



# A novel binocular vision system for accurate 3-D reconstruction in large-scale scene based on improved calibration and stereo matching methods

Daolei Wang<sup>2</sup> · Hao Sun<sup>1</sup> · Wu Lu<sup>2</sup> · Wenbin Zhao<sup>2</sup> · Yiteng Liu<sup>2</sup> · Pingping Chai<sup>2</sup> · Yang Han<sup>2</sup>

Received: 20 November 2020 / Revised: 8 March 2021 / Accepted: 10 March 2022 /

Published online: 26 March 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

The 3D reconstruction technology based on stereo vision directly acquires the 3D model of the object through two 2D images. The reconstruction is recovered highly automated. It does not require any prior information and special hardware support. However, for large outdoor scenes, the existing 3D reconstruction technology based on stereo vision often has detailed information loss and data scattering, which makes the reconstruction result less accurate. For this problem, a novel binocular vision system for 3D reconstruction in large-scale scene is proposed. This system uses the calibration rods to perform the calibration calculation based on the polar line correction, and then it utilizes the weighted least squares filter(WLSF) to denoise and smooth the depth map, finally, the point cloud is reconstructed. The results of experiment show that compared with the traditional stereo vision system, the calibration results of new system is more accurate and the calibration space is expanded. The depth map is smoother and less noisy. The system can reconstruct the 3D point cloud of large scene more stably and accurately, has high practical value.

**Keywords** Stereo vision · Camera calibration · Polar correction · 3-D reconstruction

---

✉ Hao Sun  
hallesun@stu.xjtu.edu.cn

✉ Wenbin Zhao  
zhaowenbin@shiep.edu.cn

<sup>1</sup> School of Energy and Power Engineering, Xi'an Jiaotong University, Xi'an 710049 Shaanxi, China

<sup>2</sup> College of Energy and Mechanical Engineering, Shanghai University of Electric Power, Shanghai 200090, China

## 1 Introduction

With the rapid development of technologies such as augmented reality and virtual reality, the demand for geometric 3D models is also rising. 3D reconstruction technology is a very important research topic in the field of modeling, and its innovation and optimization become very important. Today's computer hardware and software levels have been rapidly improved. More and more excellent and classic 3D reconstruction algorithms have been successfully applied in various fields, and the reconstruction and rendering accuracy of large-scale, high-complexity 3D scenes. The real-time performance is also constantly improving, and the difficulty is greatly reduced. In the early 3D reconstruction, modeling tools such as AutoCAD and 3DMAX were mainly used, but for large-scale scenes, it was impossible to solve the problem that the workload was too large and the reconstructed scene lacked reality. Therefore, a three-dimensional reconstruction method using a laser scanner has been proposed. The reconstruction accuracy of this method is relatively high, and the reconstructed three-dimensional model is more accurate and close to reality, but the scanning time required for this method is extremely long, and one scan is required [18]. Larger scale scenarios often take hours or even days. In addition, optical devices such as laser scanners are expensive and are not suitable for general applications, and in some special fields, such as cultural relic protection applications, it is often impossible to use lasers to illuminate objects to be reconstructed [10, 16]. Therefore, in comparison, the 3D reconstruction based on binocular stereo vision uses the photos of the real scene as the main data source, which can truly reflect the real scene under the premise of ensuring that the reconstructed model has good precision, compared with the laser scanning. The method has great advantages [5].

Although the mathematical theory based on the binocular stereoscopic 3D reconstruction method is perfect, it has great advantages compared to other methods, and binocular stereo vision has a mature and reliable implementation, but the existing embodiment is for a large-scale scene. There are few applications for 3D reconstruction, and there are still many defects and shortcomings in detail [4, 7, 14]. There are still many challenges, such as insufficient calibration space and insufficient matching accuracy in the process of large-scale scene reconstruction [15]. In view of the shortcomings of the above-mentioned 3-D reconstruction methods based on binocular stereo vision, this paper optimizes and improves the original 3-D reconstruction methods [13]. Firstly, when the camera calibration is performed, the calibration pole is selected as the calibration mark. This method overcomes the problem of too small checkerboards. Using calibration rods can expand the space that can be calibrated, and is cheaper than using checkerboards. Then, in order to ensure the speed and precision of the stereo matching, based on the traditional matching cost function, the weighted least squares filter(WLSF) method is used to optimize the obtained initial depth map, which can greatly improve the matching accuracy under the premise of ensuring matching efficiency [3]. Finally, the reconstruction of the point cloud is carried out, and the three-dimensional coordinates of the point cloud are calculated by the similar triangle method, and finally the dense point cloud of the large-scale scene is obtained [12, 19].

In detail, this paper provides the following key insights: Compared to reconstructing small objects, large-scale scene reconstruction based on binocular vision requires some improvement in the calibration and matching algorithms to prevent outliers from occurring in the resulting 3D model. For the binocular vision algorithm, smoothing the accurate depth map can effectively reduce the errors contained in the reconstructed 3D model.

## 2 Methodology

### 2.1 Technical approach for accurate 3-D reconstruction process

The Fig. 1 shows the flowchart of large-scale scene reconstruction. The improved calibration method was proposed to large-scale scene which used the calibration rod instead of the traditional checkerboard. The weight least squares filtering was used to refine the depth map. Finally, the points cloud was been recovered after camera calibration and stereo matching steps. Recovery the dense point clouds is a core issue in binocular vision algorithms, and a number of scholars have study about this topic. In the following text, we will discuss the key steps of the algorithm of points clouds reconstruction, which are the camera calibration and stereo matching algorithms.

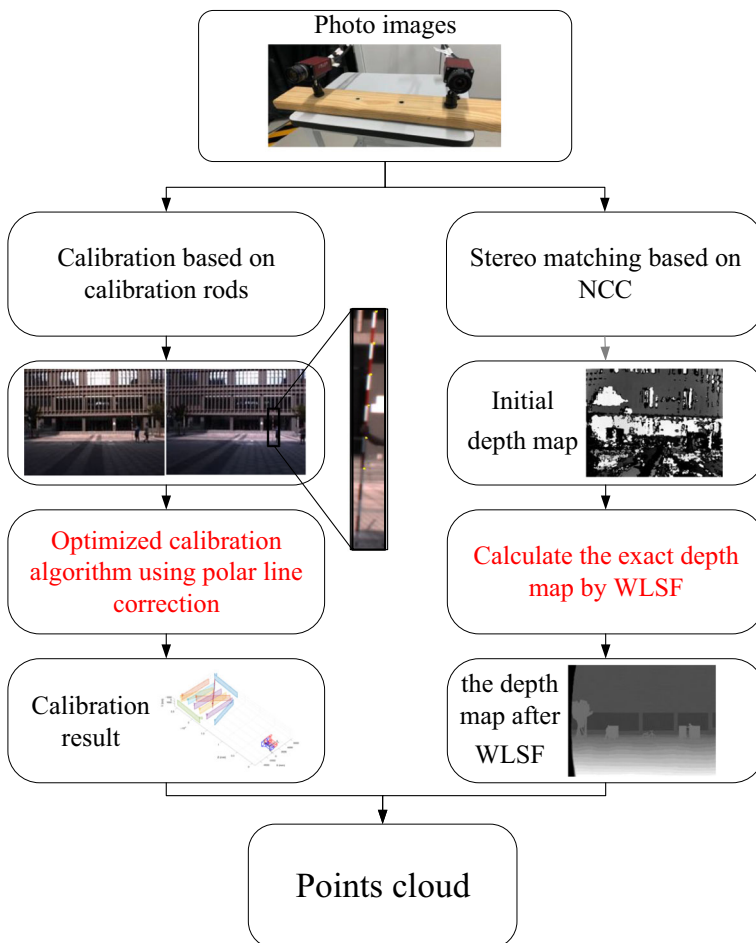


Fig. 1 Flowchart of large-scale scene reconstruction

## 2.2 Improved calibration algorithm using POLAR correction

There have many camera calibration methods such as Direct linear transformation(DLT) calibration, Radial alignment constraint(RAC) calibration and Zhang's calibration and so on. Zhang's calibration method is widely used in practical calibration experiments because of its simple calibration process and high accuracy. The chessboard grid is used to calibrate large-scale scenes, as shown in the Fig. 2.

In calibration experiments, we used standard checkerboard to calibrate, however, we found that was difficult to cover the whole calibration space for checkerboard. This is because the long baseline of binocular camera used in large-scale scenes. Because of the limited resolution of the camera, it is difficult to recognize corners when the checkerboard is far from the camera in large-scale scenes. Thus, it is difficult to accomplish calibration experiments in large-scale scenes using standard chessboard grids.

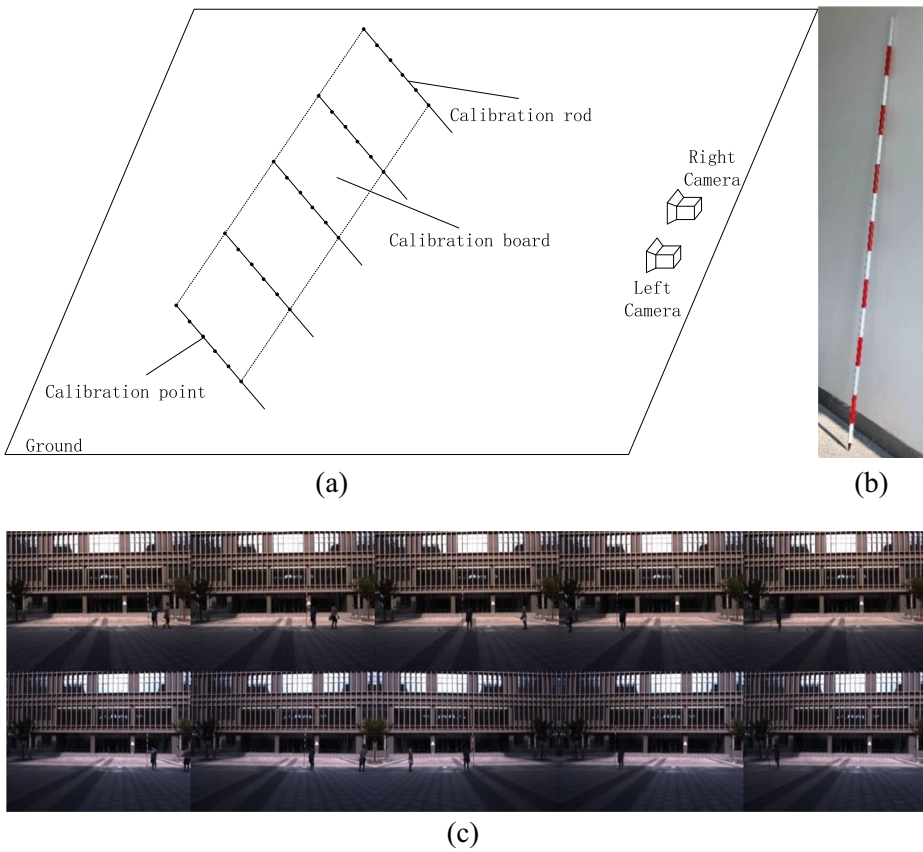
For limitation of the camera calibration method using chessboard grid, the calibration method using the calibration rod was proposed to calibrate the large-scale scene, which is a calibration algorithm based on the flat calibration method. Aiming at the inconsistency of the left and right camera focal lengths in the calibration process, Yang F et al [17] designed a calibration rod calibration algorithm to optimize the focal length difference based on Zhang's calibration method. The calibration rod is placed vertically on the ground according to a certain arrangement before the calibration image acquisition, and the fixed scale is marked on the calibration rod as the calibration point, then we can synthesize a calibration plane by using the calibration rods of several specific positions [9]. The camera parameters (the internal and external parameters of the camera) are then calculated by the relationship between the world coordinates of the calibration point and the coordinates of the picture [6]. Figure 3 is an example of the arrangement and calibration space of the calibration rod. The calibration of stereo vision system can overcome the shortcomings of the traditional checkerboard grid, which is difficult to apply to large-scale scenes, and expand the calibration space of the flat panel calibration method to meet large-scale scene.

In the calibration rod calibration method, the calibration accuracy is improved by adjusting the focal length of the calibration image. In the ideal binocular camera model, the left and right cameras with the same internal parameters are placed at the same baseline position, and the left and right camera optical axes are completely parallel. In practical applications, it is difficult for binocular cameras to have the same internal parameters due to various factors. In the calibration rod calibration algorithm, the problem of different focal lengths of the left and right cameras has been corrected. Correcting the image can eliminate the error caused by the non-parallel of the optical axis of the camera to a certain extent, which can improve the calibration accuracy and speed up the calibration speed [1].

When using the binocular industrial camera to obtain pictures, due to the distortion and aberration of the camera's own imaging, large errors will occur when performing stereo



Fig. 2 Checkerboard calibration picture in a large-scale scene



**Fig. 3** a Calibration rod and camera position diagram b Calibration rod c Calibration image

matching, and mismatching will also occur. In this paper, the polar line correction of the image is used to correct the error caused by the non-parallel of the optical axis of the camera.

The correction process for images mainly consists of two steps:

- (1) The homography matrices  $H_L$ ,  $H_R$  corresponding to the left and right images are calculated by using the internal and external parameters of the camera obtained by the initial calibration.
- (2) Warp the image and modify the camera projection matrix. The image is warped with a homography matrix, and then the camera projection matrix is modified to  $M_L^* = H_L M_L$ ,  $M_R^* = H_R M_R$ . Calibrate the camera with a homography matrix:

$$M_L^* = H_L M_L = H_L K_L R_L [I - C_L] \tag{1}$$

$$M_R^* = H_R M_R = H_R K_R R_R [I - C_R] \tag{2}$$

$K_L, K_R$  are the internal parameter matrices of the left and right cameras,  $R_L, R_R$  are the rotation matrix of the left and right cameras,  $C_L, C_R$  is the optical center of the left and right cameras. Let  $e_L, e_R$  be the poles of the left and right graphs,  $l_L, l_R$  be the polar lines, and  $u_L, u_R$  be the projections of the points in the range on the imaging plane. Let  $F^*$  be the basic matrix of the corrected image and  $\lambda \neq 0$ . The necessary conditions for the polar line to coincide with the lines in the two images are:

$$l_R^* = e_R^* \times u_R^* = \lambda F^* u_L^* \tag{3}$$

$$[1, 0, 0]^T \times [u', v, 1]^T = [1, 0, 0]^T \times [u + d, v, 1]^T = \lambda F^* [u, v, 1]^T \tag{4}$$

among them:

$$F^* \approx \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \tag{5}$$

The correcting homography matrix is not unique. In order to select the best correcting homography matrix, the following derivation is made.

- (1) Move the poles in the two images to infinity

Let  $e_L = [e_1, e_2, 1]^T$  be the pole in the graph and  $e_1^2 + e_2^2 \neq 0$ . While rotating the polar line  $e_L$  onto the u-axis, this pole maps to  $e^* \approx [1, 0, 0]^T$ , and the corresponding projection is:

$$\widehat{H}_L \approx \begin{bmatrix} e_1 & e_2 & 0 \\ -e_2 & e_1 & 0 \\ -e_1 & -e_2 & e_1^2 + e_2^2 \end{bmatrix} \tag{6}$$

- (2) Unified polar line

Since  $e_R^* = [1, 0, 0]^T$  is the left zero of  $\widehat{F}$  and the right zero space, the modified basic matrix becomes:

$$\widehat{F} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \alpha & \beta \\ 0 & \gamma & \delta \end{bmatrix} \tag{7}$$

The basic correction homograph  $\overline{H}_L, \overline{H}_R$  is selected such that  $\alpha = \delta = 0$  and  $\beta = -\gamma$ .

$$\overline{H}_L = H_S \widehat{H}_L, \overline{H}_R = \widehat{H}_R, \text{ among them, } H_S = \begin{bmatrix} \alpha\delta - \beta\gamma & 0 & 0 \\ 0 & -\gamma & -\delta \\ 0 & \alpha & \beta. \end{bmatrix}$$

such

$$F^* = \left(\widehat{H}_R\right)^{-T} F \left(H_S \widehat{H}_L\right)^{-1} \tag{8}$$

(3) Select a pair of optimal homography matrices

Let  $\overline{H}_L, \overline{H}_R$  be the basis for correcting the homography matrix.  $\overline{H}_L, \overline{H}_R$  are also corrected homography matrices, which obey the equation  $H_R F^* H_L^T = \lambda F^*, \lambda \neq 0$ , to ensure that the image remains in the corrected state.

The intrinsic parameters of  $\overline{H}_L, \overline{H}_R$  facilitate understanding of the meaning of the free parameters in the correct homography class:

$$H_L = \begin{bmatrix} l_1 & l_2 & l_3 \\ 0 & s & u_0 \\ 0 & q & 1 \end{bmatrix} \overline{H}_L, H_R = \begin{bmatrix} r_1 & r_2 & r_3 \\ 0 & s & u_0 \\ 0 & q & 1 \end{bmatrix} \overline{H}_R \tag{9}$$

Where  $s \neq 0$  is the common vertical scale;  $u_0$  is the common vertical offset,  $l_1, r_1$  are the left and right distortions,  $l_2, r_2$  are the left and right horizontal scales,  $l_3, r_3$  are the left and right horizontal offsets And  $q$  is the common projection distortion.

This article uses the polar-corrected calibration image to recalibrate to make the calibration parameters more accurate. A set of calibration images taken by the left and right cameras are shown in the Fig. 4, and the images after the polar line correction are as shown.

### 2.3 The exact depth map obtained by weighted least square filtering(WLSF) algorithm

After completing the camera calibration to obtain the camera’s internal and external parameters, we need to perform stereo matching on the left and right views of the reconstructed object to obtain the depth map [2, 8]. The local stereo matching algorithm is used to estimate the parallax in the local range, so this algorithm is called a support region based method. In the energy function of the local stereo matching algorithm, there are only data items and no smoothing terms. Therefore, the computational complexity is low, so most real-time stereo matching algorithms belong to local stereo matching, but the disadvantage is that the local stereo matching algorithm does not match the accuracy. In general, the stereo matching algorithm using local matching is far ahead of the global matching algorithm in speed, but the accuracy is slightly insufficient. After considering the speed and accuracy of reconstruction, this paper chooses to use the stereo matching algorithm based on normalized cross-correlation algorithm for initial matching, and then corrects and optimizes the initial matching

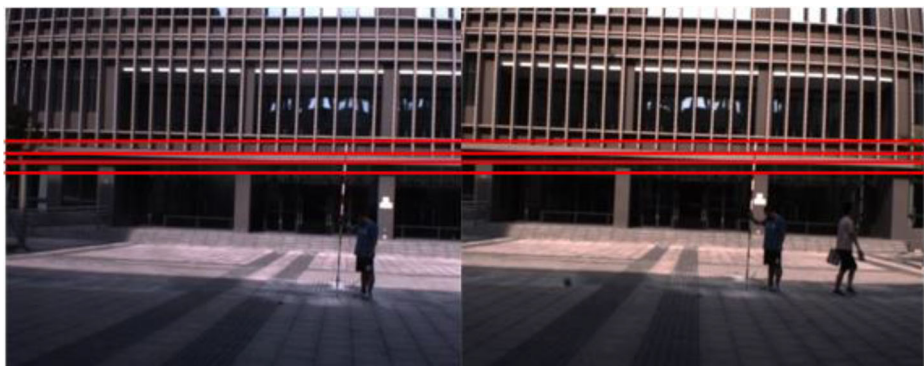


Fig. 4 Left and right polar-corrected calibration images

result by weighted least squares method. Under the premise that the accuracy can meet the requirements, this method can greatly speed up the stereo matching [11].

The matching cost function is to calculate the similarity between the corresponding points of the left and right images by some similarity measure. The cost function of the normalized cross-correlation number used in this paper is:

$$\text{NCC}(p, d) = \frac{\sum_{(x,y) \in W_p} \left( I_1(x, y) - \bar{I}_1(p_x, p_y) \right) \cdot \left( I_2(x, y) - \bar{I}_2(p_x + d, p_y) \right)}{\sqrt{\sum_{(x,y) \in W_p} \left( I_1(x, y) - \bar{I}_1(p_x, p_y) \right)^2 \cdot \sum_{(x,y) \in W_p} \left( I_2(x + d, y) - \bar{I}_2(p_x + d, p_y) \right)^2}} \quad (10)$$

Where  $\text{NCC}(p, d)$  is a measure of similarity. The closer this value is to 1, the greater the similarity between the two matching windows. The  $p$ -point represents the pixel coordinates to be matched  $(p_x, p_y)$  of the left image  $I_1$ ,  $d$  represents the distance in the horizontal direction between the pixel being searched and the point  $p$  in the right image  $I_2$ , and  $W_p$  represents the matching window centered on the  $p$ -point.

The Fig. 5 shows the initial depth map. It can be seen that there are a large number of mismatched points in the image, and the edge segmentation of the object is also very blurred.

Due to the existence of a large number of mismatched points, the reconstruction result will generate a large number of errors, and it is difficult to meet the goal of accurate reconstruction of large-scale scenes. Further noise reduction and smoothing of the initial depth map is required. The traditional filtering method is easy to lose the edge details after processing, and it is difficult to obtain a good denoising effect. Therefore, the WLSF algorithm is proposed to optimize the initial depth map. This method has strong ability to preserve edges and denoise.

In this paper, the edge of the object in the depth map is enhanced by the method of WLSF, and the rest is smoothed. The WLSF method is filtered to make the processed image as much as possible on the basis of maintaining the edge. The ground is similar to the original image. The mathematical expression of the WLSF is



Fig. 5 Initial depth map based on NCC



$$\sum_p \left\{ (O_p - I_p)^2 + \lambda \left[ a_{xp}(I) \left( \frac{\partial O}{\partial x} \right)_p^2 + a_{yp}(I) \left( \frac{\partial O}{\partial y} \right)_p^2 \right] \right\} \tag{11}$$

Where  $O$  represents the filtered output image.  $I$  represents the initial image of the input,  $I = t_b(x)$ .  $P$  denotes a pixel located at a corresponding position.  $a_{xp}$  and  $a_{yp}$  are smooth weights.  $(O_p - I_p)^2$  is a data item.  $a_{xp}(I) \left( \frac{\partial O}{\partial x} \right)_p^2 + a_{yp}(I) \left( \frac{\partial O}{\partial y} \right)_p^2$  is a regular term.  $\lambda$  is the parameter that balances these two settings, this time is set to 0.35.

In order to facilitate the calculation, the equation is written in a matrix form, set to  $b$  before the optimization process, and set to  $f$  after the optimization process. Then the equation is rewritten as:

$$(f - b)^T (f - b) + \lambda (f^T D_x^T A_x D_x f + f^T D_y^T A_y D_y f) \tag{12}$$

Where  $A_x$  and  $A_y$  are the diagonal matrices of the smoothing terms  $a_{xp}$  and  $a_{yp}$ ,  $D_x$  and  $D_y$  are the forward and backward differential matrices. When the matrix representation of the input initial image  $I$  is set to  $R$ ,  $f$  is satisfied when the minimum value is obtained by the above formula:

$$(R + \lambda L_g) f = b \tag{13}$$

Where  $L_g = D_x^T A_x D_x + D_y^T A_y D_y$ ;  $a_{xp}(g) = \left[ \left| \frac{\partial l}{\partial x}(p) \right|^\alpha + \varepsilon \right]^{-1}$ ,  $a_{yp}(g) = \left[ \left| \frac{\partial l}{\partial y}(p) \right|^\alpha + \varepsilon \right]^{-1}$ ,  $l$  is the logarithm of the luminance channel image of the input initial image  $I$ , and  $\alpha$  is an amplification factor with a value range of  $[1, 2]$ , this time is set to 1.8,  $\varepsilon$  is the gain term, used to prevent 0 from appearing, this time is set to 0.0001

From this equation, the matrix form  $f$  of the filtered image  $O$  is obtained:

$$f = F_\lambda(q) = (R + \lambda L_g)^{-1} q \tag{14}$$

In this formula,  $F_\lambda(q) \approx (R + \lambda L)^{-1} q$ ,  $L = D_x^T D_x + D_y^T D_y$

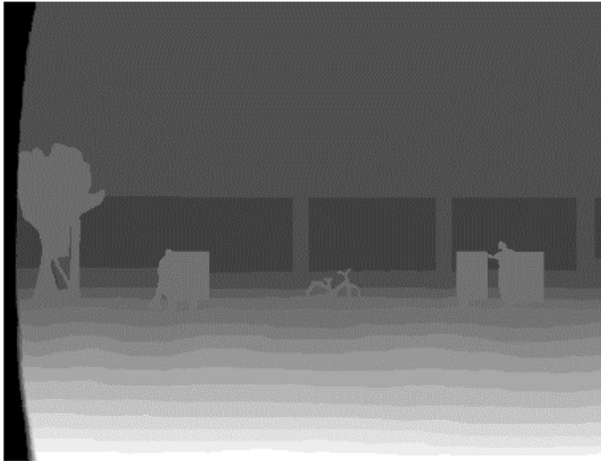
The depth image after WLSF is

$$t_w = F_\lambda(t_b) \tag{15}$$

Where  $t_w$  represents the depth image after WLSF, and  $F_\lambda(\cdot)$  represents the WLSF optimization. The Fig. 6 shows the depth map after WLSF optimization. It can be seen that after WLSF optimization, the noise of the depth map is significantly reduced, the edge of the object in the figure is smoother, and the original edge details are maintained, and no excessive smoothing occurs.

### 2.4 Points cloud reconstruction

In this paper, we need to obtain accurate 3D point cloud data of reconstructed scenes. The 3D data of objects in space includes height, width and depth. The camera inevitably loses depth information during the imaging process. If you want to restore three-dimensional geometric



**Fig. 6** The depth map after WLSF optimization

information such as the shape of an object, you need to take the three-dimensional coordinate values of the points on the surface of the object. The set of three-dimensional information of the surface points of these objects is a point cloud. The three-dimensional coordinate values of these spatial points can be solved by using the stereo vision principle to calculate the three-dimensional point coordinates, that is, using the projection relationship existing between the pixel coordinate system and the world coordinate system. Since we have obtained an accurate depth map in the previous step, the depth map contains the parallax information of the corresponding point of each pixel in the scene, so this paper uses the solution method based on the parallax principle.

As shown in the Fig. 7, it is assumed that the parallel binocular camera has a translation distance of  $b$  (i.e, the baseline length) in the  $x$ -axis direction,  $O_1$ 、 $O_2$  are the centers of the left and right cameras,  $p$  is the point on the object to be reconstructed,  $pO_1O_2$  and two The-camera-imaging planes are respectively assigned to  $p_1$ 、 $p_2$  points, and the  $p_1$ 、 $p_2$  points are on the same horizontal line, that is,  $I_1$ 、 $I_2$  are collinear.

Assuming that the coordinate systems of the left and right cameras  $C_1$ 、 $C_2$  are  $O_1x_1y_1z_1$ 、 $O_2x_2y_2z_2$ , the coordinates of the point  $p$  in the space can be obtained as  $(x_1, y_1, z_1)$ , in the left camera  $C_1$  coordinate system and  $C_2$  coordinates in the right camera. The system is  $(x_1 - b, y_1, z_1)$ , which can be obtained according to the proportion of central photography:

$$\begin{cases} u_1 - u_0 = a_x \frac{x_1}{z_1} \\ v_1 - v_0 = a_y \frac{y_1}{z_1} \end{cases} \quad (16)$$

$$\begin{cases} u_2 - u_0 = a_x \frac{x_1 - b}{z_1} \\ v_2 - v_0 = a_y \frac{y_1}{z_1} \end{cases} \quad (17)$$

Where  $u_0$ 、 $v_0$ 、 $a_x$ 、 $a_y$  are the internal parameters of the two cameras, and  $(u_1, v_1)$ 、 $(u_2, v_2)$  are the coordinates of the  $p$  point in the image. The two equations above can be used to obtain

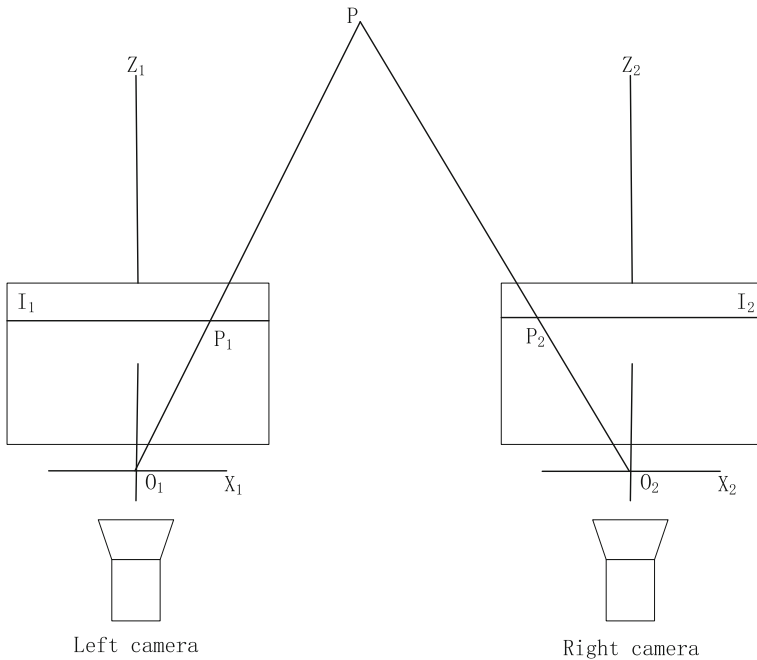


Fig. 7 Parallel binocular camera model

the three-dimensional coordinates of point  $p$ :

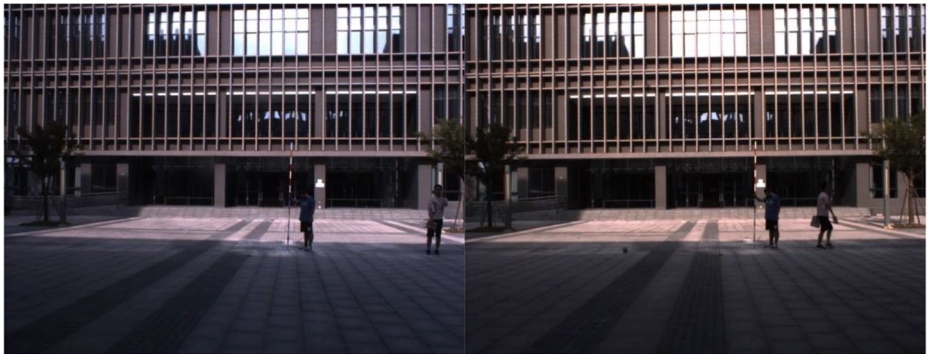
$$\begin{cases} x_1 = \frac{b(u_1 - u_0)}{u_1 - u_2} \\ y_1 = \frac{ba_x(v_1 - v_0)}{a_y(u_1 - u_2)} \\ z_1 = \frac{ba_x}{(u_1 - u_2)} \end{cases} \tag{18}$$

### 3 Results and discussion

Figure 8 shows the left and right calibration images. Figure 9 shows the calibration image after the polar line correction. After correction, the binocular camera polar lines are parallel, and the corresponding calibration points are on the same horizontal line.

We reconstruct the six calibration points in Fig. 8 by the calibration results obtained, and then analyze the accuracy of the calibration results by comparing with the actual coordinates of the calibration points. Table 1 shows the results of reconstructions before and after the correction of the six calibration points in Fig. 8. It can be seen that the corrected calibration point error is smaller and closer to the exact value.

Figure 10 shows the results of reconstruction of all calibration points. According to Zhang’s calibration principle, we have five calibration rods on the same plane to form a calibration board, and there are 30 calibration points on each calibration board. Figure 11 shows the

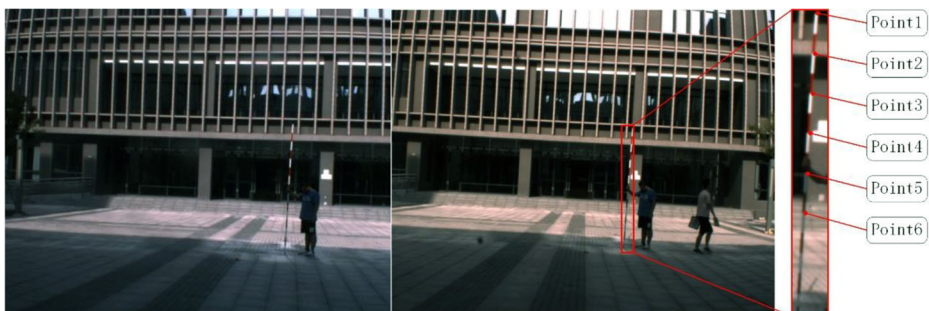


**Fig. 8** The left and right calibration images

results of reconstruction of all calibration boards. A total of 11 calibration boards are included. We calculate the average error of all calibration points contained on each calibration board as the error of this calibration board. Then shown them in Table 2. Figure 12 shows the uncorrected errors and the corrected errors of all the calibration boards, the difference of the uncorrected errors and the corrected errors, then calculates the total average of the uncorrected errors and the corrected errors. The comparison shows that the corrected coordinate point error is smaller than the coordinate point error before correction, and the actual coordinate average error of the calibration point is 1.276260 pixels, which can meet the reconstruction requirements.

Figure 13a, b show a set of left and right views taken simultaneously. There are two people (people1 and people2) and a bicycle (bicycle). Figure 13c is the disparity map. Figure 13d, e show the depth map obtained after WLSF. Figure 13d is shown in grayscale and Fig. 13e is shown in pseudo-heat map. Compared with the initial depth map (Fig. 12), the noise in the final depth map is significantly reduced, and the edge contours of each object are smoother and more accurate.

We use the calibration results and the depth map to perform point cloud reconstruction on the scene of the bicycle and the person. Figure 14 shows the point cloud reconstruction using the modified depth map and separates the point cloud within the depth range of the person. Then we use the height of the reconstructed person and the wheelbase of the bicycle to evaluate the reconstructed point cloud error.



**Fig. 9** Corrected left and right calibration images

**Table 1** Y coordinate of calibration points

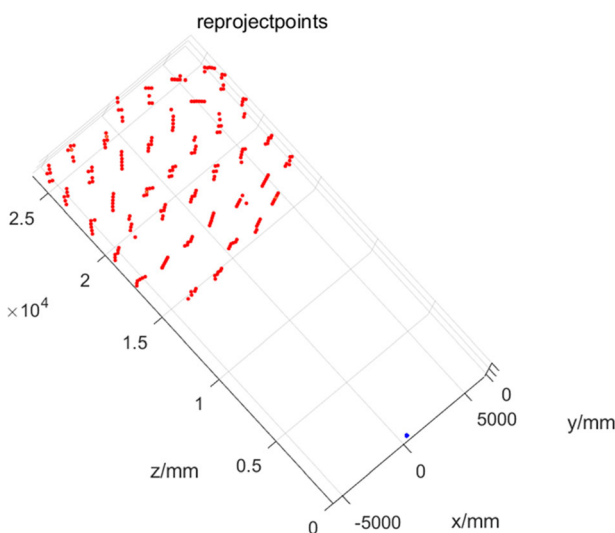
	Uncorrected left_image	Uncorrected right_image	Corrected left_image	Corrected right_image	Uncorrected Y difference	Corrected Y difference
Point1	488	483	389	387	5	1
Point2	539	534	441	441	5	0
Point3	589	584	492	491	5	1
Point4	639	634	543	543	5	0
Point5	690	684	594	594	6	0
Point6	740	735	646	646	5	0

Figure 15 and Table 3 show the results of reconstructing the bicycle shaft and the height of the person. The results show that the human height reconstruction error is within one centimeter. The wheelbase of the bicycle is slightly larger than 26.7 mm due to the identification of the center position of the bicycle shaft. But the scene we have done in 3D reconstruction has been 10 m to 20 m or even larger. We believe this error is within reasonable limits. The final reconstruction results can meet the needs of large-scale scenarios.

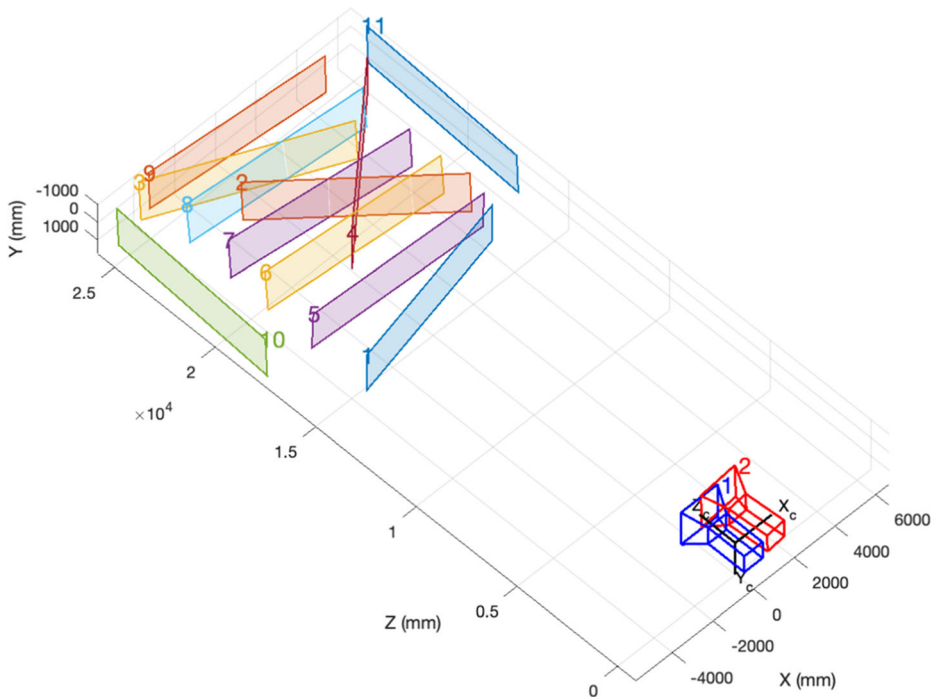
### 4 Conclusion

Through the optimization and improvement of the binocular stereo vision reconstruction algorithm, the 3D reconstruction in a large-scale scene is completed, and the reconstructed scene size is 10 m to 20 m or more, which solves the application of binocular vision based on large-scale scenes. The problem of stereoscopic 3D reconstruction.

The experimental results show that the application of the optimization algorithm in this paper can greatly expand the scope of application of Zhang’s calibration method, reduce the



**Fig. 10** Calibration point reconstruction



**Fig. 11** Calibration board reconstruction

error of depth map noise, and improve the accuracy of obtaining depth map, so as to obtain better point cloud reconstruction results.

Applying the method in this paper can be better applied to binocular 3D reconstruction in large-scale scenes than traditional methods, and expands the scope of application of binocular vision reconstruction method. It is worth mentioning that the system also has the potential to perform high-speed three-dimensional reconstruction. Compared with the laser measurement reconstruction method, the reconstruction method of binocular vision is greatly accelerated

**Table 2** Reproject points error

	Uncorrected error	Corrected error	Difference
Calib_board_1	2.537235	1.738874	-0.798361
Calib_board_2	1.724393	1.279794	-0.444599
Calib_board_3	1.621752	0.834303	-0.787449
Calib_board_4	2.096738	1.637696	-0.459042
Calib_board_5	1.885317	1.330309	-0.555008
Calib_board_6	1.764015	1.296668	-0.467347
Calib_board_7	2.030599	1.433154	-0.597445
Calib_board_8	1.641159	1.171892	-0.469276
Calib_board_9	0.8406	1.231024	0.390424
Calib_board_10	1.344919	0.875873	-0.469046
Calib_board_11	2.359082	1.209277	-1.149805
Average	1.804164	1.276260	-0.527903

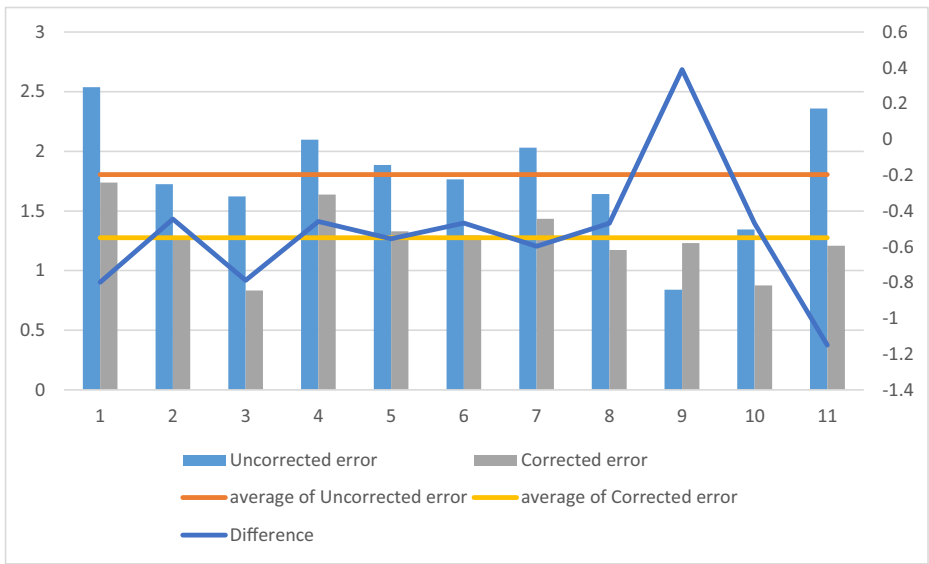


Fig. 12 Reproject points error

under the premise of ensuring accuracy, and it can completely replace the laser for high-speed three-dimensional reconstruction.

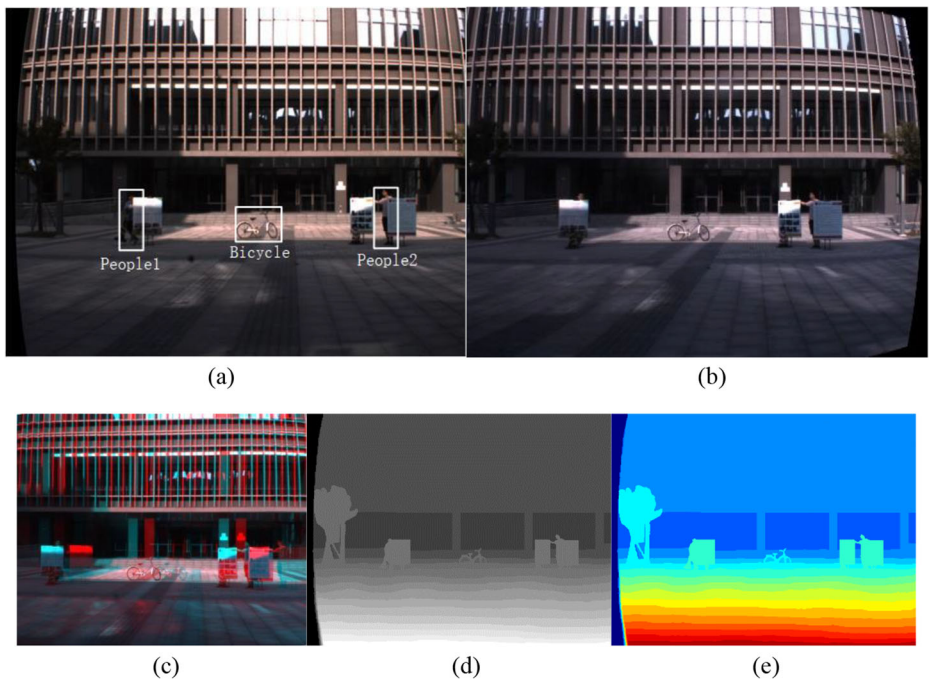
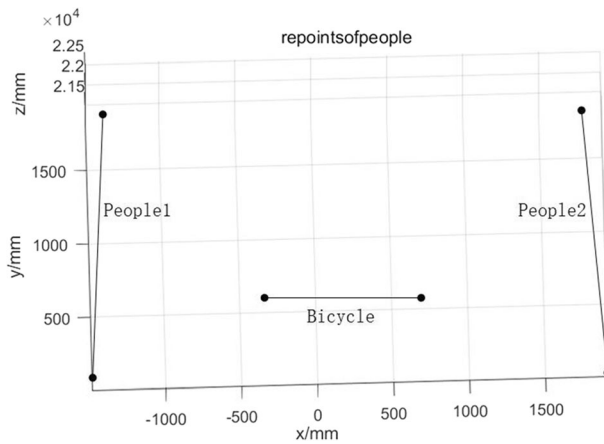


Fig. 13 a left image b right image c disparity map d depth map in grayscale e depth map in pseudo-heat map



**Fig. 14** Point cloud of people



**Fig. 15** People height and bicycle axis reconstruction point cloud

**Table 3** people height and bicycle axis reconstruction error

	Calculated value(mm)	Exact value(mm)	Error(mm)
People1	1796.1	1795	1.1
People2	1814.0	1810	4.0
Bicycle	1006.7	1000	6.7



**Acknowledgments** This work was supported by National Natural Science Foundation of China under the grant number 61502297 and 51707113.

## References

1. Burie JC, Bruyelle JL, Postaire JG (1995) Detecting and localising obstacles in front of a moving vehicle using linear stereo vision[J]. *Math Comput Model* 22(4):235–246
2. Farbman Z, Fattal R, Lischinski D, Szeliski R (2008) Edge-preserving decompositions for multi-scale tone and detail manipulation[J]. *ACM Trans Graph* 27(3):67–10
3. Gao S, Tong X, Chen P, Ye Z, Hu O, Wang B, Zhao C, Liu S, Xie H, Jin Y, Xu X, Liu S, Wei C (2019) Full-field deformation measurement by videogrammetry using self-adaptive window matching[J]. *Photogramm Rec* 34(165):36–62
4. Golodetz S, Cavallari T, Lord NA, Prisacariu VA, Murray DW, Torr PHS (2018) Collaborative large-scale dense 3D reconstruction with online inter-agent pose optimisation[J]. *IEEE Trans Vis Comput Graph* 24(11):2895–2905
5. Ha H, Han S, Lee J (2012) Fault detection on transmission lines using a microphone Array and an infrared thermal imaging camera[J]. *IEEE Trans Instrum Meas* 61(1):267–275
6. Hu Y (2011) Research on a three-dimensional reconstruction method based on the feature matching algorithm of a scale-invariant feature transform[J]. *Math Comput Model* 54(3):919–923
7. Hu Y, Chen Q, Feng S, Tao T, Asundi A, Zuo C (2019) A new microscopic telecentric stereo vision system - calibration, rectification, and three-dimensional reconstruction[J]. *Opt Lasers Eng* 113:14–22
8. Irijanti E, Nayan MY, Yusoff MZ (2011) Local stereo matching algorithm: using small-color census and sparse adaptive support weight[C]. In: National Postgraduate Conference IEEE
9. Liang X, Du Y, Wei D (2019) An Integrated Camera Parameters Calibration Approach for Robotic Monocular Vision Guidance[C]. In: 2019 34rd Youth Academic Annual Conference of Chinese Association of Automation (YAC)
10. Maolin Q, Songde MA (2000) Overview of camera calibration for computer vision[J]. *J Autom Sin* 26(1): 43–55
11. Park JH, Park HW (2006) A mesh-based disparity representation method for view interpolation and stereo image compression[J]. *IEEE Trans Image Process* 15(7):1751–1762
12. Schöps T, Sattler T, Häne C, Pollefeys M (2016) Large-scale outdoor 3D reconstruction on a mobile device[J]. *Comput Vis Image Underst* S1077314216301412:151–166
13. Shen Yue W, Qiang L, Chaoran W et al (2017) Design of 3D reconstruction system for outdoor scene based on binocular stereo camera[J]. *Comput Meas Control* 25(11):137–140 145
14. Vlamincq M, Luong H, Goeman W, Philips W (2016) 3D scene reconstruction using omnidirectional vision and LiDAR: a hybrid approach[J]. *Sensors* 16(11):1923
15. Wu P, Liu Y, Ye M, Li J, du S (2017) Fast and adaptive 3D reconstruction with extensively high completeness[J]. *IEEE Trans Multimedia* 19(2):266–278
16. Yang L, Wang B, Zhang R et al (2017) Analysis on Location Accuracy for Binocular Stereo Vision System[J]. *IEEE Photonics J* (99):1
17. Yang F, Shuaiang R, Enqi L et al (2019) Calibration method and regulation algorithm of binocular distance measurement in the large scene of image monitoring for overhead transmission lines[J]. *High Voltage Eng* 45(2):377–385
18. Zhang Z (2000) A flexible new technique for camera calibration[J]. *IEEE Trans Pattern Anal Mach Intell* 22(11):1330–1334
19. Zhang C, Zhang Q (2018) Research on volumetric calculation of multi-vision geometry UAV image volume[C]. In: 2018 Fifth international workshop on earth observation and remote sensing applications (EORSA)

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.