applied optics

Unconstrained self-calibration of stereo camera on visually impaired assistance devices

Huabing Li,[1,2] Kaiwei Wang,[1,*] Kailun Yang,[1,2,3] Ruiqi Cheng,[1] Chen Wang,[1] and Lei Fei[1]

[1]State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou 310027, China
[2]Surimage Technology Co., Ltd., Hangzhou 310023, China
[3]Department of Electronics, University of Alcalá, Madrid 28805, Spain
*Corresponding author: wangkaiwei@zju.edu.cn

Received 14 June 2019; revised 18 July 2019; accepted 18 July 2019; posted 18 July 2019 (Doc. ID 370009); published 8 August 2019

Stereo cameras are widely used in wearable visually impaired assistance devices (VIADs). However, the inevitable vibration, shock, and mechanical stress may make the camera pair become misaligned and cause a sharp decline in the quality of the acquired depth map, which significantly influences the performance of VIADs. In this paper, we propose an epipolar-constraint-based unconstrained self-calibration method that requires neither user involvement nor specific environment, while achieving a rotation accuracy of 0.83 mrad and a translation accuracy of 0.42 mm. Several approaches are proposed to address the image matching issues, including blurred images removal, mismatched key points removal, etc. Based on correctly matched key point pairs, a planar quadric-distribution approach is proposed to ensure the quality and consistency of the final key point group. These collection approaches ensure the reliability of key point pairs, which is the most important factor to realize high accuracy with minimum constraint. A comprehensive set of experiments demonstrates the high robustness of the proposed methods, which are suitable for VIADs. We also present a field test with blindfolded users to validate the flexibility and applicability of the approach.

https://doi.org/10.1364/AO.58.006377

1. INTRODUCTION

According to the World Health Organization, about 253 million people are living with vision impairments, 36 million of whom are blind [1]. For visually impaired people (VIP), daily activities such as walking in outdoor environments are difficult and even dangerous in some scenarios. With the aim of assisting perception and navigation, visually impaired assistance devices (VIADs) based on RGB-depth (RGB-D) sensors have been widely proposed [2–5]. Because depth maps are critical to navigational path detection and obstacle avoidance, the performances of VIADs are highly related to the quality of depth maps [6,7].

Currently, time-of-flight (ToF) [8], light-coding [9], and stereo matching [10,11] are three major approaches to obtain depth maps. ToF approaches usually require a projector and detector array. The depth maps are acquired by measuring the absolute time that a light pulse travels from a target to the detector array. Light-coding approaches mainly use an infrared (IR) projector to produce patterns with spatial codification and reconstruct depth maps through triangulation and other methods. However, limited by the power of the projector, the environmental illumination, such as sunlight, which is prone to submerge the light pulse and the projected patterns, ToF and light-coding approaches are not robust enough for outdoor applications. Stereo cameras acquire depth maps by matching two or more textured images derived from different lenses, which is a passive method and works well under daylight. Augmenting stereo cameras with an active light source, stereo matching can adapt to various environments, making it more suitable for VIADs [3]. In this paper, the stereo camera mounted on the VIAD is RealSense R200 [12]. Its IR laser projector produces a pseudo-random IR spot array on the target, enriching the texture and improving the performance of stereo matching in dark and low-texture environments.

When matching pixels in the image pair, stereo-matching algorithms only search pixels in the same row of the two images, which requires every row of the two images to be accurately aligned. The original images usually fail to meet the requirements because the optical axes of the two cameras are not perfectly parallel, and their relative translation is not horizontal, either. Thus, image pairs must be rectified to be aligned before stereo matching. For this reason, calibrating relative rotation and translation of the two cameras is critical. In some scenarios, the two cameras will not change their relative orientation and position, so applying calibration before leaving the factory is sufficient to meet the application requirements.

However, calibrating stereo cameras only once is not enough for VIADs. Generally, for wearable VIADs, the integrated stereo cameras may suffer from violent vibrations when users walk, run, or jump in daily usage. Specifically, head-mounted

VIADs may fall down from the height of a person, introducing heavy shock and mechanical stress to the stereo camera. Moreover, considering the limited weight and volume of VIADs, the support of integrated stereo cameras cannot be firm. All of these issues make it inevitable that the two cameras may frequently change their relative position and rotation. Although the change is relatively slight, our experiment shows that merely several mrad change of rotation may lead to depth errors as large as 100% at distances longer than 10 m. Therefore, a stereo camera in the VIAD needs to be accurately calibrated frequently, especially after falling down.

There are a lot of camera calibration methods in the literature. Some of them are based on a specific calibration pattern, special objects, or special scenes [13–20]; others are based on global optimization, which requires the environment to be static [21–26]. For visually impaired people, calibration patterns or calibration objects are inconvenient, and the scenes during navigation are unlikely to be static. Thus, these methods are not suitable for VIADs. To the best of our knowledge, no previous work has designed an unconstrained self-calibration method for the stereo camera in VIADs.

In this paper, we propose an unconstrained self-calibration method that requires neither user involvement nor any specific environment, while achieving the rotation accuracy of 0.83 mrad and the translation accuracy of 0.42 mm. Specifically, image pairs are arbitrarily acquired by the stereo camera in the VIAD; further, blurred and low texture ones will be removed. Key points are extracted from every image pair and matched based on Euclidean distances of their descriptors. To reduce mismatched key point pairs, a valid-box method is proposed, which is based on the *priori* that the disparity must be positive and the camera's misalignment in a stereo rig should be slight. To combine and select key point pairs obtained from all of the image pairs, a planar quadratic distribution method is proposed, which is to enforce area density of key points to obey quadratic distribution in the left image. Finally, the relative rotation and translation are solved by key point pairs based on epipolar constraint. To carry out the calibration, VIAD users only need to wear the device and traverse natural environments, and calibration will be accomplished quietly. Because the proposed method only depends on the features in natural scenes, it can be used for all types of stereo cameras. In this regard, the proposed method allows stereo cameras to calibrate themselves in daily usage, greatly increasing the reliability of VIAD.

To validate the flexibility and precision of the proposed method, a series of experiments were conducted. A blindfolded volunteer was invited to wear the VIAD and walk along the street to perform a field self-calibration experiment. The result shows that the accuracy of our method is close to Zhang's calibration-pattern-based method [13]. Based on the results of self-calibration, we carried out depth map recovery and depth precision experiments to evaluate the quality improvement of depth maps. The result demonstrates that our method significantly improves the density and precision of depth maps. Finally, a key point collection strategy comparison is implemented to prove that the proposed quadric distribution method is of important relevance to the performance of the proposed self-calibration.

The remainder of this paper is structured as follows: Section 2 reviews the related works that address self-calibration approaches. Section 3 introduces the principles and algorithms used in the proposed self-calibration method. Section 4 introduces experiments and analyzes the results. Section 5 concludes this paper and discusses other possible applications.

2. RELATED WORK

Stereo camera calibration approaches can be categorized into two major groups: specialized-pattern calibration approaches and self-calibration approaches. The former approaches require special calibration patterns, such as a chess pattern. Zhang's method is one of the most popular among them. Zhang's method requires taking photos of a chess board pattern from different orientations and positions. Then, the camera parameters are initialized by solving a set of homograph matrixes between the known chess board points and their projection points. Finally, intrinsic and extrinsic parameters were obtained through maximum-likelihood estimation. Zhang's method is not suitable for VIAD because carrying a chess pattern with visually impaired people is inconvenient, and taking photos as required is too difficult for them. But because Zhang's method is the most widely used one, we take its mean result as the baseline for comparison.

Stereo camera self-calibration methods fall into two groups: 1) calibration-object-dependent methods [14–20] and 2) global-optimization methods [21–26].

For the first approaches, Broggi *et al.* [15] presented an approach of calibrating the stereo cameras mounted on an autonomous vehicle. A number of marks were placed on the hood of the vehicle, which were taken as calibration patterns. The relative orientation of cameras was estimated by minimizing the reprojection error of the marks. Collado *et al.* [16] presented a self-calibration method focusing on stereo cameras mounted on vehicles, which adopted road lane boundaries as calibration patterns. They defined the fitness function according to the parallelism and coincidence of projections of the road lane boundaries. Then, pitch, roll, and height of the stereo camera were estimated through a genetic algorithm. Banglei *et al.* [17] calibrated stereo cameras through homography constraints based on image pairs of planar scenes. The authors constructed a polynomial equation system with radial distortion and solved the equation with at least five corresponding point pairs. Hu *et al.* [18] presented an approach of intrinsic parameters calibration by using two perpendicular planes as calibration objects. The authors assumed that the relationship of the two cameras was pure translation and the two cameras shared the same intrinsic parameters. Liu *et al.* [19] presented an approach to calibrate the relative pose between stereo cameras based on laser spots projected on parallel planes. Bok-Suk *et al.* [20] presented an approach to calibrate a wide-baseline stereo camera mounted on offshore facilities by exploiting sea horizon and points at infinity. These approaches require no specific calibration pattern, but some special objects or scenes are indeed adopted as calibration patterns. Those special objects or scenes may not be difficult to find in their application, but, for VIADs, they are not flexible enough.

For the second approaches, some researchers combined camera calibration with simultaneous localization and mapping [21,25,26]. These researchers regarded camera parameters as variables to be optimized rather than known parameters and applied global optimization to camera parameters and 3D structure of environments at the same time. Thao *et al.* [22] presented a framework for continuous calibration of stereo cameras. Camera parameters are estimated through epipolar constraint, trilinear constraints, and bundle adjustment. Then, they are continuously refined through a Karman filter. Fadi [23] calibrated the two cameras of a stereo rig independently rather than take them as a whole. A series of poses of the left and right camera is initialized independently through epipolar constraints. Then, the final relative pose of the two cameras is estimated through nonlinear optimization. Shuo *et al.* [24] proposed a method based on image pairs, which applies bundle adjustment to calibrate the mast mechanism, accelerometer, and gyroscope at the same time. These approaches don't require a calibration pattern, but, because the environment is required to be static, they are not suitable for VIADs.

Although all of these approaches are instructive, they cannot totally meet requirements of the stereo camera in VIADs. In the following sections, we will elaborate on the proposed self-calibration method and its performance.

3. APPROACH

A. Principles of Self-Calibration

A stereo camera has two horizontally separated cameras, as shown in Fig. 1. The rotation and translation from the right to left camera can be represented as rotation matrix $\boldsymbol{R}$ and translation vector $\boldsymbol{t}$, respectively, where $\boldsymbol{R}$ is a 3×3 matrix and $\boldsymbol{t}$ is a 3×1 matrix.

Thus, the relationship of the two cameras can be represented as Eq. (1), where $p_1 = [x_1 \;\; y_1 \;\; z_1]^T$ and $p_2 = [x_2 \;\; y_2 \;\; z_2]^T$ are the coordinates of one point P in the coordinate system of the left camera and right camera, respectively. In this paper, unless specified, subscript 1 and subscript 2 stand for the left and right camera, respectively:

$$p_2 = R \cdot p_1 + t. \tag{1}$$

The antisymmetric matrix $[t]_\times$ is defined as Eq. (2), which is the matrix form of the cross product of vector $\boldsymbol{t}$, that is, for an arbitrary vector $\boldsymbol{a}$, $|t \times a|$ can be represented as

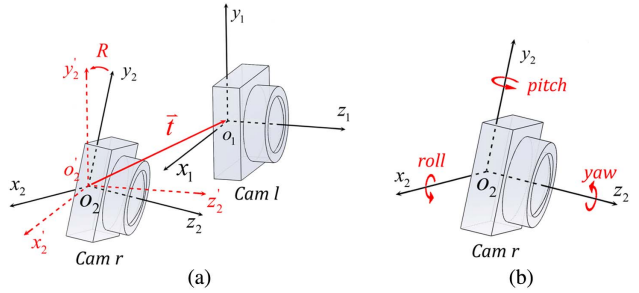

Fig. 1. Illustration of relative rotation and translation. (a) *Cam l* and *Cam r* are the left camera and right camera with coordinate systems $o_1 - x_1y_1z_1$ and $o_2 - x_2y_2z_2$, respectively. $\boldsymbol{R}$ denotes the rotation from $o_2 - x_2y_2z_2$ to $o_1 - x_1y_1z_1$, $\boldsymbol{t}$ denotes the translation between two parallel coordinate systems from $o_2' - x_2'y_2'z_2'$ to $o_1 - x_1y_1z_1$. (b) Rotation is represented by roll, pitch, and yaw.

$$[t]_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}. \tag{2}$$

Then, the epipolar constraint can be derived based on Eq. (1). The epipolar constraint is shown as Eq. (4), where E is called the essential matrix. Once $\boldsymbol{E}$ is given, the translation vector $\boldsymbol{t}$ and rotation matrix $\boldsymbol{R}$ can be obtained by applying singular value decomposition (SVD) to $\boldsymbol{E}$:

$$E = [t]_\times R, \tag{3}$$

$$p_2^T E p_1 = 0. \tag{4}$$

In practice, the locations of key point P on the two image planes $m_1(u_1, v_1)$ and $m_2(u_2, v_2)$ are known, as shown in Fig. 2, where coordinates of p_1 and p_2 are unknown. A pinhole model can be used to describe the relationship between the location on the image plane and the coordinates in the camera coordinate system as in Eq. (5):

$$s\tilde{m} = A \cdot p. \tag{5}$$

In Eq. (5), $\tilde{m} = [u \;\; v \;\; 1]^T$ is the homogeneous coordinate of point m on the image plane, and A is the intrinsic parameter matrix of the camera, which is shown in Eq. (6). Precisely, A describes the focus length and principal point of the camera, where all parameters are in pixels. $p = [x \;\; y \;\; z]^T$ is the coordinate of the object point in the camera coordinate system, and s is a factor, which equals to z in the coordinate of point p:

$$A = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{6}$$

Thus, Eq. (4) can be rewritten as Eqs. (7) and (8), where F is called the fundamental matrix. As shown in Eq. (8). $\tilde{m}_1 = [u_1 \;\; v_1 \;\; 1]^T$ and $\tilde{m}_2 = [u_2 \;\; v_2 \;\; 1]^T$ are two points located on the image plane that is imaged by the same object point. Once sufficient point pairs are collected, F can be solved according to Eq. (8):

$$F = A_2^{-T} E A_1^{-1}, \tag{7}$$

$$\tilde{m}_2^T F \tilde{m}_1 = 0, \tag{8}$$

where A_1 and A_2 are the intrinsic parameter matrices of the left and right camera, respectively. In this paper, the intrinsic

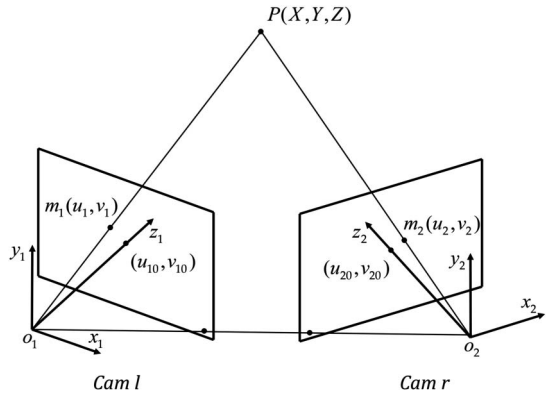

Fig. 2. Illustration of epipolar constraint.

parameters are regarded as known, and, because the lens of the RealSense R200 is thin and well-fixed with adhesive, the intrinsic parameters are unlikely to be changed. In this sense, once the fundamental matrix F is solved, essential matrix E can be easily obtained through Eq. (9):

$$E = A_2^T F A_1. \tag{9}$$

As mentioned above, the rotation matrix R and the translation vector $\boldsymbol{t}$ can be solved by applying SVD to essential matrix E. However, it should be noted that the norm of vector $\boldsymbol{t}$ cannot be solved by epipolar constraint because there is no constraint on the length of the vector $\boldsymbol{t}$ in epipolar constraint. The geometry meaning of Eq. (8) explains it clearly. Equation (8) indicates that the vector $\overrightarrow{o_1 m_1}$, $\overrightarrow{o_2 m_2}$, and $\boldsymbol{t}$ should be coplanar. If $P(X, Y, Z)$ is not specified, P may locate at any point on line $o_1 m_1$, leading to the uncertainty of the length of $|o_1 o_2|$.

Epipolar constraint cannot solve the length of vector t; for stereo cameras, however, there is another constraint, that is, the relative rotation and translation of the two cameras are fixed, which is not considered in epipolar constraint. Unfortunately, even with this additional constraint, length of vector t cannot be solved in any way without the size of objects, either. We prove this as follows: assuming there are N common key points in M pairs of images; in other words, there are N common object points in the real world, and they are imaged in M views. Formally, P^i denotes the coordinate of the ith common object point in the world coordinate system, and p_1^{ij} and p_2^{ij} indicate the coordinate of P^i in the left and right camera coordinate system at the jth view. The constraints of the stereo rig can be represented by Eq. (10):

$$\begin{aligned} p_1^{ij} &= R_1^j \cdot P^i + t_1^j, \\ p_2^{ij} &= R_2^j \cdot P^i + t_2^j, \\ p_2^{ij} &= R \cdot p_1^{ij} + t, \\ i &\in [1, N]; \qquad j \in [1, M]. \end{aligned} \tag{10}$$

Without loss of generality, we set the coordinate system of the left camera in the first view to be the world coordinate system, that is, $P^i = p_1^{i0}$ Then, Eq. (10) can be rewritten to be

$$\begin{aligned} p_1^{ij} &= R_1^j \cdot p_1^{i0} + t_1^j \qquad (a) \\ p_2^{ij} &= R_2^j \cdot p_1^{i0} + t_2^j \qquad (b) \\ p_2^{ij} &= R \cdot p_1^{ij} + t \qquad (c) \\ i &\in [1, N]; \qquad j \in [1, M]. \end{aligned} \tag{11}$$

Substituting Eqs. (a) and (b) into (c) in Eq. (11), two additional constraints can be obtained, as shown in Eq. (12), which describes that the relative pose of the two cameras is fixed:

$$\begin{aligned} R &= R_2^j (R_1^j)^{-1} \\ t &= t_2^j - R \cdot t_1^j. \end{aligned} \tag{12}$$

According to the geometry meaning of equations, there should be at least one set of solutions, which can be represented as $R_{01}^j, t_{01}^j; R_{02}^j, t_{02}^j; R_0, t_0$. It is not difficult to verify that, for any nonzero factor a, $R_{01}^j, a \cdot t_{01}^j; R_{02}^j, a \cdot t_{02}^j; R_0, a \cdot t_0$ are solutions of Eqs. (11) and (12), too. Therefore, the additional constraints cannot determine the length of vector $\boldsymbol{t}$.

B. Algorithm

The misalignment of the stereo camera in VIADs is usually slight. But it may cause a sharp decline in quality of the depth map, especially for the regions with large depth. Although the quality of the depth map declines, the VIAD still can assist perception and navigation in relatively low performance. Therefore, the accuracy and robustness of self-calibration are extremely important, while the real-time performance of self-calibration is comparably less important. Because the two cameras of a stereo camera are relatively fixed, in every pair of images, matched key points should represent the same epipolar constraint. Thus, sufficient image pairs can be taken and combined by certain strategies to achieve high accuracy. Based on this idea, we designed the self-calibration algorithm as shown in Fig. 3.

Infrared image pairs are acquired through the stereo camera, but only clear and high-texture ones remain. For every image pair, key points are extracted through SURF [27] and matched based on the Euclidean distance of their descriptors. Because of the noise of image sensors, parallax of the two cameras, limitation of descriptors, and other factors, key points matching is not completely reliable. To increase robustness of key point matching, a valid-box method is proposed to remove mismatched key point pairs.

To comply with epipolar constraint, the two matched key points in one image pair should represent the same point in object space. However, this situation is not so ideal in practice. On the one hand, because of aberration of the lens, the imaging model is not compatible with the pin-hole model completely; on the other hand, location error of the matched key point is

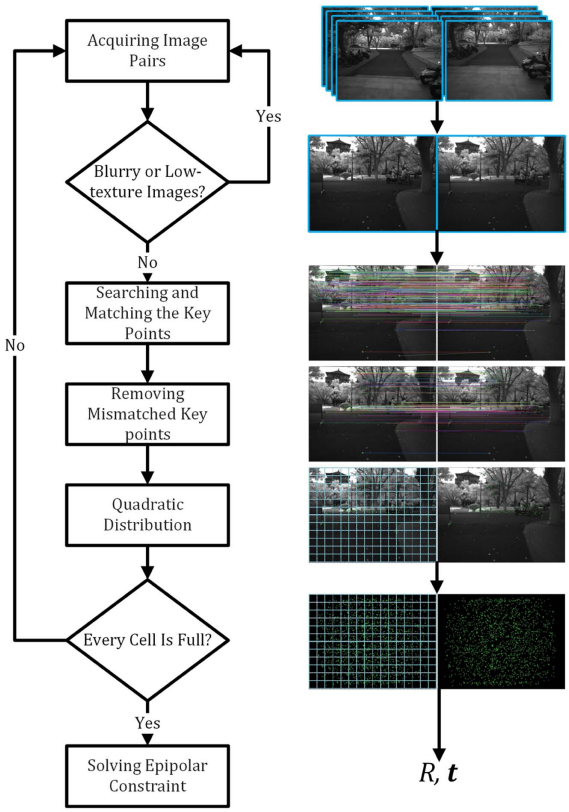

Fig. 3. Flow chart of the proposed algorithm.

inevitable. To improve the robustness and accuracy of the self-calibration, a quadric distribution strategy is applied to combine key points in different image pairs. Finally, the relative rotation and translation are solved by these key point pairs.

1. Blurred and Low-Texture Images Removal

Because blurred and low-texture images are the disadvantage of key points extraction, it is necessary to remove them. In general, clear and high-texture images are rich in high-frequency variation in intensity, while blurred or low-texture images are poor in this regard. Thus, a gradient feature can be used to distinguish blurred or low-texture images. Specifically, we convolve a third-order Laplacian operator with the original image to obtain the Laplace response, which stands for the gradient feature map of the original image. For clear and high-texture images, their Laplace responses are rich, resulting in a large gray-scale variance; if the image is blurred or low texture, however, the variance should be small. Therefore, a variance threshold can be set to remove the blurred image. To strike a good trade-off between the collection difficulty and the image quality, it is necessary to collect image samples and analyze their variances. Then, the variance value can be chosen as the threshold when most of the clear images remain and most of the blurry images are removed. Figure 4 shows the difference of clear images, blurred images, and low-texture images. It is obvious that the gray-scale histogram of the clear and rich-texture image spreads widely, while the gray-scale histograms of low-texture and blurry images are narrow.

2. Mismatched Key Point Pairs Removal

For every image pair, key points are extracted through SURF and matched based on the Euclidean distance of their descriptors. In this paper, descriptors of key point are 64-dimensional vectors. Because the Euclidean distance of two descriptors represents the similarity of the two points, the larger their Euclidean distance is, the more likely they are mismatched. Thus, a common matching strategy is to simply set a Euclidean distance threshold as Eq. (13) to remove mismatched key point pairs, where $Coef$ is a constant, and $D_{\min}$ and $D_{\max}$ are the minimal and maximal Euclidean distance of all the key point pairs in the image pair, respectively:

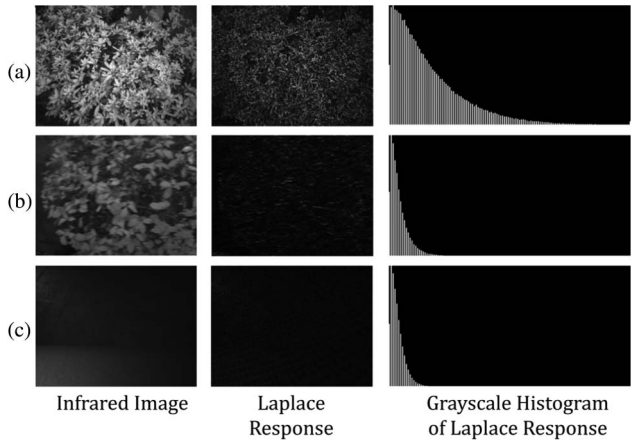

Fig. 4. Laplace responses and gray-scale histograms of different images. (a) Clear and rich-texture image. (b) Blurred image. (c) Low-texture image.

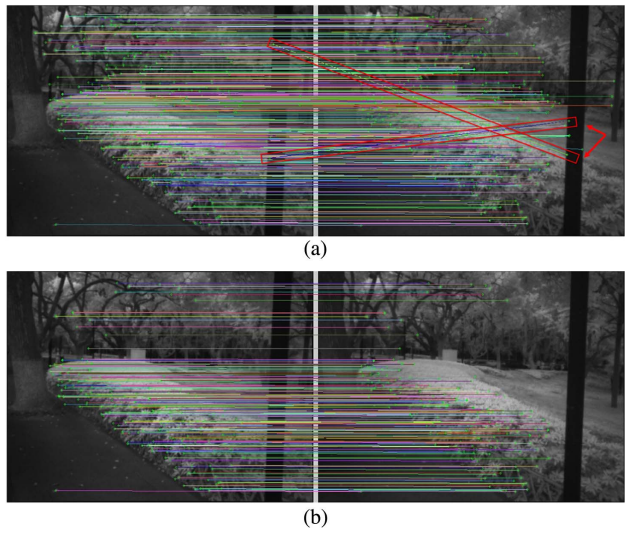

Fig. 5. Example of mismatched key point pair removal. (a) Matching result by Euclidean distance of descriptors only, where the matched pairs indicated by the red arrow are obviously wrong. (b) Matching result refined by proposed method.

$$th = Coef \cdot (D_{\min} + D_{\max}). \tag{13}$$

Straightforwardly, a stricter threshold can reduce mismatched key point pairs but cannot remove all of them. In this paper, although the two cameras of the stereo camera are not completely parallel and horizontal, their relative rotation and translation misalignment are quite slight. Thus, a pair of matched key points in left and right images must fall in a narrow horizontal band. In addition, disparities are always positive; the matched key points in the right image must locate at the right side of the corresponding key points in the left image. Thus, for a given key point in the left image, a narrow rectangle valid box can be set in the right image, and, once the corresponding right key point falls out of the valid box, the matched key point pair can be regarded as invalid and removed. This way, the matching result obtained by Euclidean distance threshold can be refined by the valid box.

In this paper, we set the rectangle valid box to be 20 pixels in width and 10 pixels in height. As shown in Fig. 5, the obviously wrong matched key point pairs pointed by red arrows in (a) can be effectively removed. Furthermore, the key point pairs at the top of (a) are points with large depth, and, if they are correctly matched, their disparities should be small positive value. But they are removed, as shown in (b), meaning that their disparities are either negative or over 20 pixels. Either one of the two possibilities is wrong and should be removed.

3. Planar Quadric Distribution

As discussed at the beginning of the section, matched key point pairs in every image pair should comply with the same epipolar constraint, and the strategies to select and combine key points over pairs of images to achieve optimal accuracy are the key to the problem.

In order to address the above issue, a key points selection and combination method based on planar distribution and matching quality is proposed. Planar distribution means to

enforce the key points pattern to obey a certain rule. More specifically, the relationship between key points area density and the distance to the center of image is set to obey a quadratic curve. Because key points are discrete, density must be discrete. To approximately fit the quadratic curve, the image is divided into a $N_w \times N_h$-cell grid. The center density ρ_{ij} of a certain cell C_{ij} is determined by Eq. (14), where r_{ij} is the distance between the centers of C_{ij} and the image, as shown in Fig. 6. In Eq. (14), ρ_0 and α are determined through two boundary conditions: 1) the density at the edge of the image is zero; 2) the integral of area densities is one:

$$\rho_{ij} = \rho_0 - \alpha \cdot r_{ij}^2, \tag{14}$$

where ρ_{ij} is regarded as the mean area density of C_{ij} approximately. Thus, the expected amount of key points m_{ij} in C_{ij} can be calculated as follows, where w and h are width and height of the image, respectively:

$$m_{ij} = \frac{w \cdot h}{N_w N_h} \rho_{ij}. \tag{15}$$

For every new image pair, key point pairs are added to the $N_w \times N_h$ cells according to the locations of key points in the left image. Once key points falling into C_{ij} *are* more than m_{ij}, they will be sorted by their matching qualities, and only the best m_{ij} key point pairs will remain. The quality of two matched key points is defined to be Euclidean distance of their descriptors. Obviously, the smaller the Euclidean distance, the more similar the two points are. Continuously acquiring new images and extracting key point pairs from them add to the grid until all the cells are full. Thus, the key points on the left image will obey the quadratic distribution approximately, as shown in Fig. 7.

The proposed quadric distribution method ensures that key points distribute symmetrically and prevents key points from concentrating at the edge of the image. In addition, quadric distribution ensures that key points reach the largest density at the center of image where aberration is the smallest. Moreover, sorting according to match quality ensures that key points in every cell are the best-matched ones falling into the cell over all of the image pairs; thus, the object point of every pair of key points coincides better in space.

4. Relative Rotation and Translation Calculation

The relative rotation and translation are solved based on epipolar constraint. The essential matrix E is solved according to

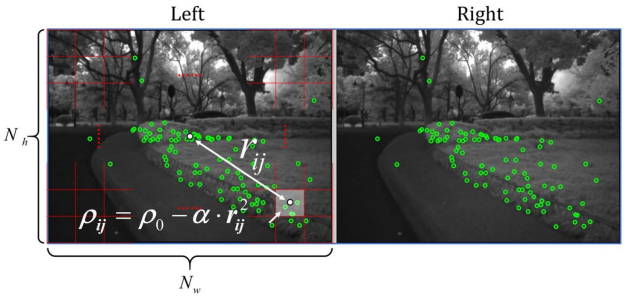

Fig. 6. Area densities calculation. The green circles are key points in one image pair that have been matched. The red grid divides the left image into $N_h \times N_w$ cells; for a certain one of them, the key point area density is defined as the equation in the image.

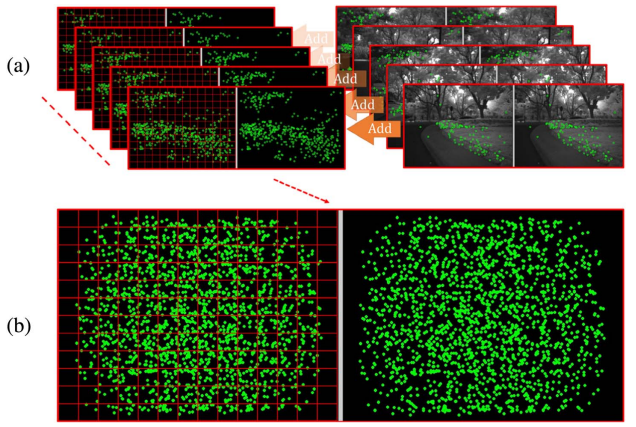

Fig. 7. Illustration of key points collection and quadric distribution. (a) The right images are the sequence of image pairs and the matched key points; the left images show how key points are filled into the grid. (b) The result of quadratic distribution.

Eqs. (8) and (9); then, rotation matrix R and unit vector $\boldsymbol{e}$ of translation vector $\boldsymbol{t}$ are obtained by applying SVD to essential matrix. The key lies in the method used to solve fundamental matrix F. Among those methods, eight-point [28], LMed [29], and RANSAC [30] are popular. In this paper, the RANSAC method with nonstrict parameters is adopted to remove some outlier points; then, the eight-point method is applied to solve the fundamental matrix.

In the previous section, we proved that only unit vector $\boldsymbol{e}$ of vector $\boldsymbol{t}$ could be calculated through the proposed self-calibration method. But because the stereo rig in the VIAD is not likely to be stretched or compressed, the distance of the cameras should be constant, meaning that the norm of $\boldsymbol{t}$ should be constant. The change of translation vector $\boldsymbol{t}$ is mainly caused by the rotation of the left camera, which changes the components of $\boldsymbol{t}$. Thus, we get t_x, t_y, t_z by multiplying unit vector $\boldsymbol{e}$ by the distance of the two cameras obtained in the previous section.

4. EXPERIMENTS

In this section, the approaches and results of experiments are elaborated. First, Zhang's method was applied, and the average results were taken as the baseline. Then, field self-calibration experiments were implemented at the Yuquan Campus of Zhejiang University, Hangzhou (China). After that, depth map recovery and depth precision experiments were implemented to analyze depth map quality improvement through self-calibration. Finally, several distribution strategies were compared to prove that the quadric distribution was of great significance to the accuracy.

A VIAD named "Intoer" [31], as shown in Fig. 8, was used to carry out the experiments. The Intoer is a glasses-like device, where a RealSense R200 stereo camera is mounted on the front, a pair of bone conduction headphones are mounted on the back, and a USB 3.0 cable is connected to a portable computer. Images and depth maps are acquired by the stereo camera and transferred by the USB cable. Obstacles and traversable area detections are implemented on the portable computer based on the acquired images and depth maps [3,4]. Finally, the detection results are notified to the VIP through customized stereo sound.

Fig. 8. Intoer VIAD. Intoer contains a RealSense R200 stereo camera, a pair of bone conduction headphones, and a USB 3.0 cable, which mainly assists in navigation and obstacle avoidance based on depth maps and images.

A. Zhang's Calibration Method

Because Zhang's method is the most widely used calibration method, its calibration results were taken as the baseline to evaluate the performance of the proposed self-calibration method. To ensure the performance of Zhang's method, a high-precision chess pattern was used to be the calibration pattern, as shown in Fig. 9. The chess pattern is a 400 mm width square glass plate. Grids of the pattern are 25 mm width thin metal films, whose size error is superior to 10 μm parallelism, and perpendicularity errors are lower than 0.09 mrad.

We implemented Zhang's calibration method seven times. The calibration results are shown in Table 1, where yaw, pitch, roll are the rotation angles around axis Z, Y, X, respectively, and t_x, t_y, t_z are components of translation vector $\boldsymbol{t}$ along the X, Y, Z axis, respectively. The mean value of the seven calibration results is taken as the baseline in the following sections.

B. Field Self-Calibration Experiment

We set up the experiment as shown in Fig. 10(a). A volunteer was invited to wear the VIAD with a misaligned stereo camera, and the proposed self-calibration algorithm ran on the handheld laptop. In the field test, the volunteer walked along streets at the Yuquan campus of Zhejiang University, where routes were recorded by GPS during the experiment. When the volunteer walked, the self-calibration algorithm ran at the same time, which automatically acquired one pair of images every 0.7 s until the calibration finished. The resolution of images is 640 × 480, which is determined by the stereo camera.

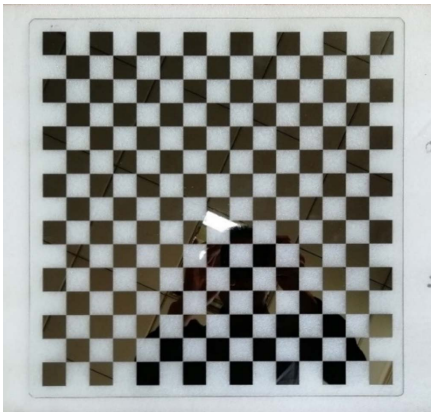

Fig. 9. High-precision glass chess pattern.

Table 1. Calibration Results of the Stereo Camera Through Zhang's Method

	Rotations/mrad			Translation/mm		
Index	Yaw	Pitch	Roll	t_x	t_y	t_z
1	−0.17	−6.52	5.99	−69.88	−1.42	−0.89
2	−0.11	−6.82	5.53	−69.89	−1.15	−0.24
3	−0.02	−5.85	5.34	−69.88	−1.36	−1.05
4	0.06	−6.18	5.35	−69.88	−1.15	−1.55
5	0.02	−4.71	5.26	−69.89	−1.32	−0.64
6	0.00	−6.70	5.51	−69.89	−1.33	−0.36
7	−0.04	−4.97	5.06	−69.88	−1.45	−1.03
Mean	−0.04	−5.97	5.43	−69.88	−1.31	−0.82
S.D.	0.08	0.84	0.29	0.01	0.12	0.45

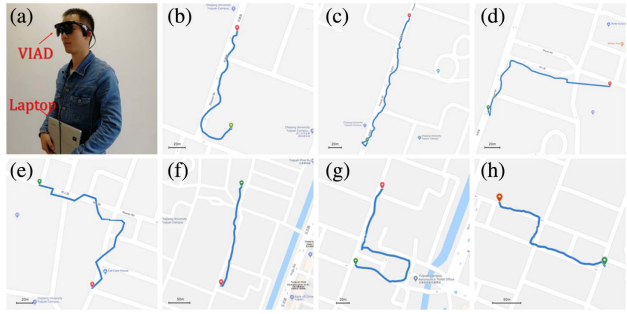

Fig. 10. Experiment setup and routes. (a) The experiment setup: the misaligned stereo camera was in the VIAD, and a laptop was used to run self-calibration. (b)–(h) Seven routes where the volunteer walked to implement self-calibrations; the volunteer merely walked along the road, and images of scenes are collected automatically, while the self-calibration ran at the same time.

The self-calibration was implemented seven times continuously. Routes of the volunteer are shown in Fig. 10. The routes consisted of crowded streets with vehicles and pedestrians moving on the road; further, the scenes were complex and natural without artificial adjustment. To collect sufficient image pairs, the volunteer needed to walk about 300–500 m, which would take about 6 to 8 min. Reducing the time gap of image collection can reduce the time cost but may lead to repetitive distribution of key points and reduce the accuracy of calibration. With sufficient image pairs, the relative rotation and translation were solved within several seconds. The distance and time were determined by walking speed, richness of the environment texture, and other factors.

Results of self-calibration are shown in Table 2. According to the table, the absolute differences between the mean angles and baseline values are all no more than 0.38 mrad. The maximum root mean square error (RMSE) of seven measures with respect to the baseline value is only 0.83 mrad, while the maximum standard deviation (S.D.) by Zhang's method is up to 0.84 mrad. As for the translation vector, the absolute differences between the mean t_x and t_y and baseline values are all no more than 0.2 mm, and the maximum RMSE with respect to the baseline value is only 0.42 mm, quite close to 0.12 mm of Zhang's method. t_z is not considered because it has few effects on improving the quality of depth maps but

Table 2. Results of Self-Calibration

	Rotation/mrad			Translation/mm		
Index	Yaw	Pitch	Roll	t_x	t_y	t_z
1	0.43	−6.79	5.83	−70.02	−1.46	0.33
2	0.22	−5.43	6.66	−70.02	−1.05	0.97
3	0.16	−6.39	4.92	−69.89	−0.88	4.37
4	0.36	−7.51	5.16	−69.97	−1.32	−2.50
5	0.16	−5.64	6.90	−69.96	−0.37	3.20
6	0.28	−5.69	5.04	−69.99	−1.48	−1.83
7	0.51	−7.01	5.87	−69.99	−1.56	−1.75
8	0.43	−6.79	5.83	−70.02	−1.46	0.33
Mean	0.30	−6.35	5.77	−69.98	−1.16	0.40
Baseline	−0.04	−5.97	5.43	−69.88	−1.31	−0.82
RMSE	0.36	0.83	0.80	0.10	0.42	2.74

may lead to large rotation transformation on both images, reducing the valid region of the stereo images. Thus, it can be set to 0 to reduce parameters and improve robustness.

The self-calibration results are basements of stereo image rectification. In this paper, the method proposed by Fusiello *et al.* [30] is adopted to rectify image pairs, which can be regarded as twice of rotation transformation based on R and $\boldsymbol{t}$ as follows:

$$\begin{aligned} s\tilde{m}_2' &= A_s R_2 R_t A_s^{-1} \tilde{m}_2, \\ s\tilde{m}_1' &= A_s R_1 R_t A_s^{-1} \tilde{m}_1, \end{aligned} \tag{16}$$

where R_1 and R_2 are the transformations caused by R; further, to simply the following derivation, we set $R_1 = I_3$ and $R_2 = R$, where R_t is the transformation based on $\boldsymbol{t}$, A_s is the common intrinsic parameter for the both images, and $\tilde{m}_1 = [\,u_1 \quad v_1 \quad 1\,]^T$ and $\tilde{m}_2 = [\,u_2 \quad v_2 \quad 1\,]^T$ are original points on the left and right images.

For a pair of rectified points, $\tilde{m}_1'$ and $\tilde{m}_2'$, the disparity $\Delta u = u_1' - u_2'$ determines the depth value of the point as Eq. (17), where D is the baseline of the stereo rig. Therefore, $d(\Delta u)/\Delta u$ can be used to evaluate the depth accuracy by the image rectification. $\Delta v = v_1' - v_2'$ is the vertical difference between a pair of points. Because the stereo-matching algorithm has only matching points on the same row, Δv influences the matching quality of the two images; thus, $d(\Delta v)$ indicates the density and reliability of the depth map:

$$\begin{aligned} d &= \frac{f_x \cdot D}{\Delta u}, \\ E_d &= \frac{\Delta d}{d} = \frac{d(\Delta u)}{\Delta u}. \end{aligned} \tag{17}$$

In this case, to analyze the impact of self-calibration accuracy on depth maps, the relationship between $d(\Delta u)/\Delta u$, $d(\Delta v)$ and calibration results should be derived.

Directly applying Eq. (16) to obtain $d(\Delta u)$ and $d(\Delta v)$ is too difficult; to simplify the calculation, yaw, roll, pitch, $\boldsymbol{t_x}$, $\boldsymbol{t_y}$, and $\boldsymbol{t_z}$ are applied to rectify the image pair separately. Their absolute sums are taken as final expressions of $d(\Delta u)$ and $d(\Delta v)$, as shown in Eq. (18) according to [32]

$$\begin{aligned} d(\Delta u) &= |f_x TH_2 TV_2 d(\text{roll})| + |f_x(1+TH_2)d(\text{pitch})| \\ &\quad + |f_x TV_2 d(\text{yaw})| + |\Delta u_{\text{org}} d(\theta_z)| + |\Delta v_{\text{org}} d(\theta_y)|; \\ d(\Delta v) &= |f_y(1+TV_2)d(\text{roll})| + |f_y TH_2 TV_2 d(\text{pitch})| \\ &\quad + \left| f_y TH_2 d(\text{yaw}) \right| \\ &\quad + \left| f_y \left(TH_2 \frac{\Delta v_{\text{org}}}{f_y} + TV_1 \frac{\Delta u_{\text{org}}}{f_x} \right) d(\theta_z) \right| \\ &\quad + |\Delta u_{\text{org}} d(\theta_y)|; \\ TH_i &= \frac{c_x - u_i}{f_x}, \quad TV_i = \frac{c_y - v_i}{f_y}, \quad \theta_z = \left| \frac{t_z}{t} \right|, \quad \theta_y = \left| \frac{t_y}{t} \right| \\ |t| &= \sqrt{t_x^2 + t_y^2 + t_x^2}; \quad \Delta u_{\text{org}} = u_1 - u_2; \quad \Delta v_{\text{org}} = v_1 - v_2. \end{aligned} \tag{18}$$

Taking the parameters of RealSense R200 in this paper as an example. In Eq. (18), $f_x = f_y = 580$, $c_x = 320$, $c_x = 240$, and $0 \le u_1 < u_2 < 640$; $0 \le v_1 < 480$, $0 \le v_2 < 480$;. Because the stereo matching only processes the disparity less than 64, thus $0 < \Delta u \le 64$, the relationship between Δu_{org} and Δu is derived in [32]. As mentioned in Section 3.B.2, one pair points must locate in the narrow horizontal band, that is, $-10 < \Delta v_{\text{org}} < 10$. To analyze the worst situation, we set $\Delta v_{\text{org}} = 10$ and take RMSEs as d(angle) into Eq. (18); the results are shown in Eq. (19), which means the worst situation of $d(\Delta u)/\Delta u$ and $d(\Delta v)$ at the center of the images:

$$\begin{cases} \frac{d(\Delta u)}{\Delta u} = \begin{cases} \frac{0.41}{\Delta u} + 0.039; & \Delta u > 3.45 \\ \frac{0.68}{\Delta u} - 0.039; & \Delta u \le 3.45 \end{cases} \\ d(\Delta v) = \begin{cases} 0.0066\Delta u + 0.44; & \Delta u > 3.45 \\ -0.0066\Delta u + 0.49; & \Delta u \le 3.45 \end{cases} \end{cases}. \tag{19}$$

In the same way, we derive the mathematical expression of E_d and $d(\Delta v)$ under Zhang's calibration results and unrectified results and draw the curves, as shown in Fig. 11. The curves are the worst E_d and $d(\Delta v)$ at the center of the image, showing that the self-calibration improves the E_d and $d(\Delta v)$ greatly as

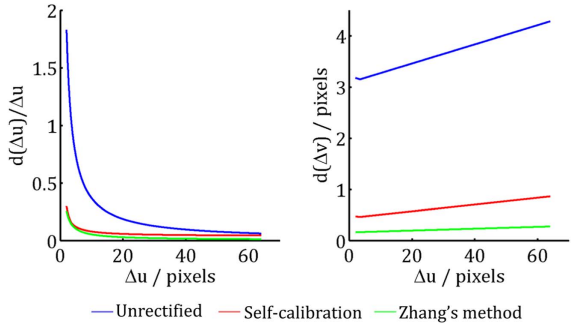

Fig. 11. Worst situation of E_d and $d(\Delta v)$ at the center of the images. The red and green curves are rectified by self-calibration and Zhang's method, respectively, while the blue curves are the unrectified ones. Compared with unrectified ones, the proposed self-calibration shows much better results. Compared with Zhang's chess-pattern-based method, the self-calibration's performance is similar.

well as Zhang's method. In this regard, the accuracy, density, and reliability of depth maps are improved by self-calibration.

C. Depth Maps Recovery

In the experiment, Fusiello's method was used to rectify misaligned image pairs based on self-calibration and Zhang's calibration results, respectively. The stereo-matching method SGM was applied to obtain depth maps from image pairs. The depth maps obtained by misaligned image pairs and two groups of rectified image pairs are shown in the second to fourth columns of Fig. 12.

Comparing depth maps in the second and third columns, it is not difficult to find that the self-calibration greatly improves the reliability and density of depth maps. For example, in Fig. 12(a), the wall behind the sculpture should be too far to calculate disparity, meaning that depth of the wall should be invalid in the depth maps. But the depth maps without rectification show that the wall is even closer than the sculpture, while the depth maps rectified by self-calibration are much more reliable; in Figs. 12(b)–12(f), depth maps in the second column are obviously much more sparse than in the third column, especially the depth of far objects, showing the density improvement with self-calibration. The depth maps in the third and the fourth column are quite similar, showing that the proposed self-calibration improves the quality of depth maps as well as Zhang' calibration.

The experiment results meet the curves of $d(\Delta v)$ in Fig. 11, the proposed self-calibration decreases the $d(\Delta v)$ as much as Zhang's method, leading to great improvement of the quality on depth maps. Based on our approach, a dense and reliable depth map is ensured, which is preferred to assist the visually impaired so as not to leave out potential obstacles. In this regard, the safety of the navigation assistance system has been enhanced.

D. Depth Precision Experiment

To quantitatively analyze the quality improvement of depth maps, a depth precision experiment was set up, as shown in Fig. 13. The VIAD and a laser ranger were parallelly fixed on a movable tripod. A rich-texture wall was taken as the target object, and the distance between the target and the tripod would be measured by a laser ranger and stereo camera, respectively. Because the accuracy of the laser ranger is up to ± 1.5 mm, its measurement result was taken as ground truth.

The depth measuring started at a distance of 0.5 m from the wall and increased at intervals of about 0.5 m until it was about 18 m. At every step, the depth measured by a laser ranger was recorded, while three pairs of images were collected at the same time. Figure 14 shows four samples of the left images of acquired image pairs, where the brick wall beside the door is the rich-texture wall. The red bounding boxes were manually drawn to sample the depth values. When drawing the box, we obeyed the following rules: 1) the boxes should be rectangles with the aspect ratio of 4:3; 2) the width of the boxes should be equal to the width of the wall; 3) the boxes should locate at the bottom of the brick wall. Thus, the boxes represented the same region of the wall in every image. Furthermore, it is arranged that the spots of the laser ranger always located at the central area of the boxes; thus, the distance could represent the mean depth of the area of the wall.

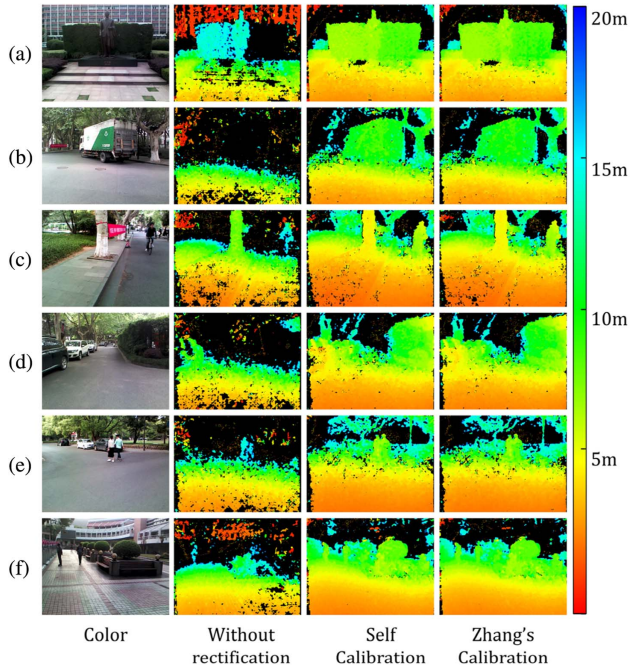

Fig. 12. Qualitative examples of depth map quality improvement through calibration. The first column images are color images; the second column images are depth maps obtained by misaligned images; the third and the fourth column show depth maps obtained by images rectified based on the proposed self-calibration and Zhang's calibration.

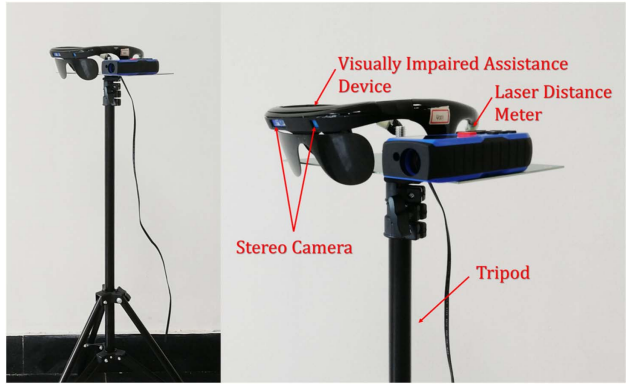

Fig. 13. Depth precision experiment setup. The VIAD and a laser ranger with an accuracy of ± 1.5 mm were parallelly fixed on a moveable tripod. The distance measured by a laser ranger was regarded as the ground truth to evaluate the precision of depth maps.

Fig. 14. Four samples of the left images of acquired image pairs. The distances from the wall to the VIAD are 1.066, 6.082, 12.104, and 18.110 m, respectively.

For every measurement, the three misaligned image pairs were rectified and matched to obtain depth maps as in Section 4.C. To reduce the influence of noise, we applied a median filter to the sample box area of depth maps and took the average value of the sample boxes in the three depth maps to be the final measurement result.

The result of the precision experiment is shown in Fig. 15. The abscissa is ground truth obtained by the laser ranger, and error ratios in the figure are all calculated through dividing the measurement value by ground truth. In Fig. 11, the unrectified curve shows that E_d is larger than 100% when the disparity is less than 4 pixels, which is about 10 m in distance. In Fig. 15, the orange curve is the depth error ratio of mismatched image pairs, which has been over 100% at the distance of 9 m, illustrating that the experiment results are consistent with the derivation results. Because of the large $d(\Delta v)$, as shown in Fig. 11, the depth maps of unrectified images become too sparse to be measured at a distance over 11 m, as shown in Fig. 15. While the green curve is the depth error ratio of image pairs rectified by the mean results of self-calibration, whose maximum error is only 25.2%, it is much lower than the unrectified one. The red curve is the depth error ratio of image pairs rectified by the mean results of Zhang's calibration. It can be easily seen that the green curve is only a little higher than the red curve, indicating that the accuracy of the proposed self-calibration is quite close to Zhang's calibration.

The performances of each self-calibration and Zhang's calibration are shown in the dark green and dark red regions. More specifically, we rectified the images and calculated the depth error ratio based on seven times self-calibration and obtained seven curves, respectively, and the region was drawn by the highest and the lowest ones. Therefore, this region shows the variation range in the accuracy of the depth measure of one single calibration result. The upper boundary of the dark green region corresponds to the worst performance of the rectification in seven times self-calibration, but it still greatly improved the depth accuracy. This illustrates that, once the stereo camera is misaligned, it can be rectified to work normally even with the worst one single time of self-calibration. Moreover, the dark green and dark red regions are highly overlapped, further

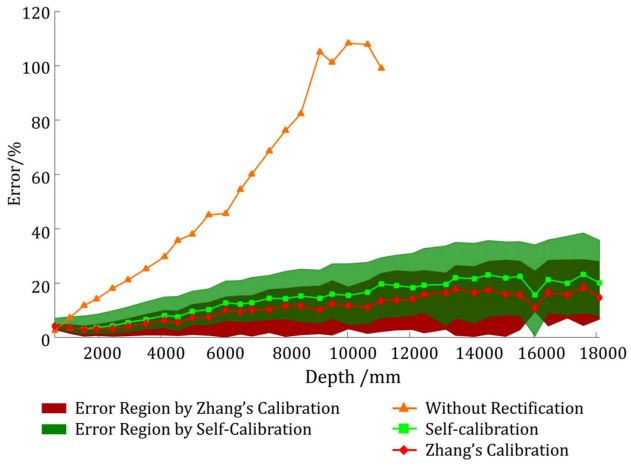

Fig. 15. Depth error ratios of depth maps obtained by misaligned image pairs and rectified images.

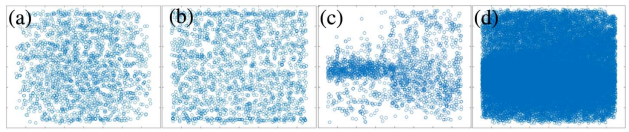

Fig. 16. Samples of different key points collection strategies. (a) Key points with quadratic distribution. (b) Key points with even distribution. (c) Key points without distribution. (d) All of the key points collected in a single time calibration.

verifying that the performances of our self-calibration and Zhang's calibration methods are close.

E. Key Points Collection Strategies Comparison

To demonstrate the effectiveness of the proposed quadric distribution strategy, different kinds of key points collection strategies are evaluated on the same image pairs. In Fig. 16, four strategies are shown. Specifically, the strategy used in Fig. 16(c) is simple: collecting key point pairs until there are 2000 of them, so that the key points distribution is totally random. Figure 16(d) shows all of the key points collected during one calibration, that is, the quadric distribution is to select some key points pairs from them.

The RMSEs of self-calibration results based on four distribution strategies are shown in the Table 3. For the nondistribution strategy, the distribution of key points is completely random, and key points may gather around the corners, as shown in Fig. 16(c), resulting in huge RMSEs, as shown in the third row of Table 3. Simply increasing the key point amount leads to a better result, but the improvement is quite limited. In Fig. 16(d), there are more than 30,000 key points collected by the nondistribution strategy, 15 times more than that of others. However, the calibration result shown in the last row of Table 3 is much worse than the results in the first two rows. Furthermore, once the environment of self-calibration is more adverse, the nondistribution strategy may lead more key points to gather around corners, which may result in worse performance. A suitable distribution strategy ensures that key points have a stable and symmetrical distribution pattern, as shown in Figs. 16(a) and 16(b), leading to much better RMSEs compared with nondistribution strategies. Compared with even distribution, our proposed quadratic distribution makes key points concentrate more on the center of image,

Table 3. RMSE of Different Key Point Collection Strategies

	Rotation/mrad				Translation/mm			
	Yaw	Pitch	Roll	Mean	t_x	t_y	t_z	Mean
QD[a]	0.36	0.83	0.80	0.66	0.10	0.42	2.74	1.09
ED[b]	0.33	1.52	1.42	1.09	0.49	1.55	6.62	2.89
ND[c]	1.12	4.99	2.99	3.03	4.24	2.94	17.09	8.09
AP[d]	0.51	2.76	1.57	1.61	0.85	1.15	8.81	3.60

[a]QD: Quadric Distribution.
[b]ED: Even Distribution.
[c]ND: Nondistribution.
[d]AP: All Points.

reducing the influence of distortion. Thus, a higher accuracy is achieved.

5. CONCLUSION

In this paper, we present an unconstrained self-calibration method for the stereo camera in VIADs based on epipolar constraint. This method entails minimum participation of the user and minimum requirement on environments but achieves high precision and robustness. In this method, image pairs are acquired arbitrarily. The key point collection mechanisms, including blurred image removal, valid-box based mismatched key points removal and quadric-distribution-based key points selection, are the keys to achieve few constraints and high accuracy at the same time. As the experiments and field tests illustrate, the proposed method achieves an accuracy of 0.83 mrad error on rotation and 0.42 mm on translation vector. The accuracy is comparable with the calibration-pattern-based method such as Zhang's method. Based on the proposed approach, the reliability, density, and precision of depth maps are ensured, enhancing the safety and robustness of navigation assistance system with the VIAD.

In the future, we aim to apply the proposed self-calibration method to the multicamera and multimodal systems, such as the multistereo camera system capturing the depth maps within the 360° field of view the visual system with polarization cameras or RGB-IR cameras.

Funding. Zhejiang Provincial Public Fund (NO.2016C33136); State Key Laboratory of Modern Optical Instrumentation (MOI) (111303-I21805); Hangzhou SurImage Technology Co., Ltd; Krvision Technology Co., Ltd.

REFERENCES

1. R. R. Bourne, S. R. Flaxman, T. Braithwaite, M. V. Cicinelli, A. Das, J. B. Jonas, J. Keeffe, J. H. Kempen, J. Leasher, and H. Limburg, "Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis," Lancet Global Health **5**, e888–e897 (2017).
2. R. Cheng, K. Wang, K. Yang, N. Long, J. Bai, and D. Liu, "Real-time pedestrian crossing lights detection algorithm for the visually impaired," Multimedia Tools Appl. **77**, 20651–20671 (2018).
3. K. Yang, K. Wang, W. Hu, and J. Bai, "Expanding the detection of traversable area with RealSense for the visually impaired," Sensors **16**, 1954 (2016).
4. K. Yang, K. Wang, L. Bergasa, E. Romera, W. Hu, D. Sun, J. Sun, R. Cheng, T. Chen, and E. López, "Unifying terrain awareness for the visually impaired through real-time semantic segmentation," Sensors **18**, 1506 (2018).
5. X. Zhao, K. Wang, K. Yang, and W. Hu, "Unconstrained face detection and recognition based on RGB-D camera for the visually impaired," Proc. SPIE **10225**, 1022509 (2017).
6. K. Yang, K. Wang, H. Chen, and J. Bai, "Reducing the minimum range of a RGB-depth sensor to aid navigation in visually impaired individuals," Appl. Opt. **57**, 2809–2819 (2018).
7. K. Yang, L. M. Bergasa, E. Romera, and K. Wang, "Robustifying semantic cognition of traversability across wearable RGB-depth cameras," Appl. Opt. **58**, 3141–3155 (2019).
8. R. Lange and P. Seitz, "Solid-state time-of-flight range camera," IEEE J. Quantum Electron. **37**, 390–397 (2001).
9. J. Salvi, J. Pages, and J. Batlle, "Pattern codification strategies in structured light systems," Pattern Recogn. **37**, 827–849 (2004).
10. H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," IEEE Trans. Pattern Anal. Mach. Intell. **30**, 328–341 (2008).
11. A. Geiger, M. Roser, and R. Urtasun, "Efficient large-scale stereo matching," in *Asian Conference on Computer Vision* (Springer, 2010), pp. 25–38.
12. L. Keselman, J. Iselin Woodfill, A. Grunnet-Jepsen, and A. Bhowmik, "Intel RealSense stereoscopic depth cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2017), pp. 1–10.
13. Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1330–1334 (2000).
14. H. Zhuang, "A self-calibration approach to extrinsic parameter estimation of stereo cameras," Robot. Auton. Syst. **15**, 189–197 (1995).
15. A. Broggi, M. Bertozzi, and A. Fascioli, "Self-calibration of a stereo vision system for automotive applications," in *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation* (IEEE, 2001), pp. 3698–3703.
16. J. M. Collado, C. Hilario, A. de la Escalera, and J. M. Armingol, "Self-calibration of an on-board stereo-vision system for driver assistance systems," in *IEEE Intelligent Vehicles Symposium* (2006), pp. 156–162.
17. B. Guan, Y. Shang, and Q. Yu, "Planar self-calibration for stereo cameras with radial distortion," Appl. Opt. **56**, 9257–9267 (2017).
18. Z. Hu and Z. Tan, "Calibration of stereo cameras from two perpendicular planes," Appl. Opt. **44**, 5086–5090 (2005).
19. Z. Liu, Y. Yin, S. Liu, and X. Chen, "Extrinsic parameter calibration of stereo vision sensors using spot laser projector," Appl. Opt. **55**, 7098–7105 (2016).
20. B.-S. Shin, X. Mou, W. Mou, and H. Wang, "Vision-based navigation of an unmanned surface vehicle with object detection and tracking abilities," Mach. Vis. Appl. **29**, 95–112 (2018).
21. J. Sola, "Multi-camera VSLAM: from former information losses to self-calibration," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, San Diego, CA, USA (2007).
22. T. Dang, C. Hoffmann, and C. Stiller, "Continuous stereo self-calibration by camera parameter tracking," IEEE Trans. Image Process. **18**, 1536–1550 (2009).
23. F. Dornaika, "Self-calibration of a stereo rig using monocular epipolar geometries," Pattern Recogn. **40**, 2716–2729 (2007).
24. S. Zhang, S. Liu, Y. Ma, C. Qi, H. Ma, and H. Yang, "Self calibration of the stereo vision system of the Chang'e-3 lunar rover based on the bundle block adjustment," ISPRS J. Photogramm. Remote Sens. **128**, 287–297 (2017).
25. J. Sola, A. Monin, M. Devy, and T. Vidal-Calleja, "Fusing monocular information in multicamera SLAM," IEEE Trans. Robot. **24**, 958–968 (2008).
26. L. Heng, G. H. Lee, and M. Pollefeys, "Self-calibration and visual slam with a multi-camera system on a micro aerial vehicle," Auton. Robots **39**, 259–277 (2015).
27. H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: speeded up robust features," in *European Conference on Computer Vision* (Springer, 2006), pp. 404–417.
28. R. I. Hartley, "In defense of the 8-point algorithm," in *Proceedings of IEEE International Conference on Computer Vision* (IEEE, 1995), pp. 1064–1070.
29. Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," Artif. Intell. **78**, 87–119 (1995).
30. A. Fusiello, E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs," Mach. Vis. Appl. **12**, 16–22 (2000).
31. The "Intoer" VIAD, http://www.krvision.cn/.
32. "The derivation of stereo rectification error," http://www.wangkaiwei.org/file/publications/Stereo_Rectification_Errors_Caused_by_Calibration.pdf.