

## RESEARCH ARTICLE

# 3D reconstruction method based on second-order semiglobal stereo matching and fast point positioning Delaunay triangulation

Yongbing Xu<sup>1</sup>, Kaixiang Liu<sup>2\*</sup>, Jinyan Ni<sup>2</sup>, Qingwu Li<sup>2</sup><sup>1</sup> Shandong Survey and Design Institute of Water Conservancy, Jinan, China, <sup>2</sup> College of Internet of Things, Hohai University, Changzhou, China\* [kaixiang2017@outlook.com](mailto:kaixiang2017@outlook.com)

## Abstract

Binocular vision uses the parallax principle of the human eye to obtain 3D information of an object, which is widely used as an important means of acquiring 3D information for 3D reconstruction tasks. To improve the accuracy and efficiency of 3D reconstruction, we propose a 3D reconstruction method that combines second-order semiglobal matching, guided filtering and Delaunay triangulation. First, the existing second-order semiglobal matching method is improved, and the smoothness constraint of multiple angle directions is added to the matching cost to generate a more robust disparity map. Second, the 3D coordinates of all points are calculated by combining camera parameters and disparity maps to obtain the 3D point cloud, which is smoothed by guided filtering to remove noise points and retain details. Finally, a method to quickly locate the insertion point and accelerate Delaunay triangulation is proposed. The surface of the point cloud is reconstructed by Delaunay triangulation based on fast point positioning to improve the visibility of the 3D model. The proposed approach was evaluated using the Middlebury and KITTI datasets. The experimental results show that the proposed second-order semiglobal matching method has higher accuracy than other stereo matching methods and that the proposed Delaunay triangulation method based on fast point location requires less time than the original Delaunay triangulation.

## OPEN ACCESS

**Citation:** Xu Y, Liu K, Ni J, Li Q (2022) 3D reconstruction method based on second-order semiglobal stereo matching and fast point positioning Delaunay triangulation. PLoS ONE 17(1): e0260466. <https://doi.org/10.1371/journal.pone.0260466>

**Editor:** Claudionor Ribeiro da Silva, Universidade Federal de Uberlandia, BRAZIL

**Received:** October 29, 2020

**Accepted:** November 9, 2021

**Published:** January 25, 2022

**Copyright:** © 2022 Xu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data underlying the results presented in the study are available from <http://www.cvlibs.net/datasets/kitti/index.php> and <https://vision.middlebury.edu/>.

**Funding:** This research was supported by National Key R&D Program of China (No. 2018YFC0406900), Jiangsu Key R&D Program (No. BE2018066) and Shandong Provincial Water Conservancy Research Project (No. SDSLKY201905) The funders had no role in study

## 1 Introduction

Binocular simulation of human vision matches the corresponding pixels of image pairs to obtain disparity maps and calculates the 3D coordinate of each point in the 3D scene. This process has wide applications in 3D reconstruction, industrial inspection, robotics navigation and virtual reality. However, binocular vision still has great limitations in obtaining the 3D information of a scene, although its ability to roughly calculate the depth range is sufficient for humans. For 3D reconstruction tasks, the approximate depth range is not sufficient. Recovering a realistic 3D scene in a computer requires accurate 3D coordinates. In addition, efficient and realistic texture mapping is also indispensable. We improved the key components of the 3D reconstruction pipeline based on binocular vision, including stereo matching, point cloud

design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

filtering and triangulation. Finally, the accuracy and visibility of 3D reconstruction are improved. Specifically, the proposed method includes a semiglobal stereo matching method based on second-order smoothness constraints, a point cloud smoothing method based on guided filtering, and a Delaunay triangulation method based on fast point positioning.

Stereo matching is vital in binocular vision, and directly influences the 3D reconstruction results. Stereo matching methods can be divided into four steps: cost calculation, cost aggregation, disparity calculation and disparity refinement [1]. The methods are classified into three categories according to cost aggregation: local matching methods [2–5], global matching methods [6–11], and semiglobal matching (SGM) methods [12]. The proposed second-order semiglobal stereo matching method aggregates multiple cost loss functions to enhance the robustness of the method, and the matching error of the method is decreased by pooling cost loss in different directions.

The 3D coordinates of all points are calculated according to camera parameters and disparity maps. The 3D point cloud obtained by the above steps contains considerable noise, and needs to be smoothed. Point cloud filtering methods include bilateral filtering [13] and guided filtering [14]. Bilateral filtering combines spatial proximity and intensity similarity, which can not only remove the noise points but also preserve the detailed features. However, the method has higher computational complexity. Guided filtering is a local linear filtering method with lower computational complexity than bilateral filtering that introduces guided images. The method can achieve good results. To restore the visual surface of a 3D scene, surface reconstruction and texture mapping are also required. Surface reconstruction methods include the distance field contour surface method [15], Poisson reconstruction [16], and Delaunay triangulation [17]. The distance field contour surface method derives the initial tangent plane from the K-nearest neighborhood points and extracts the contour surface by forming the distance field according to normal vector uniformity. Poisson reconstruction adopts invisible fitting, and obtains an invisible equation corresponding to the point cloud model by solving the Poisson equation. Delaunay triangulation connects 3D points into triangles according to certain rules and has high stability. After reconstructing the surface of the scene, texture mapping is performed to enhance the realism of the 3D model.

In this paper, we propose a 3D reconstruction method based on binocular stereo matching and Delaunay triangulation based on fast point positioning to improve the accuracy and efficiency of 3D reconstruction. The contributions of this paper are as follows: (1) A semiglobal stereo matching method combining multiple matching costs is proposed to obtain disparity maps, which are applied to subsequent 3D reconstruction. (2) A 3D point cloud smoothing method based on guided filtering is used to effectively remove noise points and retain the details of the 3D point cloud. (3) The Delaunay triangulation method based on fast point positioning is proposed to accelerate the Delaunay triangulation. Finally, the surface of the point cloud is accurately reconstructed, and the 3D reconstruction of the binocular image pair is realized.

The remainder of this paper is organized as follows: Section 2 introduces related work similar to the proposed method in the field of stereo matching and Delaunay triangulation. Section 3 introduces the proposed 3D reconstruction method. Section 4 discusses the overall evaluations and discussions on the proposed method. Section 5 draws conclusions and outlines directions for future research.

## 2 Related work

The 3D reconstruction method based on binocular stereo matching obtains the parallax of the left and right images through stereo matching to restore depth information, and then

constructs a surface triangle mesh and implements texture mapping to achieve 3D reconstruction. In the following, the latest developments in related research will be explained in terms of stereo matching and the construction of surface triangular meshes.

## 2.1 Stereo matching

Early stereo matching was based on feature point matching [18–20]. The sparse disparity map that was obtained must be converted into a dense disparity map through interpolation calculation, however, the interpolation process is complicated. Feature extraction and positioning have a great influence on the matching result. The current sparse stereo matching is mostly applied to tasks such as camera pose estimation in SLAM. To avoid the complex process and errors caused by interpolation, most of the existing methods for recovering the 3D information of a scene directly obtain the dense disparity map [21, 22].

The current research on stereo matching mainly focuses on matching strategies, matching cost calculation and cost aggregation. Stereo matching methods are divided into global methods [7, 9], local methods [4, 5] and semiglobal methods [12] according to the different matching cost aggregation processes. The global method is more robust to occlusion and weak texture areas, which is more complex and time-consuming. Compared with the global method, the local method is faster but less accurate. The semiglobal method (SGM) converges matching costs from multiple directions to achieve a balance of accuracy and efficiency and is widely used in vision-based 3D reconstruction. SGM uses single-pixel mutual information (HMI) as the matching cost and performs one-dimensional energy minimization along multiple directions to approximately replace the two-dimensional global energy minimization. Woodford et al. [23] proposed a quadratic pseudo-Boolean optimization to calculate the second-order smoothness constraint and achieved superior performance in experiments. The proposed second-order semiglobal stereo matching method combining multiple matching costs can take into account the time efficiency and robustness of weak texture regions.

## 2.2 Mesh reconstruction based on triangulation

Triangulation divides discrete and disordered point clouds into mesh grids in 3D space. Usually, the point cloud is projected onto a two-dimensional plane to form a discrete point set in the plane area. Then, an irregular triangulated network of the point set is constructed. Among the methods for generating triangle meshes, the Delaunay triangulation method is the best, as it avoids the appearance of ill-conditioned triangles. Common methods for constructing Delaunay triangular meshes include the divide and conquer method [24], point-by-point insertion method [25], sweep surface line method [26], and triangulation growth method [27]. The triangulation growth method has been gradually eliminated due to its low efficiency. The divide-and-conquer method recursively divides the set of points and merges it level by level from bottom to top to generate the final triangulation. This is the most efficient method, but due to its use of recursion, the running process consumes considerable memory and cannot process a large amount of data. The point-by-point interpolation method constructs a convex polygon containing all points, generates the initial triangulation, and inserts the remaining points one by one. This method occupies a small amount of memory and can handle a large amount of data. However, as the triangle increases, the operating efficiency of the insertion point gradually decreases. The proposed method of quickly locating insertion points can improve the efficiency of Delaunay triangulation based on point-by-point insertion.

### 3 The proposed method

#### 3.1 Matching cost of semiglobal stereo matching

In the actual scene, the binocular image has large differences in the gray value of pixels near the matching point due to inconsistent light intensity, camera exposure, or radiation intensity on the surface of the object. In the proposed method, the census metric and the gradient metric are used as the matching cost. The census transformation reflects the local structural features of the matching point domain, and we denote the census transformation as  $T(p)$ , where  $p$  represents the currently matched pixel. In Formula 1,  $C_{census}(p, d)$  represents the census metric between the pair of matching points with the disparity value  $d$  in the binocular image pair. This is obtained by calculating the Hamming distance between  $T(p)$  and  $T(p_d)$ .

$$C_{census}(p, d) = \text{Hamming}[T(p), T(p_d)] \tag{1}$$

The gradient metric  $C_{gradient}(p, d)$  can be obtained from Formula 2, where  $C_{gradient}(p, d)$  represents the gradient metric of a matching point pair with a disparity  $d$ .  $\nabla I(p)$  is the gradient value at pixel point  $p$ , which is obtained by the Sobel operator.

$$C_{gradient}(p, d) = \nabla I_l(p) - \nabla I_r(p_d) \tag{2}$$

Since both the census metric and the gradient metric are cost metrics for a single pixel, they will be mismatched due to the effects of noise and lighting. The proposed second-order smoothness-constrained stereo matching cost aggregation method based on multidirectional angles can improve the matching accuracy of weak texture regions. This is achieved by constraining the parallax difference in multiple angle directions to adapt to inclined or curved surfaces. In Fig 1, the direction  $r$  of pixel  $p$  is represented as a triangular area composed of pixels  $p, p - r$ , and  $p + r$ . Pixel  $p$  in direction  $r$  is expressed as:

$$\alpha_r(p) = \arccos\left(\frac{a^2 + b^2 + c^2}{2ab}\right) \tag{3}$$

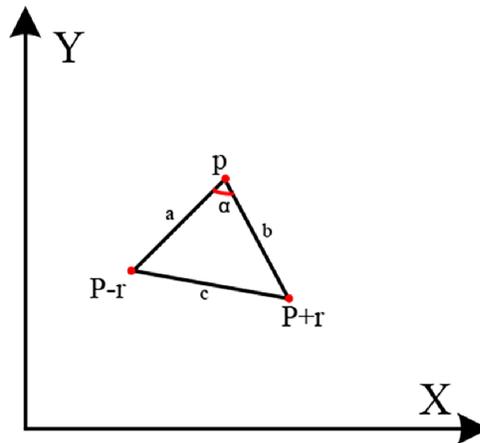
Formula 4 gives the definition of the disparity smoothness constraint  $P_{3r}(\cdot)$  of pixel  $p$  in direction  $r$ . When the angle  $\alpha$  of direction  $r$  is small, the smoothness constraint is increased to maintain the discontinuity of the parallax; when the angle  $\alpha$  of direction  $r$  is large, the smoothness constraint is reduced to adapt to the inclined or curved plane;

$$P_{3r}(p) = \left[ \frac{\pi}{\alpha_r(p)} - 1.0 \right] \cdot \tau \tag{4}$$

In Formula 4,  $\tau$  is the threshold used to prevent the parallax smoothness constraint from falling into the local minimum. The final second-order smoothness constraint is defined as:

$$V(d, d') = \begin{cases} P_{3r}(p) & \text{if } |d - d'| = 0 \\ P_1 + P_{3r}(p) & \text{if } |d - d'| = 1 \\ P_2 + P_{3r}(p) & \text{if } |d - d'| > 1 \end{cases} \tag{5}$$

where  $d$  is the disparity of pixel  $p$  and  $d'$  is the disparity of pixel  $p - r$ .  $P_1$  and  $P_2$  are two constants,  $P_2 > P_1$ . The  $P_2$  penalty is applied to pixel  $p$  with large parallax smoothness to adapt to the inclined or curved surface; the  $P_1$  penalty applied to pixel  $p$  with small parallax smoothness to preserve the discontinuity of the edge. Therefore, the final matching cost must also make



**Fig 1. Angular direction of pixel  $p$ .**

<https://doi.org/10.1371/journal.pone.0260466.g001>

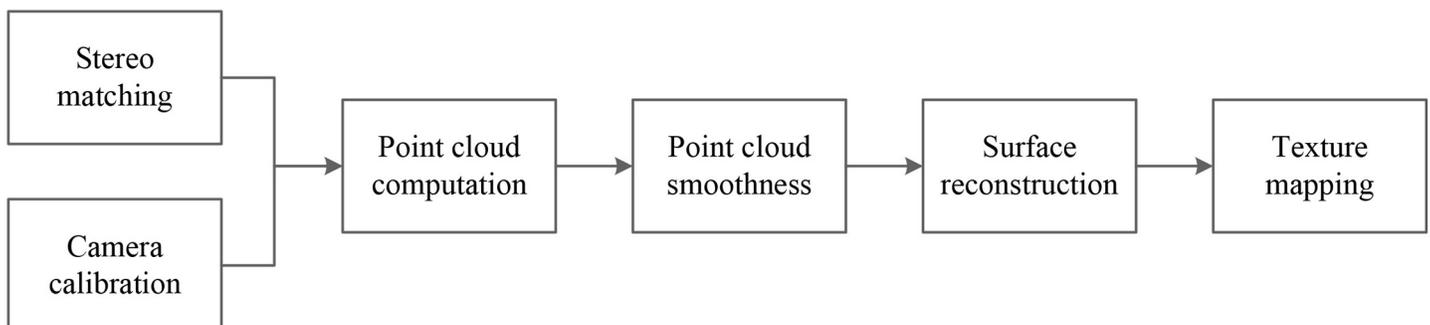
the cost aggregation of the current pixels affected by all pixels in multiple directions. Eight directions are selected from around the current pixel, and the matching cost in each direction is calculated using dynamic programming methods. Finally, the parallax is determined by the WTA(Winner takes all) rule. The path cost of a single pixel in a certain direction is defined as follows:

$$L_r(p, d) = C(p, d) + \min[L_r(p - r, d), V(d, d), L_r(p - r, d - 1) + V(d, d - 1), L_r(p - r, d + 1) + V(d, d + 1), \min_i L_r(p - r, i) + V(d, dmp)] - \min_k L_r(p - r, k) \quad (6)$$

where  $dmp = \operatorname{argmin}_i L_r(p - r, i)$ . In Formula 6,  $L_r(p, d)$  represents the path cost of a single pixel in direction  $r$ ,  $d$  is the parallax of the pixel, and  $C(p, d)$  represents the aggregation of  $C_{census}(p, d)$  and  $C_{gradient}(p, d)$ .

### 3.2 3D reconstruction based on guided filtering and Delaunay triangulation

The flowchart of the proposed 3D reconstruction method is shown in Fig 2. It includes point cloud computation, point cloud smoothing, surface reconstruction, and texture mapping. The point cloud is obtained by combining the disparity maps and camera parameters, the disparity maps are obtained by stereo matching and the camera parameters are obtained by camera



**Fig 2. Flow chart of the 3D reconstruction method.**

<https://doi.org/10.1371/journal.pone.0260466.g002>

calibration. The point cloud is smoothed by filtering to remove the noise points. Finally, surface reconstruction and texture mapping are performed to acquire an authentic 3D model. This section focuses on point cloud smoothing and surface reconstruction during 3D reconstruction.

**3.2.1 Point cloud smoothing based on guided filtering.** Due to errors in camera calibration and stereo matching, the obtained 3D point cloud includes much noise. Therefore, the point cloud must be smoothed before surface reconstruction. This section adopts guided filtering to smooth the point cloud. Guided filtering is a local linear filter method that effectively removes noise points and preserves the detail features. Guided filtering introduces the guiding image  $G$ . Here,  $G$  is the input itself and the relationship between the output image  $O$  and the input image  $I$  can be obtained as follows:

$$O(p) = \sum_{q \in W_p} W_{pq}(G)I(q) \tag{7}$$

where,  $q$  is the neighborhood pixel of pixel  $p$ , and the weight  $W_{pq}(G)$  is defined as follows:

$$W_{pq}(G) = \frac{1}{|w|^2} \sum_{(p,q) \in W_k} \left\{ 1 + \frac{[G(p) - \mu_k](G(q) - \mu_k)]}{\sigma_k^2 + \epsilon} \right\} \tag{8}$$

where,  $\mu_k$  and  $\sigma_k^2$  are the mean and variance of all pixels' gray value in the window  $w_k$ ,  $|w|$  is the size of the window, and  $\sigma$  is an adjusting parameter. Guided filtering is introduced to 3D point clouds, and the original point cloud and smoothed point cloud satisfy the local linear relationship in depth. After guided filtering, all the vertices move along their normal direction, which is defined as follows:

$$V' \leftarrow V + d \cdot n \tag{9}$$

where,  $V$  is a vertex in the point cloud,  $V'$  is the vertex after smoothing by guided filtering,  $d$  is a weighting factor. The normal direction  $n$  of vertex  $V$  is calculated by local surface fitting. A plane is established to estimate the local geometry of the vertex  $V$ , and is defined as follows:

$$F(x, y, z) = Ax + By + Cz + D = 0 \tag{10}$$

where,  $A, B, C$  and  $D$  are plane parameters. The KD-tree algorithm is adopted to select ten points near vertex  $V$  as the neighborhood spatial points. The coordinates of these points are substituted into Formula 10 so that the plane parameters can be estimated by the least squares method. Then, the updated normal direction is obtained as follows:

$$n = \left( \frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z} \right) \tag{11}$$

The weighting factor  $d$  is determined by the depth similarity between the center point and the local neighborhood points. The depth similarity refers to the distance vector that is projected onto the normal direction of the center point and is defined as follows:

$$d = \frac{1}{|s|^2} \sum_{(p,q) \in S_V} \left[ 1 + \frac{(\langle n, V - p \rangle - \mu_v)(\langle n, V - q \rangle - \mu_v)}{\sigma_v^2 + \epsilon} \right] \tag{12}$$

where  $S_V$  is a neighborhood point set centered on vertex  $V$ ,  $p$  and  $q$  are the neighborhood points of vertex  $V$ , and  $S$  is the size of set  $S_V$ .  $\langle n, V - p \rangle$  and  $\langle n, V - q \rangle$  indicate the depth similarity between the two points  $p$  and  $q$  and vertex  $V$ .  $\mu_v$  and  $\sigma_v$  indicate the mean and variance of all points' depth similarity in the set  $S_V$ , respectively.



**Fig 3. Triangulation.**

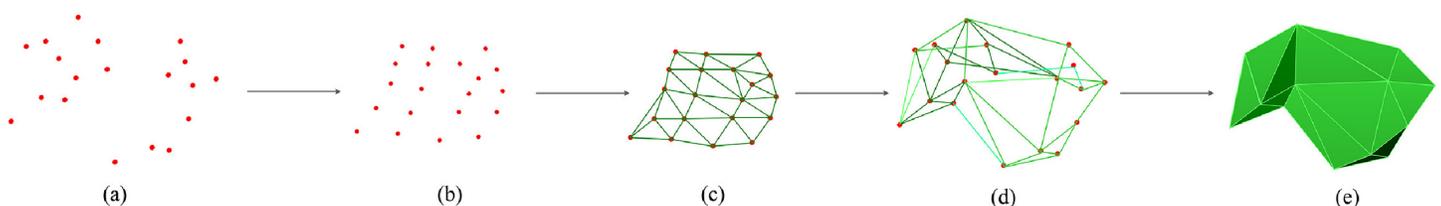
<https://doi.org/10.1371/journal.pone.0260466.g003>

**3.2.2 Projection-based Delaunay triangulation.** To improve the visibility of the 3D model, a surface patch must be added on the point cloud. The surface patch can be any shape, generally a triangle or quadrangle is selected. Since a triangle is the smallest unit that constitutes the plane, it does not easily deform during the rotation process. Therefore, a triangle is selected as the basic unit to perform triangulation. As shown in Fig 3, projection-based Delaunay triangulation can be roughly divided into the following three steps:

1. Map a 3D point cloud to a 2D plane using orthogonal projections based on the normal direction.
2. Triangulate the 2D point set obtained from the mapping according to Delaunay criterion to determine the topological connection relationship between points.
3. Determine the topological connection between the original 3D points according to the topological connection relationship of the projection points in the plane. The obtained triangular mesh is the reconstructed surface model.

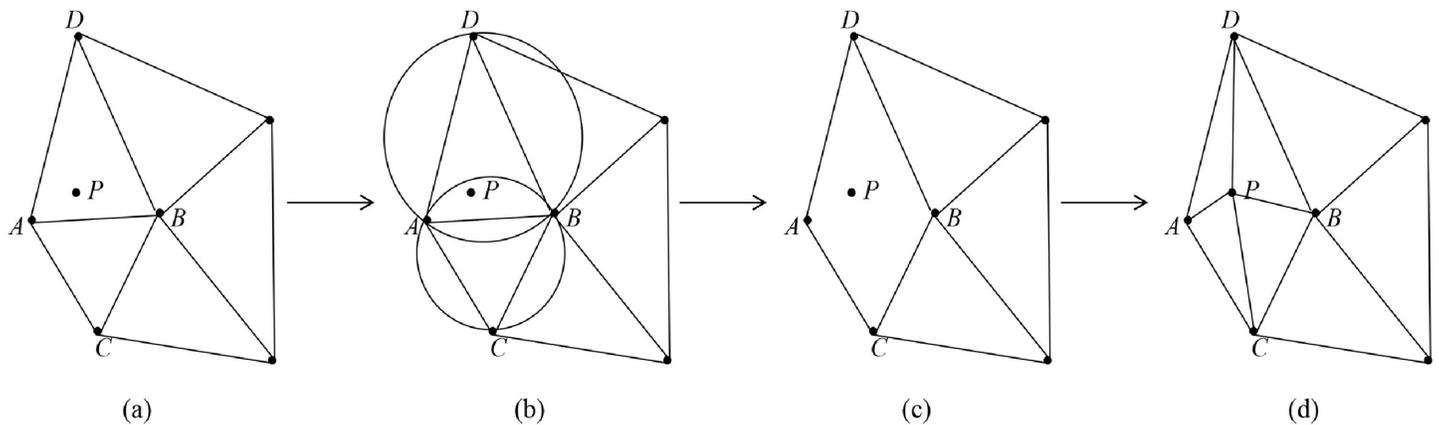
Among the triangulation methods, Delaunay triangulation has the best mathematical features such as the closest, uniqueness, optimality, most regular, and convex polygon surface, so it is adopted for surface reconstruction. A projection-based Delaunay triangulation method is performed. First, the 3D point set is mapped to the  $xOy$  plane. Second, the plane point set is triangulated according to the Delaunay triangulation method based on the Bowyer-Watson algorithm [28]. Finally, the plane triangle mesh is mapped to the 3D space to generate a 3D model. Then, texture mapping is performed to restore the surface features of the 3D model. A flow chart of the projection-based Delaunay triangulation method is shown in Fig 4.

The Bowyer-Watson method is currently the most common point-by-point insertion method, as shown in Fig 5. First, a supertriangle containing all of the scatter points is constructed. Second, each point is inserted in turn, and the influence triangles containing the point are found. Third, the common edges of the influence triangles are deleted, and the point and all the vertices of the influence triangles are joined. Finally, the unique triangular mesh model is obtained by adjusting the diagonal lines until all the points are inserted.



**Fig 4. Flow chart of the projection-based Delaunay triangulation.** (a) 3D point set (b) plane point set (c) plane triangulation (d) 3D triangulation (e) texture mapping.

<https://doi.org/10.1371/journal.pone.0260466.g004>

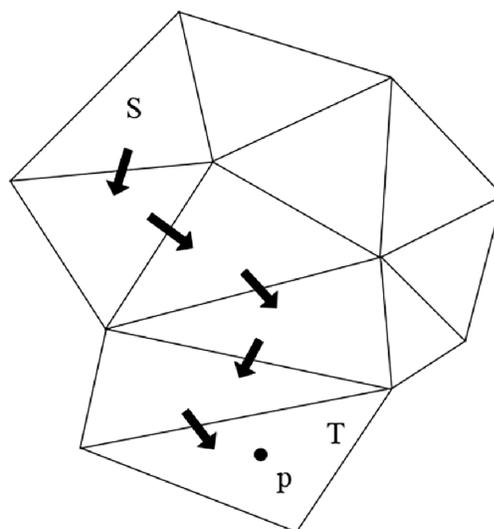


**Fig 5. Flow chart of the Bowyer-Watson algorithm.** (a) inserting the point  $P$  (b) searching for the influence triangles (c) deleting common edge  $AB$  (d) forming triangles.

<https://doi.org/10.1371/journal.pone.0260466.g005>

In the process described above, the triangle containing the insertion point and the influence range of the insertion point must be found every time a point is inserted. The original Bowyer-Watson method needs to traverse the edges and vertices of all triangles. As the triangle increases, the time cost of these two steps will increase significantly. A method for quickly locating the insertion point based on direction search is proposed, which can quickly locate the triangle containing the point to be inserted. In the following,  $G$  indicates the barycentric of the triangle and  $P$  indicates the point to be inserted. Take the newly generated triangle as the initial triangle. Starting from the initial triangle, the search direction is determined by the relative position of  $G$  and  $P$ . When  $G$  and  $P$  are on different sides of a certain edge, the next triangle to be searched is the triangle adjacent to this edge. Stop searching when  $G$  and  $P$  are on the same side of the three edges of the triangle. This triangle contains the point to be inserted. Fig 6 shows the direction search method.

In Fig 6,  $P$  is the point to be inserted,  $S$  is the initial triangle, and  $T$  is the target triangle. In terms of implementation, the vertices of the triangle are arranged clockwise, such that the



**Fig 6. Method of quickly locating insertion point based on direction search.**

<https://doi.org/10.1371/journal.pone.0260466.g006>

barycentric of the triangle is always on the right side of the edge. It is only necessary to determine whether the point to be inserted is on the left edge. For some special cases:

1. The point to be inserted is on a certain edge of the triangle: it is regarded as inside the triangle.
2. The point to be inserted is the vertex of the triangle: do not insert the point.
3. The point to be inserted is on the extension line of a certain edge of the triangle: it is considered to be on the same side of the edge as the barycentric. Continue to evaluate the next edge.

The method of quickly locating the insertion point based on direction search can be used to locate the triangle containing the points very quickly due to the point-by-point insertions according to the principle of spatial proximity.

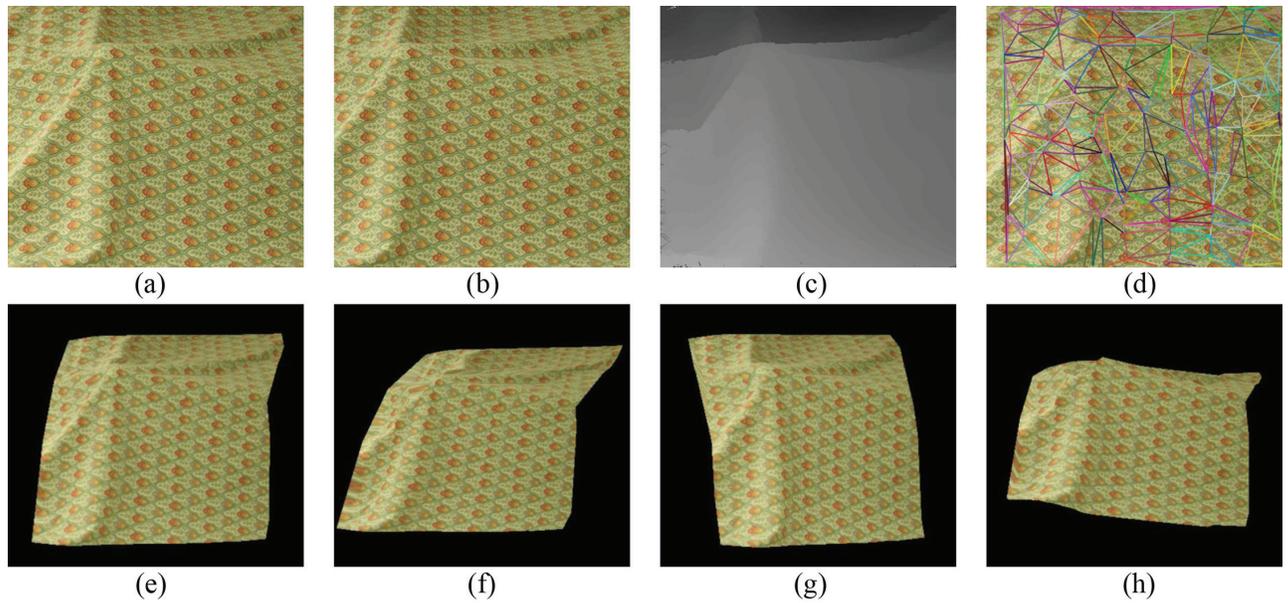
## 4 Results and discussion

To verify the effectiveness of the proposed method, the C++ language is combined with OpenCV and OpenGL to realize the proposed method. The hardware platform is an Intel Core i5-4210H@2.90GHz. The Middlebury dataset [29, 30] and KITTI dataset [31] are adopted to evaluate the performance of the proposed method, and the depth information is compared with the ground truth to perform quantitative analysis of the reconstruction performance. The scene of the Middlebury dataset is an indoor environment, so a qualitative analysis of 3D reconstruction visualization for the Middlebury dataset is selected. In addition, our results are evaluated on the KITTI 2012 and KITTI 2015 benchmarks to obtain a quantitative analysis of the reconstruction effect. The contrast methods have certain similarities with the proposed semiglobal stereo matching method. Second, we compare the proposed Delaunay triangulation of fast point positioning with the original Delaunay triangulation-based on insertion points to show the speed advantage of the proposed fast point location method.

### 4.1 Middlebury dataset

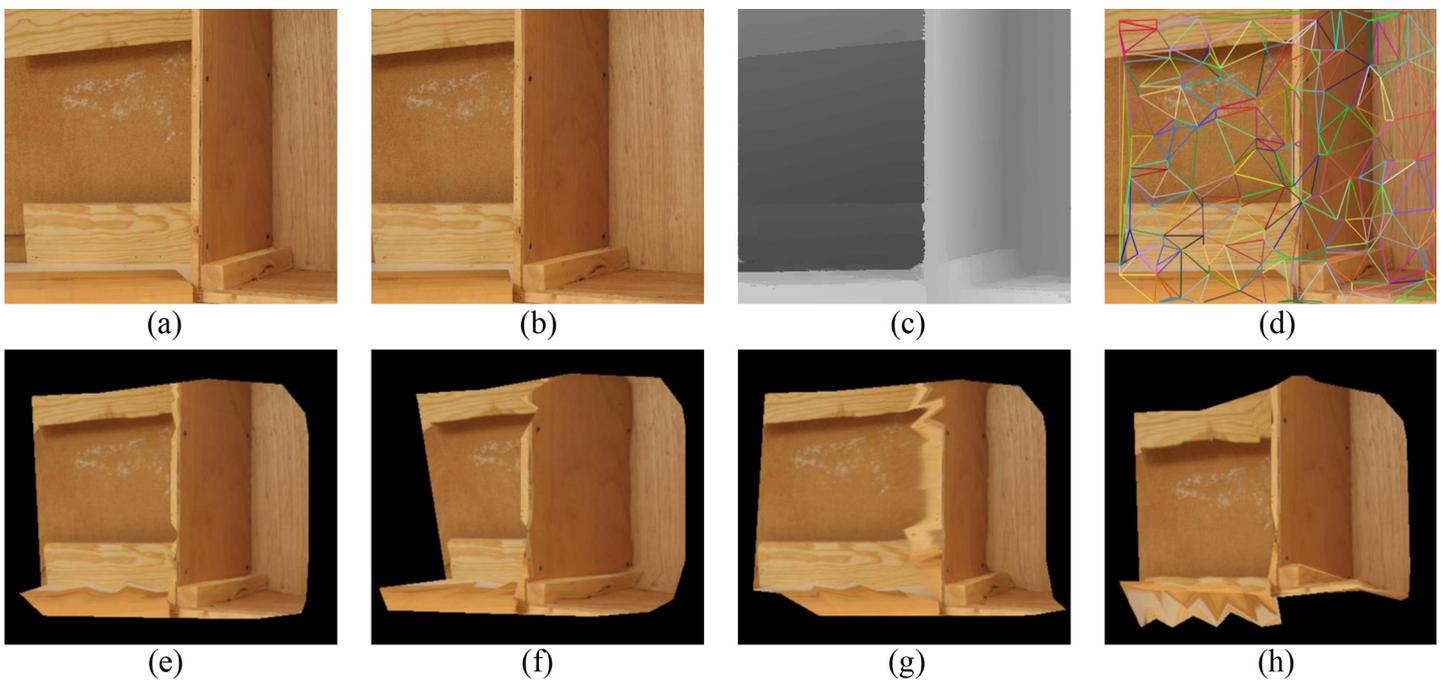
In this section, six image pairs in the Middlebury dataset are selected to perform a 3D reconstruction experiment. The experimental results are shown in Figs 7–11, for the Cloth1, Wood1, Djembe, Piano, and Shelves scenes, respectively. In these figures, (a)-(d) are the left image, right image, disparity image, and Delaunay triangulation, while (e)-(h) are the 3D reconstruction results of these scenes from different angles. The 3D reconstruction results of scenes Cloth1 and Cloth3 are accurate, and the proposed method can accurately restore the surface of the cloth. The 3D reconstruction results of other scenes are not ideal due to the occurrence of depth discontinuity in these scenes.

The proposed method was also compared with other semiglobal stereo matching methods and the comparison results are shown in Figs 12 and 13. We compare the disparity maps generated by different stereo matching methods as a qualitative analysis of 3D reconstruction because comparing the texture mapping models of different methods will make distinguishing the contrast method from the proposed method difficult. Fig 12 is the Wood1 scene in the Middlebury2 dataset, where the comparison method is ARW [32] and MST-CD2 [33]. Fig 13 shows the Piano scene and Shelves scene in of the Middlebury3 dataset, where the comparison method is CSCA [34], REAF [35] and SGM [12]. In these four scenes, the proposed method has more accurate disparity edges than other methods, so that regions of different depths can be distinguished more accurately.



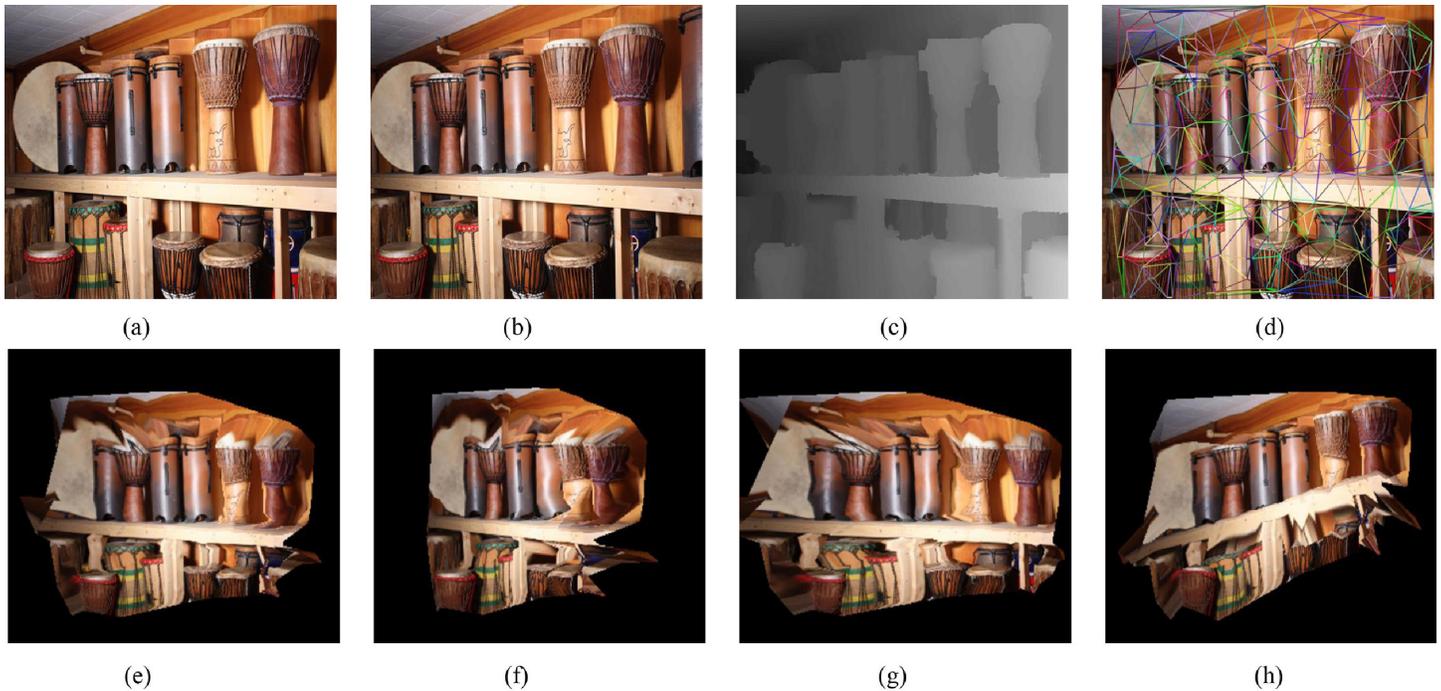
**Fig 7. 3D reconstruction experiment of scene Cloth1.** (a) left image, (b) right image, (c) disparity image, (d) Delaunay triangulation, (e) view 1, (f) view 2, (g) view 3, (h) view 4. (a) and (b) are republished from [vision.middlebury.edu/stereo/] under a CC BY license, with permission from [D. Scharstein], original copyright [2002].

<https://doi.org/10.1371/journal.pone.0260466.g007>



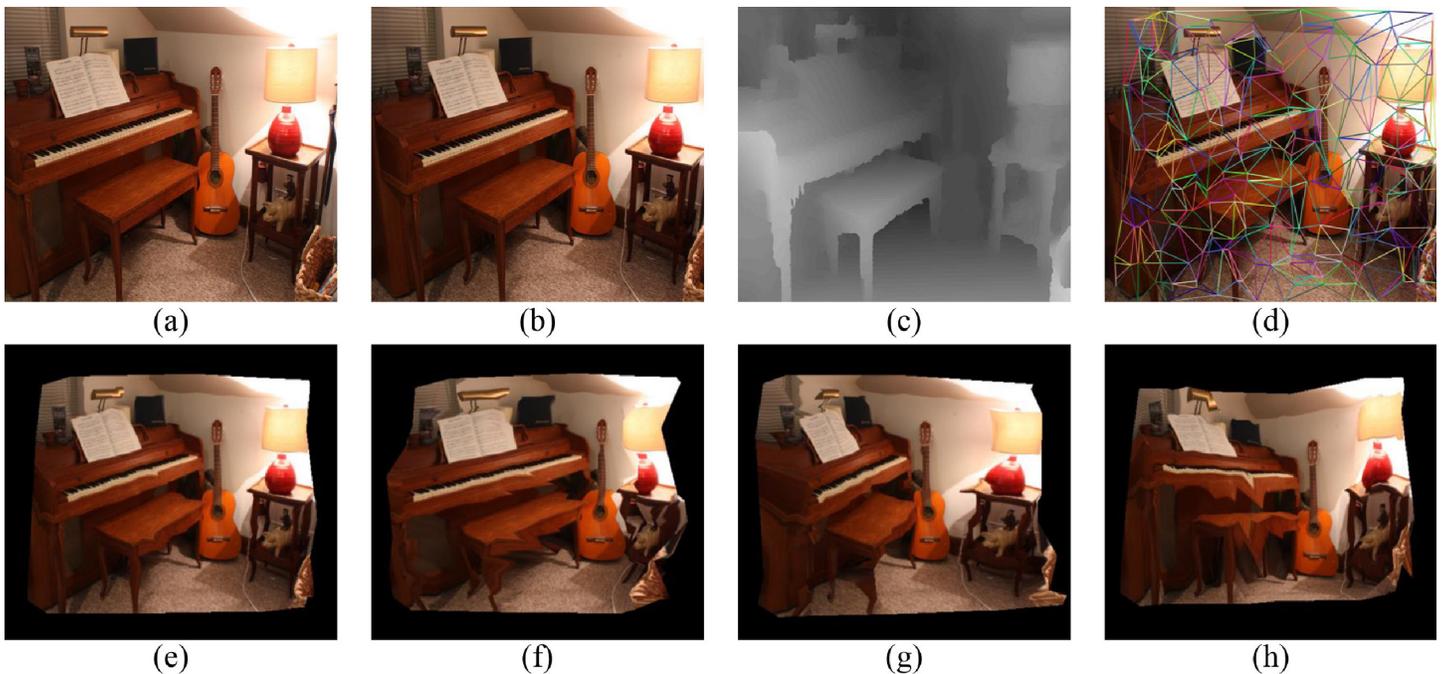
**Fig 8. 3D reconstruction experiment of scene Wood1.** (a) left image, (b) right image, (c) disparity image, (d) Delaunay triangulation, (e) view 1, (f) view 2, (g) view 3, (h) view 4. (a) and (b) are republished from [vision.middlebury.edu/stereo/] under a CC BY license, with permission from [D. Scharstein], original copyright [2002].

<https://doi.org/10.1371/journal.pone.0260466.g008>



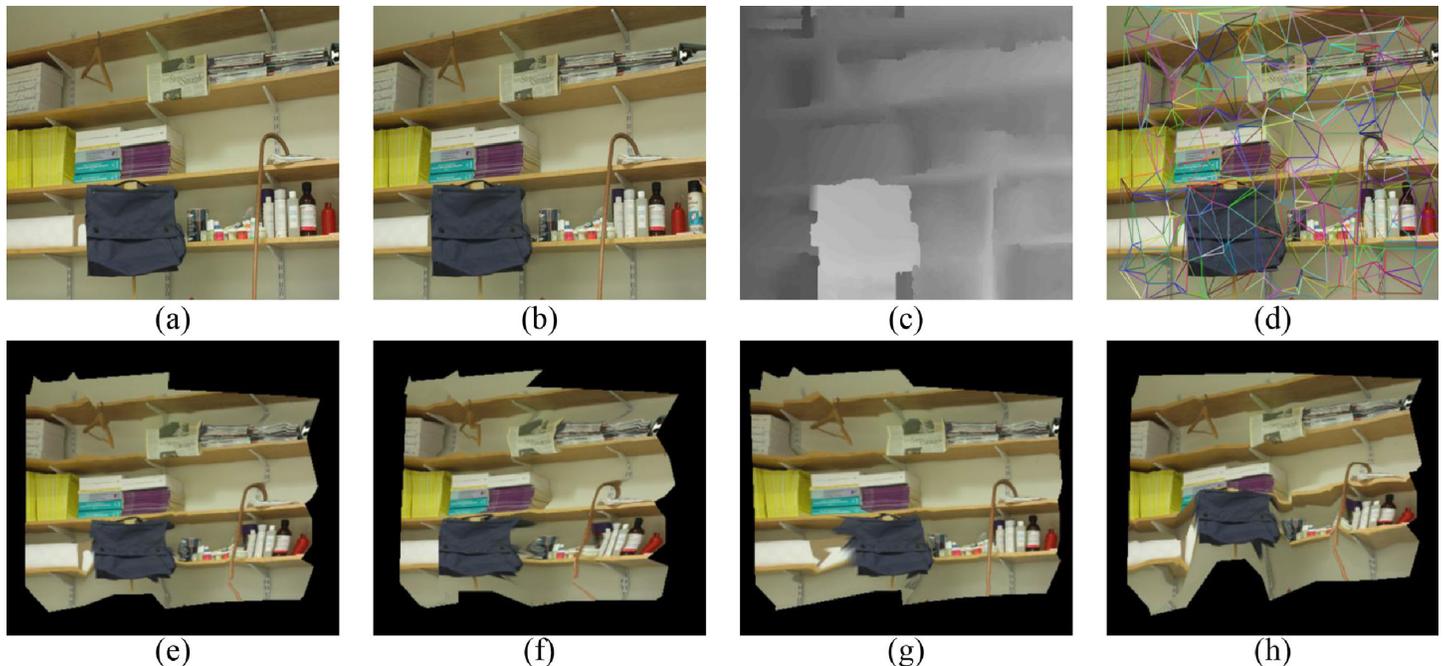
**Fig 9. 3D reconstruction experiment of scene Djembe.** (a) left image, (b) right image, (c) disparity image, (d) Delaunay triangulation, (e) view 1, (f) view 2, (g) view 3, (h) view 4. (a) and (b) are republished from [vision.middlebury.edu/stereo/] under a CC BY license, with permission from [D. Scharstein], original copyright [2002].

<https://doi.org/10.1371/journal.pone.0260466.g009>



**Fig 10. 3D reconstruction experiment of scene Piano.** (a) left image, (b) right image, (c) disparity image, (d) Delaunay triangulation, (e) view 1, (f) view 2, (g) view 3, (h) view 4. (a) and (b) are republished from [vision.middlebury.edu/stereo/] under a CC BY license, with permission from [D. Scharstein], original copyright [2002].

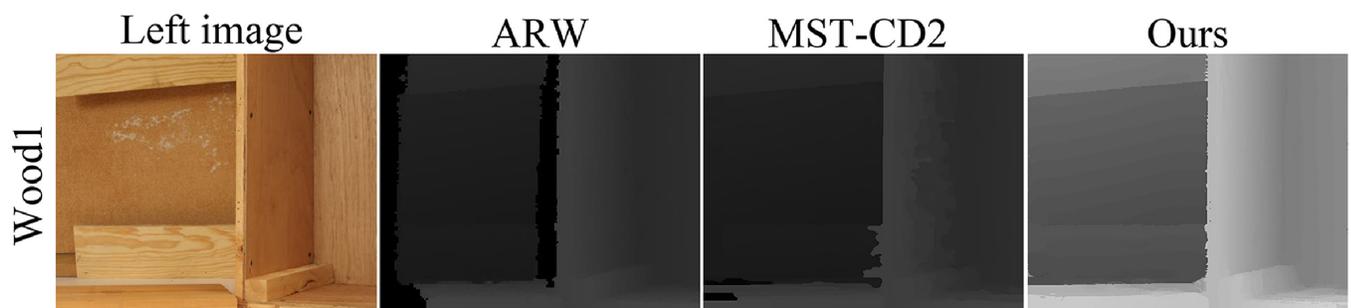
<https://doi.org/10.1371/journal.pone.0260466.g010>



**Fig 11. 3D reconstruction experiment of scene Shelves.** (a) left image, (b) right image, (c) disparity image, (d) Delaunay triangulation, (e) view 1, (f) view 2, (g) view 3, (h) view 4. (a) and (b) are republished from [vision.middlebury.edu/stereo/] under a CC BY license, with permission from [D. Scharstein], original copyright [2002].

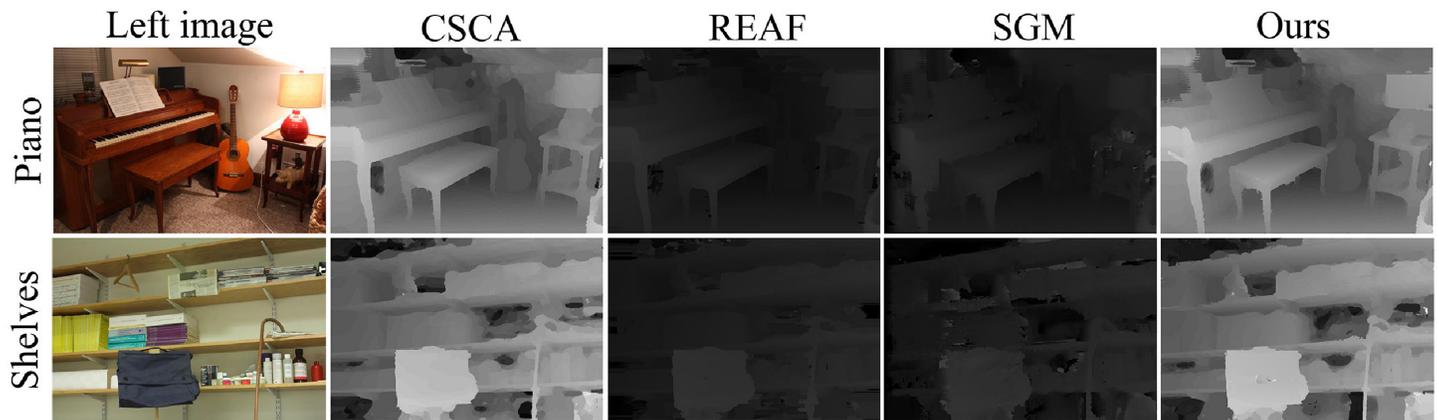
<https://doi.org/10.1371/journal.pone.0260466.g011>

The quantitative experimental results on Middlebury2 and Middlebury3 also show the superior performance of the proposed method. Three scenes (Cloth1, Cloth3 and Wood1) in the Middlebury2 dataset and three scenes (Djembe, Piano and Shelves) in the Middlebury3 dataset are used to conduct comparative experiments. We evaluate 3D reconstruction effects of the proposed method by calculating the average error distance between the point cloud generated by different methods and its ground truth. The ground truth of the point cloud is generated from the ground truth of the disparity map. Each scene in Middlebury2 and Middlebury3 contains more than 150,000 points and more than 350,000 points, respectively. Table 1 shows the experimental results. The data in Table 1 represent the average distance between each point and the ground truth in different scenes, in units of millimeters.



**Fig 12. Wood1 scene in the Middlebury2 dataset.** The left image in the figure is republished from [vision.middlebury.edu/stereo/] under a CC BY license, with permission from [D. Scharstein], original copyright [2002].

<https://doi.org/10.1371/journal.pone.0260466.g012>



**Fig 13. Piano scene and the Shelves scene in the Middlebury3 dataset.** The left image in the figure is republished from [vision.middlebury.edu/stereo/] under a CC BY license, with permission from [D. Scharstein], original copyright [2002].

<https://doi.org/10.1371/journal.pone.0260466.g013>

**Table 1. Comparison of 3D reconstruction errors for six scenes in Middlebury2 and Middlebury3 (measurement unit: mm).**

	Cloth1	Cloth3	Wood1	Djembe	Piano	Shelves
ARW [32]	<b>0.40</b>	<b>0.62</b>	1.11	0.83	2.54	1.66
CSCA [34]	1.69	1.12	0.71	1.12	2.46	6.59
MST-CD2 [33]	0.58	2.30	11.3	1.37	2.64	13.9
REAF [35]	31.0	1.95	22.5	6.26	11.1	21.2
SGM [12]	16.8	13.7	17.9	8.53	11.9	20.5
Proposed Approach	0.68	1.31	<b>0.59</b>	<b>0.00</b>	<b>1.82</b>	<b>1.47</b>

<https://doi.org/10.1371/journal.pone.0260466.t001>

## 4.2 KITTI dataset

The proposed method was evaluated on KITTI 2012 and KITTI 2015 benchmarks, and compared with other methods. Since the scene of the KITTI dataset is an outdoor environment, texture mapping was not implemented. Implementing texture mapping in places where the depth is not continuous will distort the reconstruction results. Fig 14 shows the three scenes in KITTI 2012, which can be downloaded via <https://osf.io/7srj4>. Table 2 is the evaluation result of the proposed method on the KITTI 2012 benchmark. Compared with other semiglobal stereo matching methods, the proposed second-order semiglobal stereo matching method has advantages in various evaluation indicators. Fig 15 shows the three scenes in KITTI 2015, which can be downloaded via <https://osf.io/9kzma>. Table 3 shows the evaluation results of the proposed method on the KITTI 2015 benchmark.



**Fig 14. 3D reconstruction experiment of scenes in KITTI 2012.** It can be downloaded via <https://osf.io/7srj4>.

<https://doi.org/10.1371/journal.pone.0260466.g014>

Table 2. Depth information of 3D reconstruction in the KITTI 2012 benchmark.

	Out-Noc(%)	Out-All(%)	Avg-Noc(px)	Avg-All(px)
ARW [32]	5.2	6.87	1.2	1.5
SGM [12]	5.76	7.0	1.2	1.3
iSGM [36]	5.11	7.15	1.2	2.1
wSGM [37]	4.97	6.18	1.3	1.6
pSGM [38]	4.68	6.13	1.0	1.4
Proposed Approach	<b>3.45</b>	<b>4.22</b>	<b>0.9</b>	<b>1.0</b>

<https://doi.org/10.1371/journal.pone.0260466.t002>

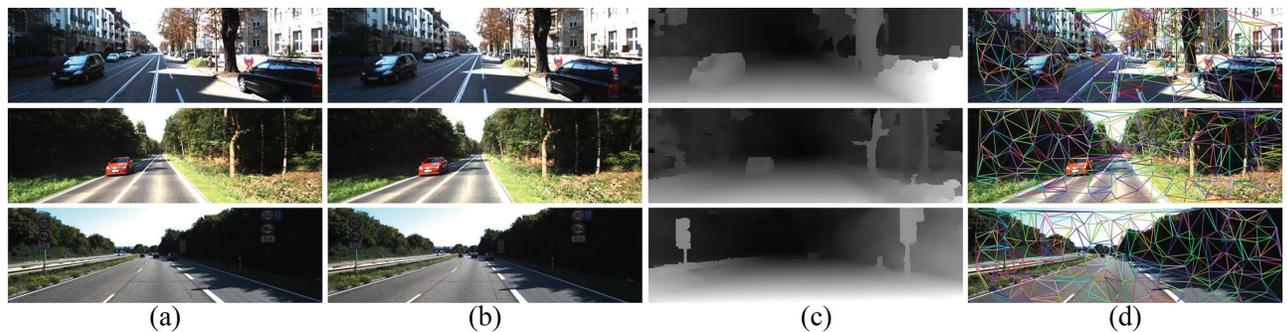


Fig 15. 3D reconstruction experiment of scenes in KITTI 2015. It can be downloaded via <https://osf.io/9kzma>.

<https://doi.org/10.1371/journal.pone.0260466.g015>

Table 3. Depth information of 3D reconstruction in the KITTI 2015 benchmark.

	D1-bg(%)	D1-fg(%)	D1-all(%)
REAF [35]	8.43	18.51	10.11
SGM [12]	5.15	15.29	6.84
pSGM [38]	4.84	11.64	5.97
SPS+FF++ [39]	5.47	12.19	6.59
Proposed Approach	<b>4.08</b>	<b>11.07</b>	<b>5.24</b>

<https://doi.org/10.1371/journal.pone.0260466.t003>

### 4.3 Results of Delaunay triangulation based on fast point positioning

The proposed Delaunay triangulation based on fast point positioning was evaluated on the Middlebury dataset and the KITTI dataset. Table 4 shows the experimental results. The original Delaunay triangulation based on point-by-point insertion uses the implementation in OpenCV, which is denoted as Ori-D in Table 4.

In Table 4,  $F_{avg}$  represents the average number of feature points of each image, which is the average size of the point set processed by Delaunay triangulation. K-12 represents the KITTI 2012 dataset, which contains 388 images; K-15 represents the KITTI 2015 dataset, which

Table 4. Results of Delaunay triangulation based on fast point positioning (measurement unit: ms).

	K-12( $F_{avg}$ :115)	K-15( $F_{avg}$ :115)	M-2( $F_{avg}$ :95)	M-3( $F_{avg}$ :124)
Ori-D	2.59	2.45	1.87	3.26
Fast-D	<b>0.08</b>	<b>0.12</b>	<b>0.13</b>	<b>0.17</b>

<https://doi.org/10.1371/journal.pone.0260466.t004>

contains 398 images; M-2 represents the Middlebury2 dataset, which contains 31 images; M-3 represents the Middlebury3 dataset, which contains 23 images. Fast-D in Table 4 represents the proposed approach. Compared with the original Delaunay triangulation based on point-by-point insertion, the proposed fast point positioning method is significantly improved in terms of time efficiency. When processing a large amount of data (for example, 1000 points), the proposed method processes the data within 1 ms, but the Delaunay triangulation implementation in OpenCV cannot handle the data amount.

## 5 Conclusions

3D reconstruction based on binocular vision combines camera parameters and disparity maps to calculate all of the point coordinates in a scene and perform a surface reconstruction of the point cloud. In this paper, a 3D reconstruction method based on binocular stereo matching and fast point positioning of Delaunay triangulation is proposed to improve the accuracy and efficiency of 3D reconstruction. First, a second-order semiglobal stereo matching method combining multiple matching costs is used to obtain a more robust disparity map. Second, the 3D point cloud is obtained with camera parameters and disparity maps, which is smoothed by guided filtering to remove the noise points and retain details. Finally, the Delaunay triangulation method based on fast point positioning is used to reconstruct the surface of the point cloud and improve the visibility of the 3D model.

The Middlebury and KITTI datasets are used to evaluate the performance of the proposed method. The proposed method can obtain accurate depth information and reconstruct the surface of the point cloud to obtain a 3D model. The depth information obtained thus far only comes from visible light images. In future work, we will combine the accurate 3D information of LIDAR and the guidance of visible light images to achieve high-precision 3D information perception. To further improve the authenticity of the 3D model, we will also study triangulation in cases with shadows and occlusions in the future.

## Supporting information

### S1 Dataset.

(EML)

### S2 Dataset.

(EML)

### S3 Dataset.

(EML)

### S1 File.

(PDF)

### S2 File.

(ZIP)

## Author Contributions

**Conceptualization:** Qingwu Li.

**Formal analysis:** Kaixiang Liu, Jinyan Ni.

**Investigation:** Qingwu Li.

**Methodology:** Kaixiang Liu, Jinyan Ni.

**Project administration:** Yongbing Xu, Qingwu Li.

**Software:** Jinyan Ni.

**Supervision:** Yongbing Xu, Qingwu Li.

**Writing – original draft:** Kaixiang Liu, Jinyan Ni.

**Writing – review & editing:** Kaixiang Liu, Qingwu Li.

## References

1. Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*. 2002; 47(1):7–42. <https://doi.org/10.1023/A:1014573219977>
2. Zhang K, Lu J, Lafruit G. Cross-based local stereo matching using orthogonal integral images. *IEEE transactions on circuits and systems for video technology*. 2009; 19(7):1073–1079. <https://doi.org/10.1109/TCSVT.2009.2020478>
3. Li L, Yu X, Zhang S, Zhao X, Zhang L. 3D cost aggregation with multiple minimum spanning trees for stereo matching. *Applied optics*. 2017; 56(12):3411–3420. <https://doi.org/10.1364/AO.56.003411> PMID: 28430267
4. Lu H, Xu H, Zhang L, Ma Y, Zhao Y. Cascaded multi-scale and multi-dimension convolutional neural network for stereo matching. In: 2018 IEEE Visual Communications and Image Processing (VCIP). IEEE; 2018. p. 1–4.
5. Yang J, Wang H, Ding Z, Lv Z, Wei W, Song H. Local stereo matching based on support weight with motion flow for dynamic scene. *IEEE Access*. 2016; 4:4840–4847. <https://doi.org/10.1109/ACCESS.2016.2601069>
6. Veksler O. Stereo correspondence by dynamic programming on a tree. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). vol. 2. IEEE; 2005. p. 384–390.
7. Suhr JK, Jung HG. Dense stereo-based robust vertical road profile estimation using Hough transform and dynamic programming. *IEEE Transactions on Intelligent Transportation Systems*. 2014; 16(3):1528–1536. <https://doi.org/10.1109/TITS.2014.2369002>
8. Kolmogorov V, Zabih R. Computing visual correspondence with occlusions using graph cuts. In: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. vol. 2. IEEE; 2001. p. 508–515.
9. Tani T, Matsushita Y, Naemura T. Graph cut based continuous stereo matching using locally shared labels. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2014. p. 1613–1620.
10. Sun J, Zheng NN, Shum HY. Stereo matching using belief propagation. *IEEE Transactions on pattern analysis and machine intelligence*. 2003; 25(7):787–800. <https://doi.org/10.1109/TPAMI.2003.1206509>
11. Besse F, Rother C, Fitzgibbon A, Kautz J. Pmbp: Patchmatch belief propagation for correspondence field estimation. *International Journal of Computer Vision*. 2014; 110(1):2–13. <https://doi.org/10.1007/s11263-013-0653-9>
12. Hirschmuller H. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on pattern analysis and machine intelligence*. 2007; 30(2):328–341. <https://doi.org/10.1109/TPAMI.2007.1166>
13. Tomasi C, Manduchi R. Bilateral filtering for gray and color images. In: Sixth international conference on computer vision (IEEE Cat. No. 98CH36271). IEEE; 1998. p. 839–846.
14. He K, Sun J, Tang X. Guided image filtering. In: European conference on computer vision. Springer; 2010. p. 1–14.
15. Hoppe H, DeRose T, Duchamp T, McDonald J, Stuetzle W. Surface reconstruction from unorganized points. In: Proceedings of the 19th annual conference on Computer graphics and interactive techniques; 1992. p. 71–78.
16. Kazhdan M. Reconstruction of solid models from oriented point sets. In: Proceedings of the third Eurographics symposium on Geometry processing; 2005. p. 73–es.
17. Delaunay B. Sur la sphère vide. A la mémoire de Georges Vorono. *Bulletin de l'Académie des Sciences de l'URSS Classe des sciences mathématiques et naturelles*;6:793.
18. Wang L, Yang R. Global stereo matching leveraged by sparse ground control points. In: CVPR 2011. IEEE; 2011. p. 3033–3040.

19. Sarkis M, Diepold K. Sparse stereo matching using belief propagation. In: 2008 15th IEEE International Conference on Image Processing. IEEE; 2008. p. 1780–1783.
20. Schauwecker K, Klette R, Zell A. A new feature detector and stereo matching method for accurate high-performance sparse stereo matching. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE; 2012. p. 5171–5176.
21. Kim JC, Lee KM, Choi BT, Lee SU. A dense stereo matching using two-pass dynamic programming with generalized ground control points. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). vol. 2. IEEE; 2005. p. 1075–1082.
22. Stentoumis C, Grammatikopoulos L, Kalisperakis I, Karras G. On accurate dense stereo-matching using a local adaptive multi-cost approach. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2014; 91:29–49. <https://doi.org/10.1016/j.isprsjprs.2014.02.006>
23. Woodford O, Torr P, Reid I, Fitzgibbon A. Global stereo reconstruction under second-order smoothness priors. *IEEE transactions on pattern analysis and machine intelligence*. 2009; 31(12):2115–2128. <https://doi.org/10.1109/TPAMI.2009.131> PMID: 19834135
24. Shamos MI, Hoey D. Closest-point problems. In: 16th Annual Symposium on Foundations of Computer Science (sfcs 1975). IEEE; 1975. p. 151–162.
25. Lewis B, Robinson J. Triangulation of planar regions with applications. *The Computer Journal*. 1978; 21(4):324–332. <https://doi.org/10.1093/comjnl/21.4.324>
26. Shewchuk JR. Triangle: Engineering a 2D quality mesh generator and Delaunay triangulator. In: Workshop on Applied Computational Geometry. Springer; 1996. p. 203–222.
27. Green PJ, Sibson R. Computing Dirichlet tessellations in the plane. *The computer journal*. 1978; 21(2):168–173. <https://doi.org/10.1093/comjnl/21.2.168>
28. Bowyer A. Computing dirichlet tessellations. *The computer journal*. 1981; 24(2):162–166. <https://doi.org/10.1093/comjnl/24.2.162>
29. Scharstein D, Szeliski R. High-accuracy stereo depth maps using structured light. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. vol. 1. IEEE; 2003. p. I–I.
30. Scharstein D, Hirschmüller H, Kitajima Y, Krathwohl G, Nešić N, Wang X, et al. High-resolution stereo datasets with subpixel-accurate ground truth. In: German conference on pattern recognition. Springer; 2014. p. 31–42.
31. Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2012. p. 3354–3361.
32. Lee S, Lee JH, Lim J, Suh IH. Robust stereo matching using adaptive random walk with restart algorithm. *Image and Vision Computing*. 2015; 37:1–11. <https://doi.org/10.1016/j.imavis.2015.01.003>
33. Yao P, Zhang H, Xue Y, Zhou M, Xu G, Gao Z. Iterative color-depth MST cost aggregation for stereo matching. In: 2016 IEEE International Conference on Multimedia and Expo (ICME). IEEE; 2016. p. 1–6.
34. Zhang K, Fang Y, Min D, Sun L, Yang S, Yan S, et al. Cross-scale cost aggregation for stereo matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2014. p. 1590–1597.
35. Cigla C. Recursive edge-aware filters for stereo matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2015. p. 27–34.
36. Hermann S, Klette R. Iterative semi-global matching for robust driver assistance systems. In: Asian Conference on Computer Vision. Springer; 2012. p. 465–478.
37. Spangenberg R, Langner T, Rojas R. Weighted semi-global matching and center-symmetric census transform for robust driver assistance. In: International Conference on Computer Analysis of Images and Patterns. Springer; 2013. p. 34–41.
38. Lee Y, Park MG, Hwang Y, Shin Y, Kyung CM. Memory-efficient parametric semiglobal matching. *IEEE Signal Processing Letters*. 2017; 25(2):194–198. <https://doi.org/10.1109/LSP.2017.2778306>
39. Schuster R, Wasenmüller O, Stricker D. Dense scene flow from stereo disparity and optical flow. *arXiv preprint arXiv:180810146*. 2018;