtain the horizontal parallax of pixels in the overlapping area through stereo matching. Finally, we calculated the depth information of pixels in the scene using the corrected focal length

and baseline length. The process of binocular calibration was similar to that of monocular calibration. Based on monocular calibration, it was only necessary to take additional pictures of multiple groups of the pattern in the overlapping area for use as input data, and then to substitute the data into the perspective projection model in order to obtain the external parameters of left and right cameras relative to the pattern. The first component t_x of translation vector t is the distance between the optical centers of the two cameras and the length of the baseline of the binocular camera. After obtaining the rotation matrix between the two cameras, according to the stereo correction principle, it was decomposed into the rotation matrix of half the rotation of the left and right views, and the overall rotation matrix was constructed through the translation matrix. The image could be corrected to the attitude whereby the imaging surfaces of the two cameras were coplanar and the optical axis was parallel by using the matrix for coordinate transformation. The result is shown in Figure 7.

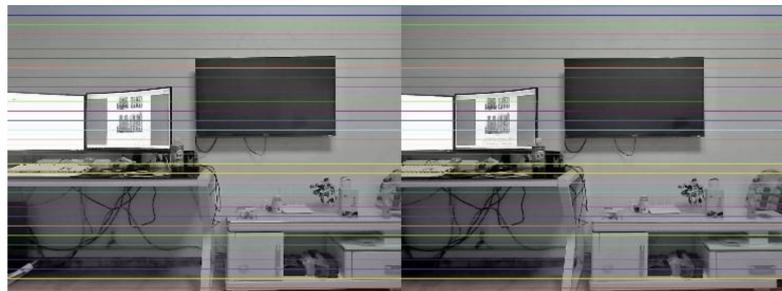


Figure 7. Stereo correction effect.

After stereo correction, each pixel in the left and right viewing angles was almost on the same horizontal line. The stereo matching algorithm then searched for the matching pixel on the corresponding horizontal line in the right-hand side image in Figure 7. The search method involved setting an odd-size sliding window, using the minimum and maximum parallax in the two images to determine the starting point and endpoint of the search, and calculating the sum of the absolute value of the gray value difference of the corresponding pixel points in the two image windows as the matching basis and selecting the point with the minimum value in the process from the start to the endpoint as the best matching point. Subsequently, the pixel coordinates of the pixel points corresponding to the left and right views in the overlapping area could be obtained. Parallax d of the point could be obtained by subtracting the abscissa of the two points. The calculated disparity map is shown in Figure 8 with the parameters and the stereo-corrected image as input.



Figure 8. Overlapping area disparity map.

After obtaining the parallax map of the overlapping area, the depth information was calculated by remapping the stereo-corrected pose to the original pose of the right perspective using Equation (7).

$$Z = \frac{f_x \times \text{Baseline}}{d} \quad (7)$$

where f_x is the number of pixels in the horizontal direction occupied by the focal length of the two cameras after stereo correction, *Baseline* is the distance between the optical centers of the two cameras, and d is the parallax of the pixel points. After converting the disparity map into a depth map, the image was layered; the effect is shown in Figure 9. The pixels in the scene were divided into several layers according to the depth information, and the average value of the depth information in each layer was recorded. At the same time, the original image of the camera angle on the right was also layered in this way, and the parts outside the overlapping area were incorporated into the layer of adjacent pixels in behavioral units. The effect is shown in Figure 10.

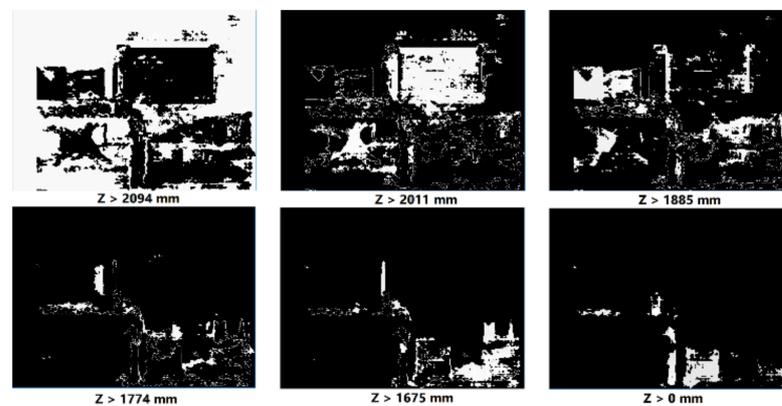


Figure 9. Image layering method based on depth information.

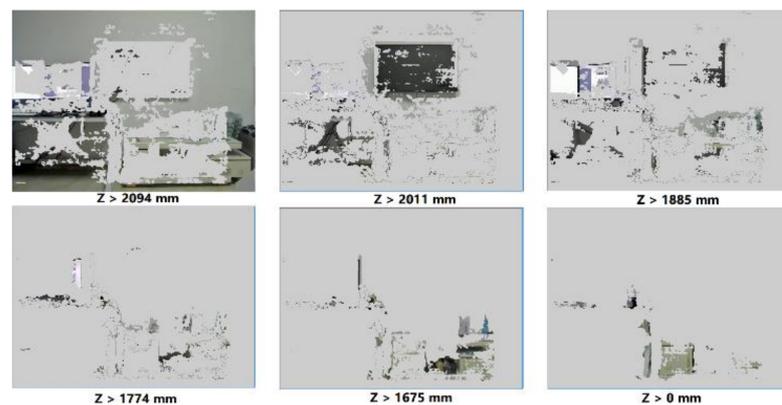


Figure 10. The original image layering effect.

After layering the original image, each layer used the pre-built projection transformation model based on the special plane and substituted the depth information of the layer into t_z in the model for calculation. Each layer used the coordinate transformation relationship calculated independently for projection, and finally, superimposed all the projection results onto the plane where the target image was located. The effect is shown in Figure 11.

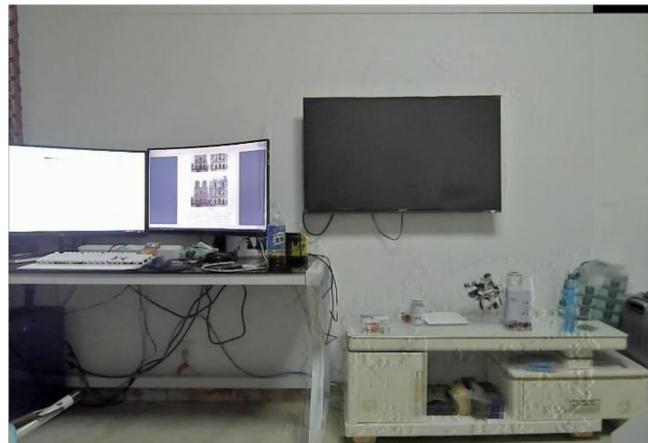


Figure 11. Stitching results based on image layering.

3.4. Process of Image Stitching Method Based on Camera Calibration

To begin, the internal parameters and the external parameters were solved by calibrating the camera. Then, using the camera calibration method based on the special plane, an additional group of pictures parallel to the camera imaging plane were taken and used as input data to solve the external parameter matrix representing the pose relationship between the pattern and the camera imaging plane. Using the external parameter matrix and the internal parameters, the coordinate transformation relationship of the binocular camera about the point on the distance plane could be constructed. In addition, the stitched image had to be layered through the layered projection method based on depth information. We then used the external parameters of the calibrated binocular camera to stereo correct the left and right viewing angles so that the corresponding pixels in the image would fall on the same horizontal line. Next, we searched the corresponding pixels in one-dimensional space using the stereo matching algorithm and obtained their horizontal parallax. After obtaining the disparity map in the above way, we calculated the depth map according to the focal length of the camera and the length of the baseline and divided the depth map according to the specific situation. Finally, the original floating map was layered according to the layered model of the depth map, and the corresponding depth information was substituted into the coordinate transformation relationship so that each layer of the image was projected according to its registration model. All projection results could then be superimposed to obtain the final image stitching result. Since the parallax of objects at the same distance imaged on the camera plane was the same, the layered projection method based on depth information could maximally eliminate the impact of parallax. The process is shown in Figure 12.

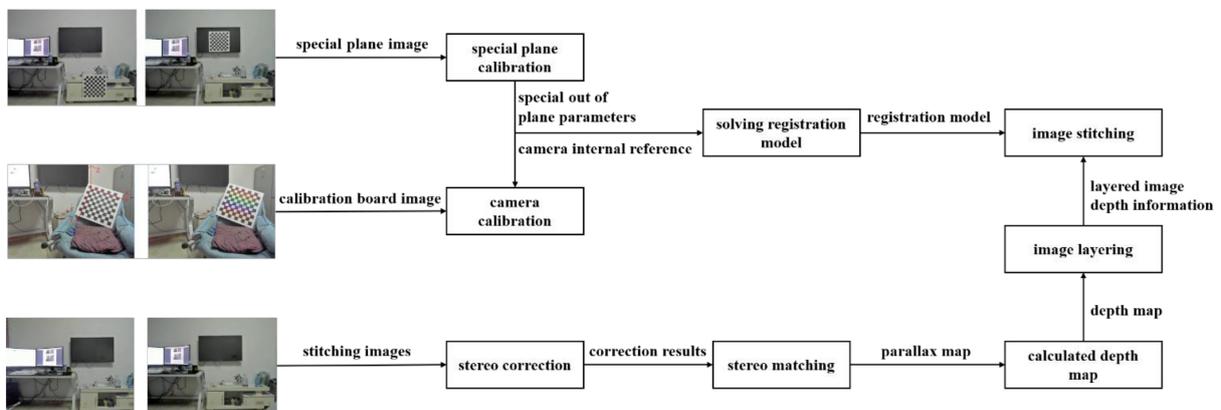


Figure 12. Image stitching process based on camera calibration model.

4. Experiment

4.1. Realization of Camera Calibration Based on Special Plane

The camera calibration algorithm flow is shown in Figure 13.

A pattern was composed of 10×10 black-and-white squares. The center of area 4×4 in the top left corner was the first corner point. There were 81 corner points on the surface of the reference object, and the horizontal distance between the points was found to be 40 mm. Before the experiment, the camera had to be monocularly calibrated many times to obtain accurate internal parameters. The internal parameter matrix of the binocular camera calibrated in the above way is shown in Table 1.

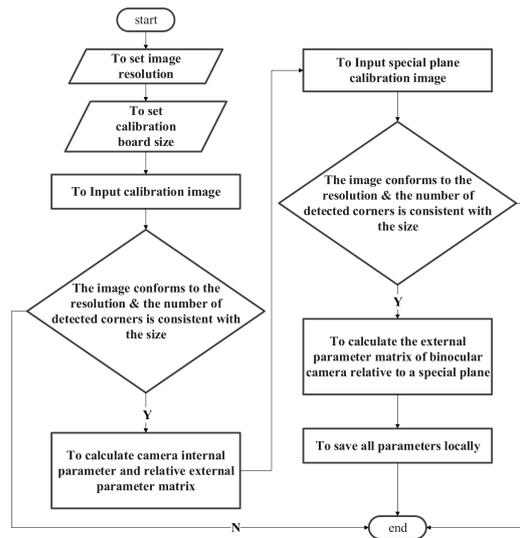


Figure 13. Camera calibration algorithm flow based on a special plane.

Table 1. Internal parameters of binocular camera.

Equipment Name	Internal Parameter Matrix	Distortion Parameter Matrix
Left camera	$\begin{bmatrix} 903.0102 & 0 & 750.2198 \\ 0 & 901.3962 & 451.7124 \\ 0 & 0 & 1 \end{bmatrix}$	$[-0.0159 \quad 0.0353 \quad -0.00127 \quad -0.0008 \quad -0.045]$
Right camera	$\begin{bmatrix} 898.9699 & 0 & 688.2513 \\ 0 & 900.6719 & 432.5414 \\ 0 & 0 & 1 \end{bmatrix}$	$[0.0022 \quad -0.027 \quad -0.0017 \quad 0.0011 \quad 0.0090]$

After monocular calibration, the relative external parameters of the binocular camera could be obtained by taking the obtained camera internal parameters and the picture group with complete reference objects in the overlapping area of the left and right viewing angles as input, as shown in Table 2.

After binocular calibration, camera calibration based on the special plane was carried out. The binocular camera was placed in a position whereby the imaging surface was parallel to the wall, and the pattern was fixed on the vertical plane in the overlapping area for photographing. The method is shown in Figure 14.



Figure 14. Camera calibration based on a special plane.

Table 2. Relative external reference of binocular camera.

Transformation Mode	Rotation Matrix	Translation Vector
From left to right	$\begin{bmatrix} 0.9999 & -0.0019 & -0.0026 \\ 0.0019 & 0.9999 & -0.0024 \\ 0.0026 & 0.0024 & 0.9999 \end{bmatrix}$	$[-16.7401 \quad -0.0788 \quad -0.0569]$
From right to left	$\begin{bmatrix} 0.9999 & -0.0016 & 0.0079 \\ -0.0017 & 0.9999 & 0.0042 \\ -0.0079 & -0.0042 & 0.9999 \end{bmatrix}$	$[16.7664 \quad 0.0526 \quad 0.0939]$

The plane of the pattern was approximately parallel to the imaging plane of the camera. The external parameters of the camera, relative to the plane of the pattern obtained using the image and known internal parameters, are shown in Table 3.

Table 3. External parameters of the binocular camera relative to the special plane.

Equipment Name	Rotation Matrix	Translation Vector
Left camera	$\begin{bmatrix} 0.0109 & 0.9998 & 0.0093 \\ 0.9993 & -0.0113 & 0.0369 \\ 0.0370 & 0.0089 & -0.9993 \end{bmatrix}$	$[9.8148 \quad -40.2235 \quad 198.7077]$
Right camera	$\begin{bmatrix} 0.0098 & 0.9997 & 0.0227 \\ 0.9999 & -0.0099 & 0.0072 \\ 0.0074 & 0.0026 & -0.9997 \end{bmatrix}$	$[-7.5794 \quad -40.2643 \quad 198.7469]$

4.2. Implementation of Hierarchical Projection of Depth Information

The implementation flow of hierarchical projection of depth information is shown in Figure 15.

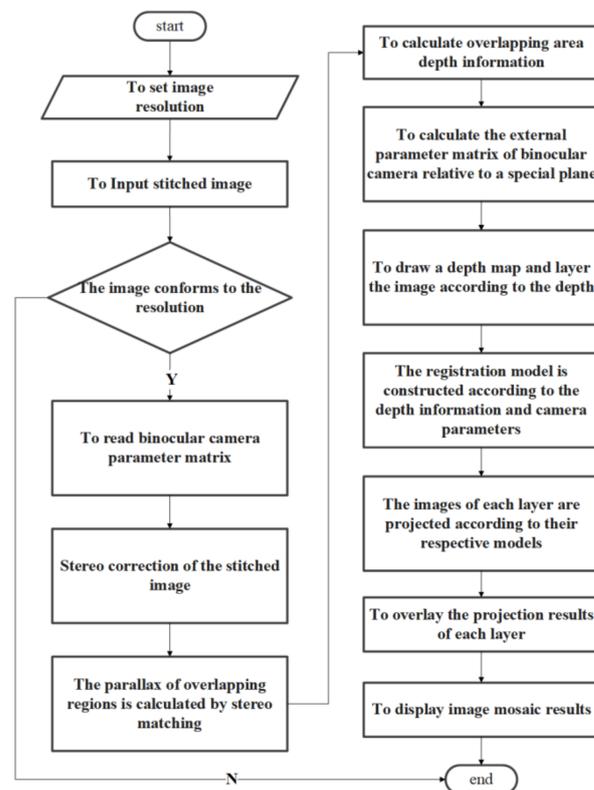


Figure 15. Hierarchical projection algorithm flow based on depth information.

To verify the hierarchical projection based on depth information, an experiment was carried out in which the variables were controlled. Firstly, the left and right camera internal parameters and special out-of-plane parameters obtained in Tables 1 and 2 were substituted into the model in Equation (6) to calculate the coordinate transformation relationship. In the experiment, different values were input for image registration. Based on the internal and external parameters of the camera, the correction matrix and projection matrix of the left and right views could be deduced, as shown in Table 4.

Table 4. Stereo correction and projection matrix.

Perspective Name	Correction Matrix	Projection Matrix
Left camera view	$\begin{bmatrix} 0.9999 & 0.0028 & 0.0008 \\ -0.0028 & 0.9999 & -0.0012 \\ -0.0008 & 0.0012 & 0.9999 \end{bmatrix}$	$\begin{bmatrix} 901.0340 & 0 & 721.6889 & 0 \\ 0 & 901.0340 & 438.3938 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$
Right camera view	$\begin{bmatrix} 0.9999 & 0.0047 & 0.0034 \\ -0.0047 & 0.9999 & -0.0012 \\ -0.0034 & 0.0012 & 0.9999 \end{bmatrix}$	$\begin{bmatrix} 901.0340 & 0 & 721.6889 & -15083.6593 \\ 0 & 901.0340 & 438.3938 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

The left and right viewing angles of the image to be stitched could then be transformed into a horizontally aligned attitude through the correction matrix, and the stereo correction results could be obtained by mapping the projection matrix to the new coordinate system. Next, we searched the corresponding points of the left and right viewing angles in the one-dimensional space, calculated the parallax and depth information, and obtained a depth map which we used to layer the image. The effect is shown in Figure 16.



Figure 16. Image layering effect based on depth information.

The image was layered in such a way that the outermost layer was greater than 2 m and each layer was separated by 100 mm. We took the average depth information of the current layer as the input to solve the respective image registration models. All layers were aligned with the corresponding models and superimposed upon one another to obtain the final stitching results.

4.3. Effect Analysis Experiment after Image Stitching

The experiments were conducted with binocular cameras. We captured close scenes several times; these were then stitched together using the method proposed in this paper, with the following results.

As shown in Figure 17, good stitching results were achieved for various planes at different distances, such as TVs, monitors, and chairs; however, objects such as table legs, which are long and thin and have different overall depth information still showed a certain degree of overlap and misalignment, and there was still blurring due to the layering in the stitched image.

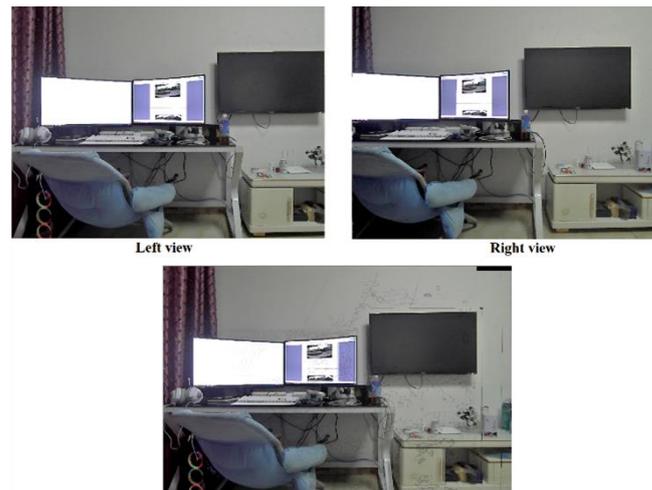


Figure 17. Image stitching results.

Figure 18 illustrates a depth span of a scene. Except for the door seam of the closet, a good stitching effect was achieved for all objects despite some blurring phenomena.



Figure 18. Image stitching results.

Similarly, Figure 19 presents to a scene with a large depth span. Once again, except for the door seam of the closet, a good stitching effect was achieved for all objects despite some blurring phenomena.

Comparative experiments were conducted to stitch the images using the method proposed in this paper and the feature point-based stitching method, respectively. The latter applies the SURF [19] and ORB (Oriented FAST and Rotated BRIEF) algorithms [20] to detect and match the feature points, purifies the matching results via the random sampling

consistency method, obtains the best matching results, constructs a global homography model to align the images, and then uses the fading-in and fading-out method to obtain the final image. The experimental results are shown in Figures 20–22.

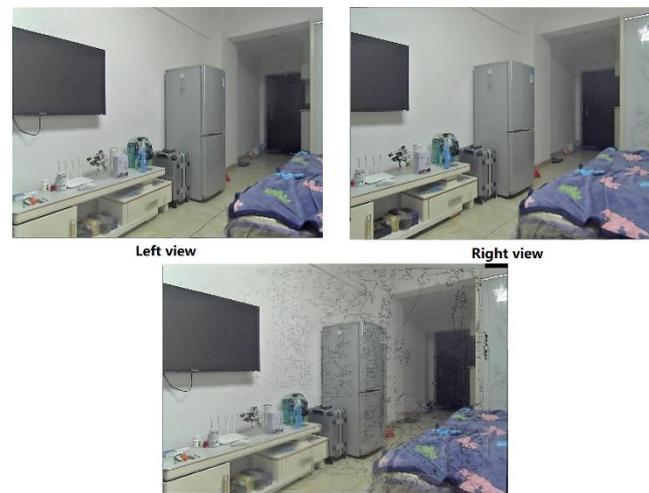


Figure 19. Image stitching results.



Figure 20. Image stitching experiment. (a) Image stitching effect based on depth information layering; (b) Image stitching effect based on SURF feature points.



Figure 21. Image stitching experiment. (a) Image stitching effect based on depth information layering; (b) Image stitching effect based on ORB feature points.

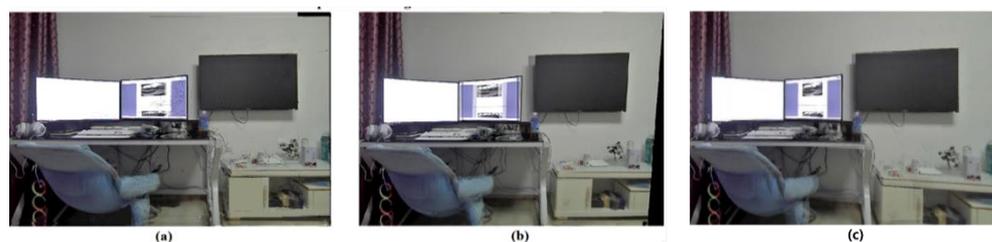


Figure 22. Image stitching experiment. (a) Image stitching effect based on depth information layering; (b) Image stitching effect based on SURF feature points; (c) Image stitching effect based on ORB feature points.

Since Figure 20 is an oblique scene, the depth information of the object surface varies linearly; as such, the objects with different depths achieved better stitching results using the method proposed in this paper (Figure 20a). In the stitching results obtained using the SURF feature points (Figure 20b), the TV, the refrigerator and the shelf all showed different degrees of misalignment.

As shown in Figure 21, when the proposed method was used for stitching (Figure 21a), good results were obtained except for the closet gap. However, in the stitching results based on ORB feature points (Figure 21b), there were mismatches, because the feature points detected for the edges of the three closet doors were too similar. As such, the result was not satisfactory.

As shown in Figure 22, using the method based on depth information layering (Figure 22a), most of the objects in the scene achieved good results, with only the seat closest to the camera showing a small amount of ghosting. Meanwhile, in the results obtained using the two image stitching method based on feature points (Figure 22b,c), most objects showed ghosting, and the upper right corner of the TV set had significant deformation.

Although the results of the method proposed in this paper demonstrated less ghosting and fewer errors, some fuzzy edge noise appeared. This may have been because the coordinate transformation relationship between the layers was not accurate enough when the image was layered. Assuming that the present registration model is not accurate enough, eliminating such edge noise will be a focus in subsequent research.

5. Conclusions

To maintain high robustness in cases of sparse feature points, uneven distribution, or obvious parallax, an image stitching method based on the camera calibration model is proposed in this paper. Based on the general camera calibration, an additional set of pictures with a vertical pattern were taken. Using the external parameters obtained from the camera calibration and the internal parameters of the camera, a spatially layered image registration model could be constructed. By adjusting the depth information in the model, the coordinate transformation relationship between the viewing angles of the two cameras concerning the vertical plane at any distance could be obtained. To apply the spatially layered image registration model, this paper also proposed an image layered projection method based on depth information. The depth information of the overlapping area in the scene was obtained through stereo correction and matching. According to this information, the original image was layered, and each layer was registered according to the coordinate transformation relationship based on the current depth information. By superimposing all the projection results, image stitching results that were resistant to parallax disturbances could be obtained.

Author Contributions: Conceptualization, M.T., Q.Z., M.Y. and B.Z.; methodology, M.T.; validation, M.T., Q.Z. and Y.J.; formal analysis, M.T.; investigation, Q.Z., M.Y., Y.J. and B.Z.; resources, M.T. and Q.Z.; data curation, M.T.; writing—original draft preparation, M.T.; writing—review and editing, M.T., Q.Z., M.Y. and Y.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by science and technology to boost the economy 2020 key project (No. SQ2020YFFO4107-66), Chongqing technology innovation and application development special general project (No. cstc2020jscx-msxmX0147), and research fund supported projects of Southwest University (No. SWU2008045).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Yuan, Y.; Fang, F.; Zhang, G. Superpixel-Based Seamless Image Stitching for UAV Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 1565–1576. [[CrossRef](#)]
2. Wang, Z.; Yang, Z. Review on image-stitching techniques. *Multimed. Syst.* **2020**, *26*, 413–430. [[CrossRef](#)]
3. Yang, Z.; Dan, T.; Yang, Y. Multi-Temporal Remote Sensing Image Registration Using Deep Convolutional Features. *IEEE Access* **2018**, *6*, 38544–38555. [[CrossRef](#)]
4. Knops, Z.F.; Maintz, J.A.; Viergever, M.A.; Pluim, J.P. Normalized mutual information based registration using k-means clustering and shading correction. *Med. Image Anal.* **2006**, *10*, 432–439. [[CrossRef](#)] [[PubMed](#)]
5. Ojansivu, V.; Heikkila, J. Image Registration Using Blur-Invariant Phase Correlation. *IEEE Signal Processing Lett.* **2007**, *14*, 449–452. [[CrossRef](#)]
6. Lucchese, L.; Leorin, S.; Cortelazzo, G.M. Estimation of Two-Dimensional Affine Transformations Through Polar Curve Matching and Its Application to Image Mosaicking and Remote-Sensing Data Registration. *IEEE Trans. Image Processing* **2006**, *15*, 3008–3019. [[CrossRef](#)] [[PubMed](#)]
7. Szeliski, R. Video mosaics for virtual environments. *IEEE Comput. Graph. Appl.* **1996**, *16*, 22–30. [[CrossRef](#)]
8. Chen, Y.; Chuang, Y. Natural image stitching with the global similarity prior. In *Computer Vision—ECCV 2016*; Springer International Publishing: Cham, Switzerland, 2016; pp. 186–201.
9. Lin, K.; Jiang, N.; Cheong, L.F.; Do, M.; Lu, J. SEAGULL: Seam-guided local alignment for parallax-tolerant image stitching. In *Computer Vision—ECCV 2016*; Springer International Publishing: Cham, Switzerland, 2016; pp. 370–385.
10. Zaragoza, J.; Chin, T.J.; Brown, M.S.; Suter, D. As-Projective-As-Possible Image Stitching with Moving DLT. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1285–1298. [[PubMed](#)]
11. Yi, K.M.; Trulls, E.; Lepetit, V.; Fua, P. LIFT: Learned invariant feature transform. In *Computer Vision—ECCV 2016*; Springer International Publishing: Cham, Switzerland, 2016; pp. 467–483.
12. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
13. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999. [[CrossRef](#)]
14. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
15. Abdel-Aziz, Y.I.; Karara, H.M.; Hauck, M. Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-Range Photogrammetry. *Photogramm. Eng. Remote Sens.* **2015**, *81*, 103–107. [[CrossRef](#)]
16. Chien, H.; Geng, H.; Klette, R. Bundle adjustment with implicit structure modeling using a direct linear transform. In *Computer Analysis of Images and Patterns*; Springer International Publishing: Cham, Switzerland, 2015; pp. 411–422.
17. Faugeras, O.D.; Luong, Q.T.; Maybank, S.J. Camera Self-Calibration-Theory And Experiments. *Lect. Notes Comput. Sci.* **1992**, *588*, 321–334.
18. Zhang, Z. Flexible camera calibration by viewing a plane from unknown orientations. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999. [[CrossRef](#)]
19. Dewanti, F.; Sumiharto, R. Purwarupa Sistem Penggabungan Foto Udara Pada UAV Menggunakan Algoritma Surf (Speeded-Up Robust Features). *IJEIS (Indones. J. Electron. Instrum. Syst.) (Online)* **2015**, *5*, 165–176. [[CrossRef](#)]
20. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011. [[CrossRef](#)]