mathematics

MDPI

Article

Three-Dimensional Reconstruction of Shoe Soles via Binocular Vision Based on Improved Matching Cost

Rui Wang [1], Lisheng Wei [1,*], Zhengyan Gu [1] and Xiaohui Liu [2]

[1] School of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, China
[2] Shanghai Ou Shuo Intelligent Packaging Technology Co., Ltd., Fengxian, Shanghai 201400, China
* Correspondence: weilsh@ahpu.edu.cn

Abstract: Aiming at the problem that the toe cap and upper part of the sole of a shoe easily appear missing when using binocular vision to reconstruct the shoe sole in the industrial production process, an improved matching cost calculation method is proposed to reconstruct shoe soles in three dimensions. Firstly, a binocular vision platform is built, and Zhang's calibration method is used to obtain the calibration parameters. Secondly, the method of fusing Census and BT costs is used to calculate the matching cost of the image, so that the matching cost calculation result is more accurate. On this basis, 4-path aggregation is performed on the obtained cost, and the optimal matching cost is selected in combination with the WTA algorithm. Finally, left–right consistency detection and median filtering are used to optimize the disparity map and combine the camera calibration parameters to reconstruct the shoe sole in three dimensions. The experimental results show that the average mismatch rate of the four images on the Middlebury website in this method is about 6.57%, the reconstructed sole point cloud contour information is complete, and there is no material missing at the toe and heel.

Keywords: binocular vision; image processing; stereo matching; 3D reconstruction

MSC: 68U10

Citation: Wang, R.; Wei, L.; Gu, Z.; Liu, X. Three-Dimensional Reconstruction of Shoe Soles via Binocular Vision Based on Improved Matching Cost. *Mathematics* **2022**, *10*, 3548. https://doi.org/10.3390/math10193548

Academic Editors: Abeer Alsadoon and Luis Coelho

Received: 1 September 2022
Accepted: 25 September 2022
Published: 28 September 2022

1. Introduction

In recent years, with rising labor costs and the development of automation technology, footwear manufacturers have begun to seek automated production of key processes in the shoemaking process [1,2]. Sole gluing is a key process in shoemaking and one that greatly affects the quality and aesthetics of shoes. In the past, sole gluing was mostly done by hand, and the volatile gas from shoe glue threatened the health of workers. Moreover, because the quality of glue applied by workers with different proficiency levels is accordingly different, more and more shoe enterprises are using robots for sole gluing instead of relying on manual operation [3]. A three-dimensional model of the sole can be obtained by 3D reconstruction, which lays a foundation for the extraction of the sole glue track. Therefore, it is of great significance to study how to use machine vision technology to realize 3D reconstruction of soles for shoe enterprises to realize automatic production.

Nowadays, machine vision technology is being gradually introduced into automatic gluing equipment. At present, most sole gluing equipment uses line-laser or binocular-structured light to reconstruct the sole in three dimensions, but the cost of line-laser and structured light is relatively high [4]. Monocular vision is affected by factors such as angle of view, and the reconstruction accuracy is low. Binocular vision can better balance hardware cost and reconstruction accuracy. At present, how to use a binocular stereo matching algorithm for 3D reconstruction of soles and improve the accuracy and speed of the algorithm has become a hot issue in the field of machine vision. Relevant scholars have conducted much research and achieved fruitful results [5–7]. On the basis of binocular vision, Li Zhenzhen et al. [8] added LED light bars, increased the contrast of sole edges, and used the edge stereo matching method for feature matching. Experiments show that the

sole edges obtained by this method are relatively complete. Ding Dukun et al. [9] adopted the threshold segmentation method based on a genetic algorithm. The experimental results show that compared with traditional methods, this method has stronger anti-interference and more accurate sole information extraction. Stefano Pagano et al. [10] used a Kinect V2 vision system to obtain point cloud data of soles and designed trajectory extraction algorithms for planar 2D objects and 3D objects, which improved the flexibility of the bonding process. Zhu et al. [11] transformed the spatial trajectory of the sole into a two-dimensional image depth map, extracted the edge contour of the sole by a two-block algorithm, and fitted and biased the contour line by B spline to obtain the spraying trajectory. Ma Xinwu et al. [12] used a canny algorithm to extract the edge of the sole and used a region-matching algorithm for stereo matching. The cubic spline interpolation was performed on the obtained three-dimensional contour line of the sole, and the glue spraying trajectory was obtained through offset. Experiments have shown that this method has good real-time performance. Luo Jiufei et al. [13] used the accelerated robust feature algorithm and the adaptive double threshold algorithm to extract initial matching pairs in left and right images, and then used distance and angle features to eliminate false matching points. Experiments have shown that the running time of this method is reduced by 40% compared with traditional methods, and the matching accuracy is higher. Yan Jia et al. [14] used a Sobel operator to extract edge information as the basis of the Census window, used the selected window to calculate the similarity cost and gradient information cost, and fused to obtain the cost matching function. Experimental results show that this method has better matching results for areas with inconspicuous texture and large depth change. Xiao Hong et al. [15] fused the gradient cost with the Census cost with the neighborhood pixel weight, and used guided filtering for cost aggregation. Experimental results show that this method can make full use of the coarse information and fine information of images, but the matching details need to be improved. Yu Chunhe et al. [16] aimed at the problem that points with high similarity tend to cause mismatching in the matching process, and used an SAD algorithm with different sizes of windows to deal with this problem. Experimental results show that the feature point matching of this method is more accurate. Zhu Jianhong et al. [17] added a third state to the traditional Census algorithm and used guided filtering for cost aggregation. Experimental results show that the time of this method is shortened by 36.60% compared with the traditional Census algorithm. Lim et al. [18] proposed a fast stereo matching algorithm based on Census transformation. The experimental results show that this method can resist radiation changes well, and the mismatching rate and running efficiency are improved compared with the traditional Census algorithm. Jia Kebin et al. [19] used a neighborhood information constraint algorithm to carry out weighted average processing of neighborhood center points, and used an adaptive color window for cost aggregation. Experiments have shown that this method has a low mismatching rate and strong anti-interference to Gaussian noise.

The above scholars have conducted much research on the extraction and 3D reconstruction of the sole gluing track [20,21] and have achieved fruitful results. However, most algorithms have high requirements on the color and texture of objects, and the matching effect is poor for objects with a single color and less texture, such as soles. Since the feature matching–based method often has missing trajectories at the toe and heel, this paper proposes an improved stereo matching method on the basis of the above research. In this paper, the algorithm that fuses Census and BT cost is used to calculate the matching cost to reduce the error caused by single-matching cost. The four-path aggregation strategy is used to aggregate the cost to improve the real-time performance of the algorithm. Left–right consistency detection and median filtering are used to optimize the disparity. Finally, the sole is reconstructed according to the calibration parameters.

2. Improved Binocular Visual Matching Cost Method

In order to realize the 3D reconstruction of the sole, this paper proposes a method to improve the accuracy of sole contour extraction on the premise of reducing hardware cost.

In this paper, a binocular vision sole 3D reconstruction algorithm with improved matching cost is studied, and its flow is shown in Figure 1.

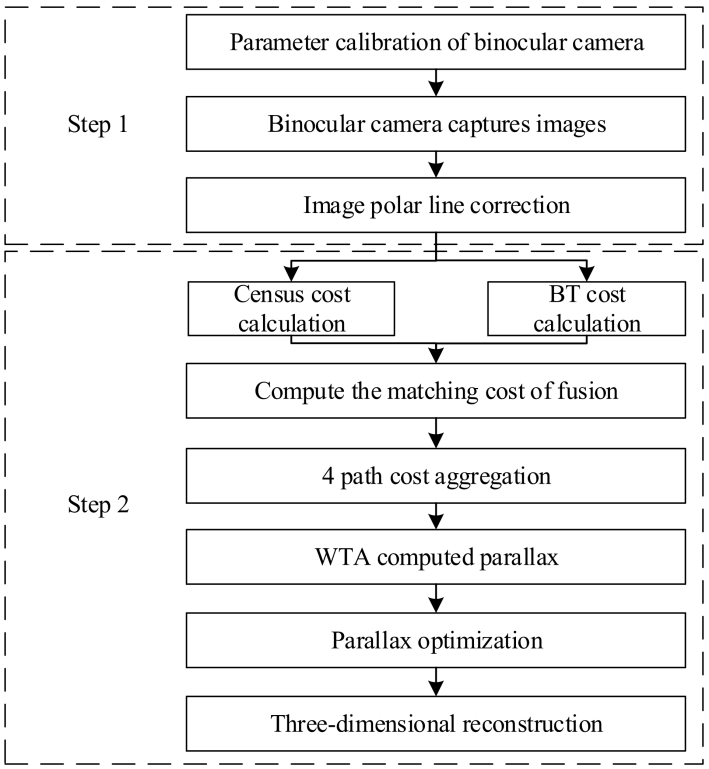


Figure 1. Algorithm flow chart of this paper.

It can be seen from Figure 1 that the 3D reconstruction algorithm of a binocular vision sole with improved matching cost has two main steps. The first step is to build a binocular vision system, use the Zhang calibration method to obtain camera parameters, and collect sole images for polar line correction. The second step is to perform stereo matching on the image. The specific steps are as follows: First, the algorithm combining BT and Census costs is used to calculate the matching cost, and the 4-path aggregation strategy is used to aggregate the cost. The obtained disparity is then post-processed using the WTA algorithm to select the final matching cost. Finally, according to the calibrated camera parameters, 3D reconstruction of the sole is performed.

2.1. Binocular Vision Image Acquisition

2.1.1. Matching Cost Calculation

Firstly, the experimental platform of the binocular camera is built, and the Zhang calibration method is used for joint calibration. In order to ensure the accuracy of the calibration parameters, an alumina calibration plate with a side length of 5 mm is used for each small square. A total of 24 pairs of calibration images are collected, and image pairs with large errors are removed. For example, samples with large calibration offsets need to be removed. Sixteen samples are used to calibrate the image, and the calibration picture is shown in Figure 2.

Use the calibration toolbox that comes with Matlab2018b to calibrate the camera. The external parameters of the camera are shown in Table 1. The internal parameters of the camera are shown in Table 2.

Figure 2. Partially calibrated pictures of the camera.

Table 1. Camera external parameter table.

Translation Matrix	$\begin{bmatrix} -40.2910 & 0.0169 & 0.0771 \end{bmatrix}$
Rotation Matrix	$\begin{bmatrix} 0.9999 & -0.1662 \times 10^{-3} & -0.0121 \\ 0.4479 \times 10^{-3} & 0.9997 & 0.0234 \\ 150.3069 & 238.3458 & 0.9997 \end{bmatrix}$

Table 2. Camera internal parameter table.

Parameter	Left Camera	Right Camera
Internal parameter Matrix	$\begin{bmatrix} 342.3530 & 0 & 0 \\ 0 & 684.0529 & 0 \\ 149.8581 & 203.7685 & 1 \end{bmatrix}$	$\begin{bmatrix} 342.5567 & 0 & 0 \\ 0 & 685.6191 & 0 \\ 150.3069 & 238.3458 & 1 \end{bmatrix}$
Distortion Coefficient	$\begin{bmatrix} 0.0521 & -0.1763 \end{bmatrix}$	$\begin{bmatrix} 0.0521 & -0.2760 \end{bmatrix}$

2.1.2. Binocular Vision Image Acquisition and Preprocessing

Binocular vision is widely used in the industry. It uses the left and right cameras to obtain images with differences, then calculates the parallax of the object, and finally obtains the 3D information of the object to be measured according to the camera parameters. In an ideal situation, the parameters of the two cameras are exactly the same and the imaging planes coincide. The schematic diagram is shown in Figure 3:

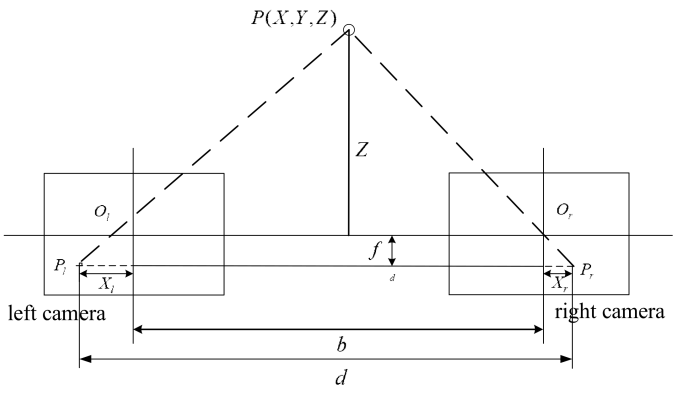


Figure 3. Principle diagram of ideal binocular vision imaging.

Where P represents the target point to be measured, and its coordinate is (X, Y, Z). O_l and O_r are the optical centers of the left and right cameras, respectively. P_l is the corresponding projection point of the target point on the left image plane, and its coordinate is (X_l, Y_l, Z_l). P_r is the corresponding projection point of the target point on the right image plane, and its coordinate is (X_r, Y_r, Z_r). d is the deviation between the two positions, called parallax. f is the distance between the optical center plane and the image plane, called the focal length. b is the baseline distance. Z is the distance from point P to the plane of the optical center, called the object distance. It can be seen from Figure 3 that:

$$d = X_l - X_r \tag{1}$$

The projection distance of the target on the left and right image planes is:

$$M_l M_r = b - X_r + X_l \tag{2}$$

where M_l and M_r represent the projection distance of the target on the left and right image planes, respectively.

According to the similarity principle of triangles, the relationship between object distance Z and parallax d satisfies:

$$\frac{Z}{Z+f} = \frac{b}{b+d} \tag{3}$$

After simplification, we get:

$$Z = \frac{f \bullet b}{d} \tag{4}$$

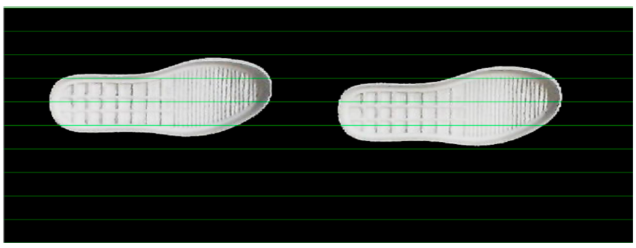
From the object distance Z, we can derive the other two coordinates X, Y of the point P in the world coordinate system.

$$X = \frac{(Z \bullet X_r)}{f} \tag{5}$$

$$Y = \frac{(Z \bullet Y_r)}{f} \tag{6}$$

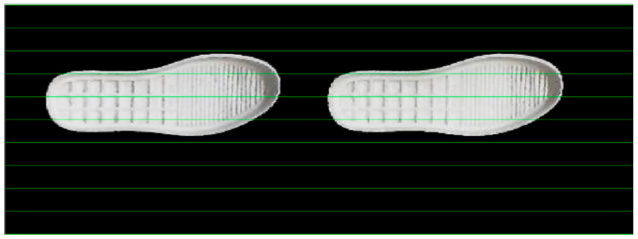
In the stereo matching algorithm, the points to be matched in the left image need to be matched with their corresponding points in the right image. Performing a two-dimensional search for all the pixels in the image on the right would take a long time. Therefore, the epipolar line constraint is adopted to make the corresponding two epipolar lines of the left and right images lie on the same horizontal line. In this way, you only need to search on the polar line corresponding to the right picture. Operations on a two-dimensional plane become operations on a one-dimensional line. This method greatly speeds up the operation. The image after epipolar correction is shown in the Figure 4.

For the convenience of observing the changes before and after the epipolar correction, equidistant horizontal straight lines are added to the figure. It can be seen from Figure 4 that the points on the edge of the sole are already at the same level after the epipolar correction.



(a) Left and right images before epipolar correction

Figure 4. *Cont.*



(b)Left and right images after epipolar correction

Figure 4. Images before and after epipolar correction.

2.2. 3D reconstruction of Improved Matching Cost Algorithm

After camera calibration and epipolar correction, the camera is used to capture left and right images of the shoe sole. In order to obtain the depth information of the sole, we need to find one-to-one corresponding pixels in the two images, and obtain the disparity map by calculating the difference between these corresponding coordinates. This process is called stereo matching.

2.2.1. Match Cost Calculation

Using only a single matching cost in the matching cost calculation usually has drawbacks. For example, the cost of BT (Birchfield & Tomasi) and Sum of Absolute Differences (SAD) relies too much on grayscale information. The Census cost relies too much on the center pixel. The sole has a single color with less texture. Therefore, it is difficult to obtain good results with a single matching cost. In order to solve this problem, the BT cost is added to the Census cost for fusion to solve the problem of depth discontinuity. In this way, the calculated matching cost is more accurate.

The Census cost compares the gray value of the center point and the other points in the neighborhood window of the point to be matched in the left and right images. This is as shown in Formula (7):

$$\zeta[IM(p), IM(q)] = \begin{cases} 0, IM(p) < IM(q) \\ 1, IM(p) \geq IM(q) \end{cases} \tag{7}$$

where $IM(p)$ represents the gray value of the center point p; $IM(q)$ represents the gray value of other points q in the field; and the center point gray value of $IM(p)$ is used as the reference value. If a point $IM(q)$ in the neighborhood is greater than $IM(p)$, it is recorded as 0; otherwise it is recorded as 1. After the pixels in the neighborhood are compared, the points in the neighborhood are arranged in order. A binary string can then be obtained. The calculation formula is shown in Formula (8):

$$C_{cen}(p) = \otimes\zeta[IM(p), IM(q)], q \in N \tag{8}$$

where $C_{cen}(p)$ represents the string corresponding to the center point, $\otimes$ means bitwise concatenation operation, and N represents the neighborhood of center point p.

The bitwise XOR and bitwise sum operations of the strings obtained in the left and right images are used to obtain the Hamming distance, which is used as the cost between the two pixels. The formula is as follows:

$$C_C(x, y, d) = Hamming[C_{cenl}(p), C_{cenr}(p - d)] \tag{9}$$

where $C_C(x, y, d)$ represents the Census matching cost; $C_{cenl}(p)$ represents the corresponding string in the left image when the parallax is d; and $C_{cenr}(p - d)$ represents the corresponding string in the right image when the parallax is d. The dimension of the Census matching cost is 5×5.

BT cost is similar to SAD cost. Both are calculated using the absolute value of the grayscale difference in the pixels, but the BT cost also performs half-pixel interpolation

on the two pixels. Taking the horizontal direction as an example, its schematic diagram is shown in Figure 5.

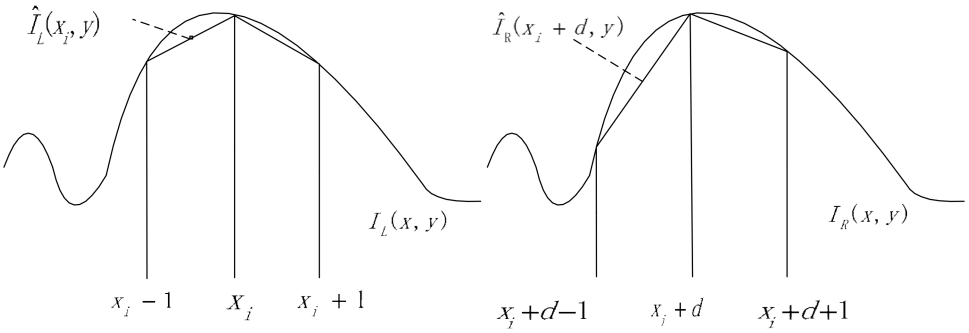


Figure 5. BT cost interpolation principal diagram.

The calculation formula of BT cost $C_B(x_i y, d)$ is shown in Formula (10):

$$C_B(x_i, y, d) = \min \begin{cases} \min \left| I_L(x_i, y) - \hat{I}_R(x + d, y) \right| \\ \min \left| I_R(x_i + d, y) - \hat{I}_L(x, y) \right| \end{cases} \tag{10}$$

where $C_B(x_i y, d)$ represents the BT matching cost, $I_L(x, y)$ represents the gray value of the pixel in the left image, and $I_R(x, y)$ represents the gray value of the pixel in the right image, $\hat{I}_L(x, y)$ represents the grayscale value at the subpixel in the left image, and $\hat{I}_R(x + d, y)$ represents the grayscale value at the subpixel in the image on the right. Among them, $x \in [x_i - 0.5, x_i + 0.5]$. The grayscale value $\hat{I}_L(x, y)$ is calculated for the subpixel location (x, y) between pixels $(x_i - 0.5, y)$ and $(x_i + 0.5, y)$ in the left image. The grayscale value $\hat{I}_R(x + d, y)$ is calculated for the sub-pixel location $(x + d, y)$ between pixels $(x_i + d - 0.5, y)$ and $(x_i + d + 0.5, y)$ in the right figure. The BT cost calculation is divided into two types because there are two left and right images.

There is a difference between the initial costs due to the adopted matching costs. Therefore, the difference needs to be normalized first and then fused. The normalized formula is:

$$\rho(c, \lambda) = 1 - \exp(-\frac{c}{\lambda}) \tag{11}$$

The normalized matching cost is fused and multiplied by the corresponding scale factor to obtain Formula (12).

$$C = \rho(C_C(x, y, d), \lambda_C) + \rho(C_B(x, y, d), \lambda_B) \tag{12}$$

Formulas (11) and (12) can be obtained simultaneously:

$$C = 2 - \exp(1 - \frac{C_C}{\lambda_C}) - \exp(1 - \frac{C_B}{\lambda_B}) \tag{13}$$

where $\rho(c, \lambda)$ represents the normalization function; c means matching cost; and λ represents the corresponding scale factor, which is used to assign the weight of the two costs. In order to balance the three costs, it has been debugged many times. Finally choose $\lambda_C = 0.8$, $\lambda_B = 0.2$.

The steps of the proposed matching cost calculation method are as follows:

(1) With the point $p(x, y)$ in the left figure as the center, build a neighborhood window of size $n \times n$.
(2) Do the same operation in the right image, and select all the pixels in the right image window at the same time.
(3) Compare the gray value of the left and right neighborhood center points and the other points, respectively. Obtain two binary strings according to the size of the gray value. Find the Hamming distance between the two strings to get the cost C_2.

(4) Calculate the sub-pixel interpolation of the center point of the left and right neighborhoods and the focus of the adjacent pixels, respectively. The cost C_3 can be obtained.
(5) Normalize C_C and C_B. Fuse them according to the corresponding scale factor. The matching cost C can then be obtained.
(6) Repeat steps 2 to 5 until the parallax search range is exceeded.
(7) Select the neighborhood with the smallest matching cost C within the disparity range of the right image. The corresponding center point is the pixel that matches the P points.

2.2.2. Cost Aggregation

After calculating the initial matching cost, if the disparity calculation is performed directly, the effect is generally poor. In order to improve the accuracy and robustness of stereo matching, cost aggregation is required. A scanline optimization method is used for cost aggregation.

Firstly, a global energy optimization strategy is used to find the minimized energy function. The optimal aggregation cost can then be found. The energy function formula is as follows:

$$E(D) = \sum_{p}(C(p, D_p) + \sum_{q \in N} P_1 T[|D_p - D_q| = 1] + \sum_{q \in N} P_2 T[|D_p - D_q| > 1]) \quad (14)$$

where Dp represents the disparity of pixel p pointsm Dq represents the parallax of q pixels, and P_1 and P_2 represent penalty coefficients. The first term on the right side of Equation (14) represents the sum of the matching costs of all pixels when the disparity is d. The second and third terms indicate that all pixels in the neighborhood N of pixel p are penalized. Among them,$P_2 > P_1$. When the parallax change is 1, P_1 is used for punishment, and when the parallax change is greater than 1, P_2 is used for punishment.

A common scheme in path aggregation is 8-path aggregation. This scheme calculates the matching cost from 8 directions, and has higher accuracy but longer operation time. Compared with the 8-path aggregation, the 4-path aggregation eliminates the four-direction paths of 5, 6, 7, and 8. Its matching speed is greatly accelerated. The strategies for 8-path aggregation and 4-path aggregation are shown in Figure 6. The calculation formula of the path cost of a pixel along a certain direction is shown in (15):

$$L_r(p, d) = C(p, d) + \min \begin{Bmatrix} L_r(p - r, d) \\ L_r(p - r, d - 1) + P_1 \\ L_r(p - r, d + 1) + P_1 \\ \min_i L_r(p - r, i) + P_2 \end{Bmatrix} - \min_i L_r(p - r, i) \quad (15)$$

where $L_r(p, d)$ represents the aggregation cost of disparity d and pixel p under the condition of path r. $C(p, d)$ represents the match cost value for pixel p. The second term on the right side of the formula indicates that the aggregation cost on path r is the value corresponding to the minimum cost when no penalty or P_1 and P_2 penalties are applied. The third term exists to prevent the path cost from being too large.

The aggregation cost in the four directions is:

$$S(p, d) = \sum_{r} L_r(p, d) \quad (16)$$

where $S(p, d)$ represents the total aggregation cost, which is the sum of the four directions of 1, 2, 3, and 4. The dimension of the aggregate cost value is 3×3.

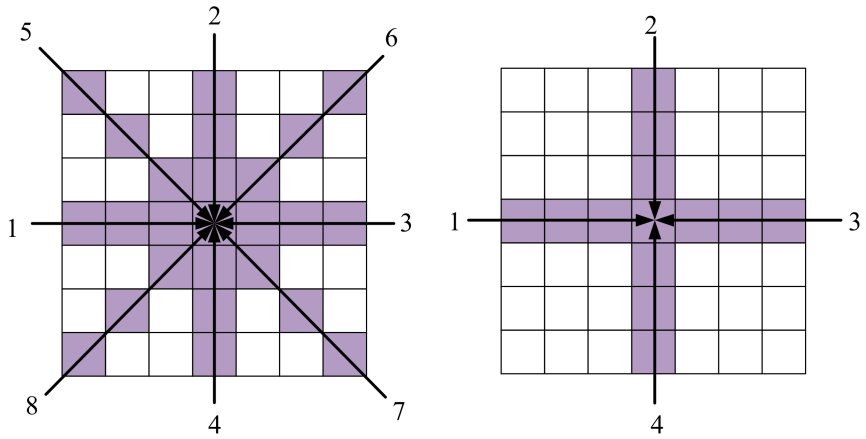


Figure 6. Schematic diagram of 8 and 4 path aggregation.

2.2.3. Parallax Optimization

After the cost aggregation for all 4 directions is obtained, the disparity corresponding to the smallest aggregated cost per pixel is used to calculate the disparity map. The winner-takes-all (WTA) algorithm is used to select the optimal cost from all possible matching costs. First, the corresponding cost value of a point in the image under all parallax d is calculated. Then the smallest cost value is found via WTA. This is the optimal parallax d_{win}. Subsequently, the picture taken by the left camera is used. The picture information corresponding to each point is replaced with its relative d_{win} for storage. The resulting picture is the desired parallax image. There are inevitably some occlusion areas due to the binocular camera. Since there are inevitably some occlusion areas in the binocular camera, it is inevitable that there will be empty areas. Parallax filling is performed on the occluded areas using left–right consistency detection. The median filtering method is used to remove excess noise and improve the accuracy of the parallax.

2.2.4. 3D Reconstruction

After the disparity map of the left and right images is obtained, the spatial coordinates corresponding to each point in the space can be calculated by combining the obtained internal and external parameters of the binocular camera. According to the basic principle of binocular vision, 3D reconstruction is performed on each point in the disparity map. The obtained set of 3D space points is the point cloud.

3. Experimental Verification and Result Analysis

(1) Experiment 1

In order to verify the effectiveness of the method proposed in this paper, we built a binocular camera experimental platform and used the proposed algorithm to reconstruct the 3D shoe sole actually used in shoe production. First, we used the four pictures of Cones, Teddy, Venus, and Tsukuba provided by the Middlebury website for testing. Twenty-six photos are used for calibration parameters, and 18 images are used for testing. The experimental platform is Matlab2018b, and the CPU is i5-4200h. In the Figure 7, from left to right are the left image to be matched, the real disparity map, the reference [16] algorithm, the Census+8 path aggregation algorithm, the reference [18] algorithm, and the disparity map obtained by the algorithm in this paper.

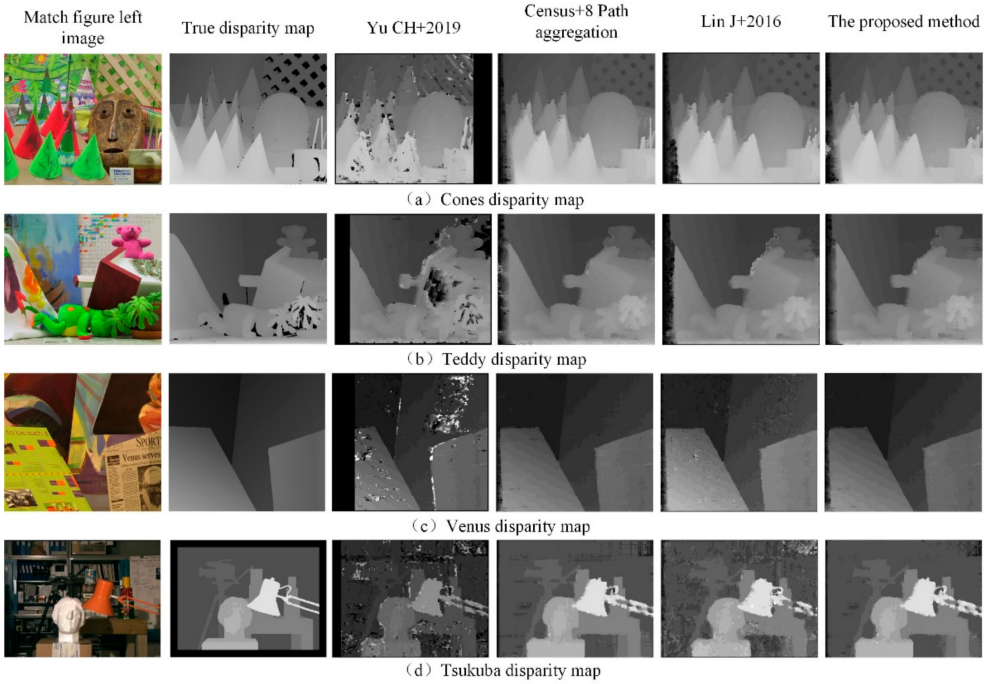


Figure 7. Comparison of test image disparity maps obtained by different algorithms [16,18].

It can be seen from the figure that the reference [16] uses local block matching and does not go through the cost aggregation step. Therefore, although the matching time is fast, there are many missing and empty areas in the disparity map, and the matching effect is poor. The other three algorithms all use the cost aggregation strategy, and the processed disparity map has significantly fewer holes. Compared with the reference [18] method and the Census+8 path aggregation method, the method in this paper retains more complete information at the edge of the image and has less noise.

The matching effect of the proposed algorithm is evaluated by using the false matching rate. The false matching rate of the above four images is calculated in the non-occlusion (non), all (all), and disparity discontinuous (disc) regions. The average value is calculated, and the mismatch rate is obtained as shown in Table 3. The smaller the false matching rate, the smaller the gap between the results obtained by the method and the real disparity map, and the more accurate the matching results. As can be seen from the table, the four disparity maps obtained by the algorithm in this paper have a false matching rate of 6.88%, 7.61%, 5.85%, and 5.94%, and an average false matching rate of 6.57%, which are all lower than in the other three methods. Disparity maps are more accurate. As regards operation time, the running time of several algorithms is shown in Table 3. The algorithm in [16] does not use the path aggregation operation; therefore, the matching time is short, but relatively more disparity map noise leads to a high false matching rate. After the 4-path cost aggregation operation, the matching time is lengthened. The average running time of the Census+8 path aggregation algorithm is 21.86 s. The average running time of the algorithm in this paper is 5.68 s. The running time of the algorithm in this paper is longer than that of the reference [16] algorithm but much shorter than the Census+8 path aggregation algorithm and the reference [18] algorithm. It can be seen that the 4-path aggregation greatly shortens the running time of the algorithm. In summary, the matching accuracy of this paper is the highest among several methods, and the running time is faster. It can meet the real-time requirements. The effectiveness of the method proposed in this paper is verified.

Table 3. Comparison table of disparity map evaluation indicators obtained by different algorithms.

Method		Reference [16]	Census+8 Path Aggregation	Reference [18]	The Proposed Method
Cones	Mismatch Rate (%)	32.24	10.39	11.11	6.88
	Running Time (s)	1.98	23.37	10.45	6.60
Teddy	Mismatch Rate (%)	36.54	14.16	15.90	7.61
	Running Time (s)	2.43	23.35	9.31	5.85
Venus	Mismatch Rate (%)	28.52	9.26	9.26	5.85
	Running Time (s)	1.60	21.38	9.77	5.62
Tsukuba	Mismatch Rate (%)	24.49	8.40	13.28	5.94
	Running Time (s)	1.69	19.34	7.24	4.65
Average	Mismatch Rate (%)	30.45	10.55	12.52	6.57
	Running Time (s)	1.93	21.86	9.19	5.68

(2) Experiment 2

This experiment involves the stereo matching of sneaker sole images using an improved matching cost-based algorithm. The resulting disparity map looks as shown in Figure 8.

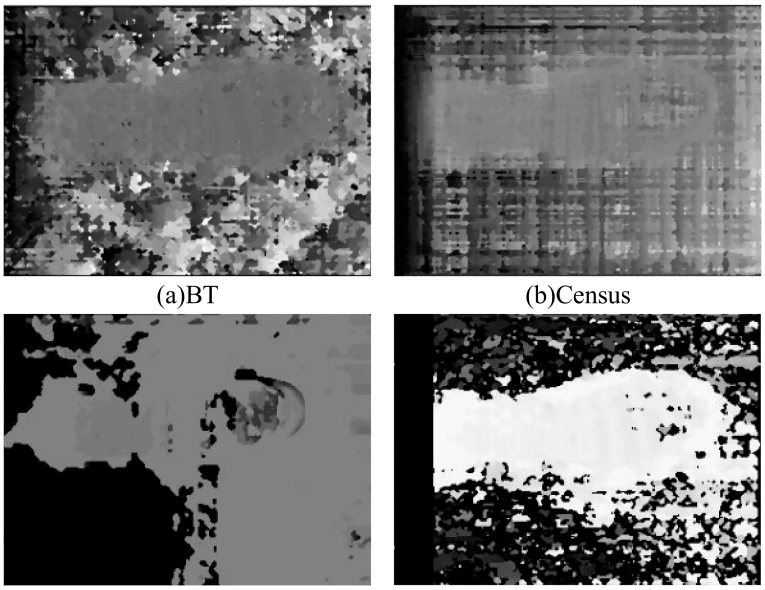
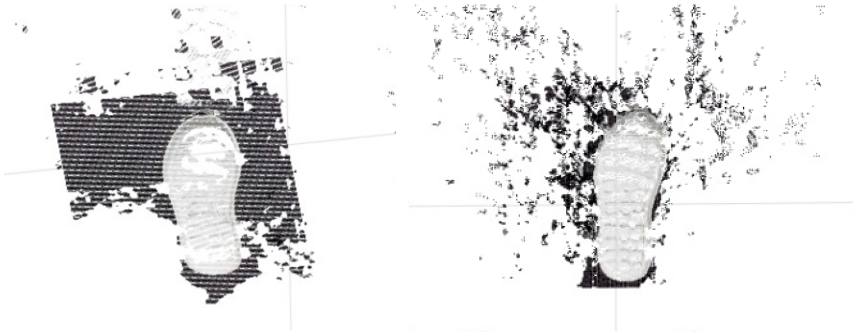


Figure 8. Disparity map of shoe sole.

It can be seen from Figure 8 that the images obtained by other methods are blurred or missing. The contours of the disparity maps obtained by our method are clear and complete. The difference in parallax between the sole and its surroundings is well represented. The method in this paper shows a good stereo matching effect. The resulting 3D reconstruction renderings are shown below:

It can be seen from Figure 9 that the soles reconstructed based on the improved matching cost algorithm retain the details better. The traditional method is prone to the phenomenon of missing details. The point cloud has an obvious layering phenomenon, and the contours of the obtained toe and heel remain intact. There is still some noise due to background distractions. According to the difference between the color of the sole and the background, the obtained point cloud image is filtered to remove the black noise. The result obtained is shown in the Figure 10:



(a)Traditional SGM (b)The method in this paper

Figure 9. 3D reconstructed point cloud image of a shoe sole.

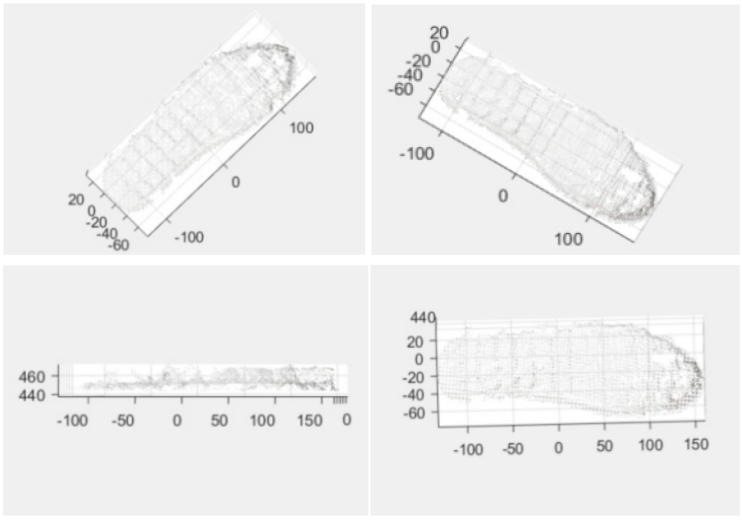
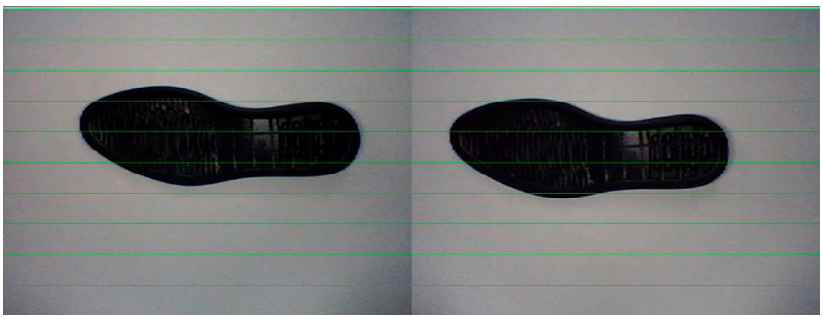


Figure 10. The filtered point cloud image of a sneaker sole.

Due to the reflection phenomenon in some areas of the sole, the middle of the reconstructed sole is partially missing. But the edge information of the sole remains intact. This part of the information is also used for the extraction of the sole gluing trajectory. Therefore, the proposed method based on improved matching cost can meet the requirements. Experiments were carried out on the leather shoe sole shown i–n Figure 11. The effectiveness of the proposed method is further verified. The results are shown in Figure 12.



(a) Left and right images before epipolar correction

Figure 11. Leather sole images before and after polar line correction.

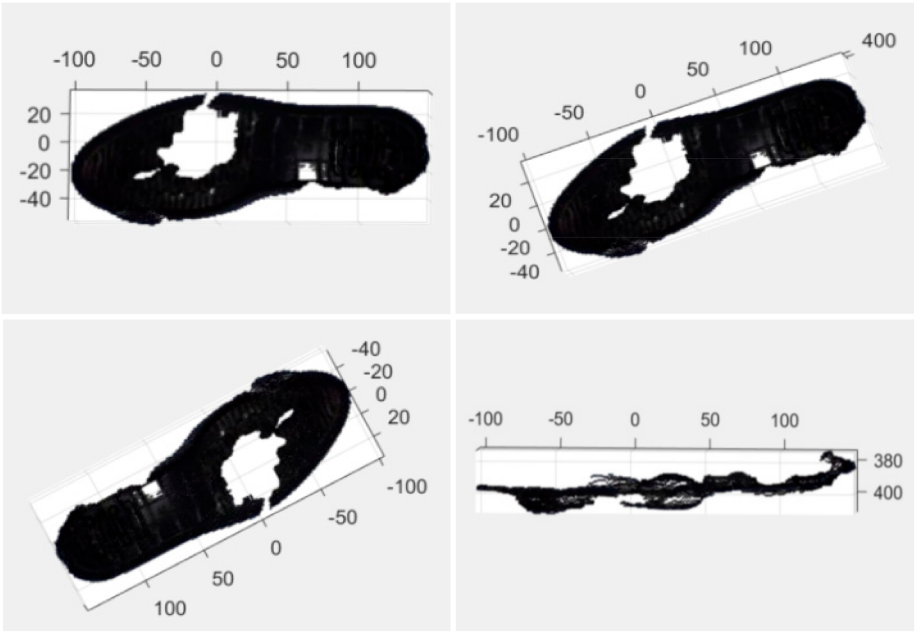


Figure 12. The filtered point cloud image of leather shoe sole.

The reconstructed point cloud images of white sneaker soles and black leather shoe soles are observed. It is obtained that the 3D reconstruction of the sole can be better achieved based on the improved matching cost algorithm. The evaluation criteria of reference [12] were used. The maximum and minimum heights of the heel and toe center were measured. The height difference was calculated. The length data of the sole was obtained by calculating the distance between the heel and the center of the toe. The accuracy of the reconstructed point cloud was evaluated. The results are shown in the table below.

It can be seen from Table 4 that the error between the reconstructed sports shoe and the leather shoe sole is within 1.30mm. Its accuracy is relatively high, and there is no missing phenomenon in important positions such as the toe and heel. This method can meet the requirements of actual production.

Table 4. Comparison table of reconstruction accuracy of different soles.

Indicators		Heel Center Height Difference (mm)	Height Difference in Toe Center (mm)	Distance between Toe and Heel Center (mm)
Sneaker Soles	Measurement	30.1	27.5	294
	Actual	29.9	26.2	293.5
Leather Shoe Sole	Measurement	31	13	253.71
	Actual	30.3	13.5	255

4. Conclusions

Due to rising labor cost, the quality requirements of shoe sole coating have gradually increased. Therefore, it is difficult for manual gluing to meet the needs of businesses. This paper adopts the method of fusing Census and BT cost to calculate the matching cost of left and right images. In this way, a more accurate matching cost can be obtained. At the same time, the 4-path aggregation strategy improves the efficiency of the algorithm. The parallax is optimized by using left–right consistency detection and median filtering methods. The proposed algorithm was evaluated with source images on the Middlebury website. The false matching rate was 6.57%, and the running speed was faster. In future research, it should be considered that when the number of feature points extracted is large, feature matching becomes difficult and there is an unnecessary matching burden. Non-maximum suppression can be cited to keep the point with the largest response and avoid the feature

set. On the other hand, deep neural networks can learn more efficient features and metric functions to replace the cost computation of traditional stereo matching methods, thus improving the accuracy of stereo matching methods. At the same time, self-calibration methods [22,23] should be studied, as they have a wide range of usage environments and are flexible and adaptable.

Author Contributions: Conceptualization, R.W. and Z.G.; methodology, Z.G.; software, R.W.; validation, R.W., L.W., Z.G. and X.L.; formal analysis, R.W. and Z.G.; investigation, Z.G.; resources, L.W.; data curation, R.W.; writing—original draft preparation, R.W.; writing—review and editing, L.W.; visualization, Z.G.; supervision, L.W.; project administration, L.W.; funding acquisition, L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by The Natural Science Research Program of Colleges and Universities of Anhui Province under grant KJ2020ZD39, by the Open Research Fund of Anhui Key Laboratory of Detection Technology and Energy Saving Devices under grant DTESD2020A02, and by the Key Project of Graduate Teaching Reform and Research of Anhui Polytechnic University.

Data Availability Statement: Not applicable.

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

References

1. Su, C.Y. In 2020, the global footwear industry production and trade both decline, and China's production and sales still rank first. *Beijing Leather* **2021**, *46*, 74.
2. Zhu, W.C. The future development trend of China's shoe industry. *Fujian Text.* **2015**, *5*, 24.
3. Ye, X.J. Challenges and opportunities coexist in China's shoe industry. *West Leather* **2015**, *13*, 14–16.
4. Ren, F.; Yu, X.; Dang, W.M. Depressive symptoms in Chinese assembly-line migrant workers: A case study in the shoe-making industry [J]. Asia-pacific psychiatry. *Off. J. Pac. Rim Coll. Psychiatr.* **2019**, *11*, 12332.
5. Huang, G. Binocular vision system realizes real-time tracking of badminton. *J. Electron. Meas. Instrum.* **2021**, *35*, 117–123.
6. Shi, L.; Zhu, H.H.; Yu, Y.; Cui, X.; Hui, L.; Chu, S.B.; Yang, L.; Zhang, S.K.; Zhou, Y. Research on wave parameter remote measurement method based on binocular stereo vision. *J. Electron. Meas. Instrum.* **2019**, *33*, 99–104.
7. Chong, A.X.; Yin, H.; Liu, Y.T.; Liu, X.B.; Xu, H.L. Research on Longitudinal Displacement Measurement Method of Continuously Welded Rail Based on Binocular Vision. *Chin. J. Sci. Instrum.* **2019**, *40*, 82–89.
8. Li, Z.Z.; Jiang, K.Y.; Lin, J.Y. Edge stereo matching algorithm of sole based on extreme constraint. *Comput. Eng. Appl.* **2016**, *52*, 217–220.
9. Ding, D.K.; Shu, Y.F.; Xie, C.X. Application of Machine Vision in the Recognition of Motion Trajectory for Shoe Machine. *Mach. Des. Manuf.* **2018**, *2*, 257–259+262.
10. Pagano, S.; Russo, R.; Savino, S. A vision guided robotic system for flexible gluing process in the footwear industry. *Robot. Comput.-Integr. Manuf.* **2020**, *65*, 101965. [CrossRef]
11. Zhu, K.Y.; Wu, J.H. *An Algorithm for Extracting Spray Trajectory Based on Laser Vision*; Institute of Electrical and Electronics Engineers: Piscataway, NJ, USA, 2017; pp. 1591–1595.
12. Ma, X.W.; Gan, Y. Research About Acquiring Sole Edge Information Based on Binocular Stereo Vision. *Electron. Sci. Technol.* **2017**, *30*, 58–62.
13. Luo, J.F.; Qiu, G.; Zhang, Y.; Feng, S.; Han, L. Surf binocular vision matching algorithm based on adaptive dual thresh-old. *Chin. J. Sci. Instrum.* **2020**, *41*, 240–247.
14. Yan, J.; Cao, Y.D.; Qu, Z. Stereo Matching Algorithm Based on Improved Census Transform. *J. Liaoning Univ. Technol. (Nat. Sci. Ed.)* **2021**, *41*, 11–14+37.
15. Xiao, H.; Tian, C.; Zhang, Y.; Wei, B. Stereo matching algorithm based on improved Census transform and gradient fusion. *Laser Optoelectron. Prog.* **2021**, *58*, 327–333.
16. Yu, C.H.; Zhang, J. Research on SAD-based stereo matching algorithm. *J. Shenyang Inst. Aeronaut. Eng.* **2019**, *36*, 77–83.
17. Zhu, J.H.; Wang, C.S.; Gao, M.F. An improved matching algorithm of Census transform and adaptive window. *Laser Optoelectron. Prog.* **2021**, *58*, 427–434.
18. Lin, J.; Kin, Y.; Lee, S. A Census transform-based robust stereo matching under radiometric changes. In Proceedings of the 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Jeju, Korea, 13–15 December 2016; pp. 1–4.
19. Jia, K.B.; Du, Y.B. Stereo matching algorithm based on neighborhood information constraint and adaptive window. *J. Beijing Univ. Technol.* **2020**, *46*, 466–475.
20. Jia, M.F.; Hu, G.Q.; Lu, C.Z. Research on automatic glue spray system based on image processing. *Manuf. Autom.* **2017**, *39*, 116–119.

21. Sun, S. *Research on Six-axis Robot Online Gluing System Based on Line Structured Light Vision Measuremen Technology*; Anhui University of Technology: Ma'anshan, China, 2020.
22. Hua, B.L.; Kai, W.W.; Kai, L.Y.; Rui, Q.C.; Chen, W.; Lei, F. Unconstrained self-calibration of stereo camera on visually impaired as-sistance devices. *Appl. Opt.* **2019**, *58*, 6377–6387.
23. Bang, L.G.; Ying, J.Y.; Ang, S.; Yang, S.; Qi, F.Y. Self-calibration approach to stereo cameras with radial distortion based on epipolar constraint. *Appl. Opt.* **2019**, *58*, 8511–8521.