

## Research Article

# Vision System of Mobile Robot Combining Binocular and Depth Cameras

**Yuxiang Yang, Xiang Meng, and Mingyu Gao**

*Department of Electronics and Information, Hangzhou Dianzi University, Hangzhou, China*

Correspondence should be addressed to Mingyu Gao; [mackgao@hdu.edu.cn](mailto:mackgao@hdu.edu.cn)

Received 27 April 2017; Revised 12 July 2017; Accepted 16 August 2017; Published 24 September 2017

Academic Editor: Wendy Flores-Fuentes

Copyright © 2017 Yuxiang Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to optimize the three-dimensional (3D) reconstruction and obtain more precise actual distances of the object, a 3D reconstruction system combining binocular and depth cameras is proposed in this paper. The whole system consists of two identical color cameras, a TOF depth camera, an image processing host, a mobile robot control host, and a mobile robot. Because of structural constraints, the resolution of TOF depth camera is very low, which difficultly meets the requirement of trajectory planning. The resolution of binocular stereo cameras can be very high, but the effect of stereo matching is not ideal for low-texture scenes. Hence binocular stereo cameras also difficultly meet the requirements of high accuracy. In this paper, the proposed system integrates depth camera and stereo matching to improve the precision of the 3D reconstruction. Moreover, a double threads processing method is applied to improve the efficiency of the system. The experimental results show that the system can effectively improve the accuracy of 3D reconstruction, identify the distance from the camera accurately, and achieve the strategy of trajectory planning.

## 1. Introduction

With the development of the society, the application of the robots is rising rapidly in the society [1]. With the increase of the application field, the accuracy of the robot movement and the stability of the movement are very necessary for the promotion of the robot. That is how to improve the accuracy of the robot has been widely concerned, and more and more schools and companies have been involved in the research on how to improve the stability of the robot. 3D imaging technologies directly affect the accuracy of the robot movement. There exists a variety of 3D imaging technologies to acquire depth information about our world. In general, they can be categorized into two major classes: stereo matching methods and direct depth measuring methods.

Stereo vision [2–4] mainly researches on how to use imaging technology to obtain the distance information of the objects in the scene from the images with different views. At present, the basic method of stereo vision is to obtain the same scene image from two or more different perspectives and to obtain the corresponding disparity between pixels in different images and then it is simple to derive the target space position in the scene through the principle of triangulation.

Binocular stereo vision simulates human visual system, using two 2D images of the same scene obtained by two cameras from different point of views. The goal of binocular stereo matching is to find out the corresponding matching points in the left and right image after calibration. The difference in  $X$  coordinate of the corresponding points is named as disparity, which can form a disparity map. The key objective of stereo matching methods is to reduce the matching ambiguities introduced by low-texture regions, and they can be generally classified into two categories: local matching methods and global matching methods. Researchers have done a wealth of work on stereo matching in the past few years. But, unfortunately, the fundamental problems in stereo such as occlusion and texture-less regions remain to be unsolved.

Direct depth measuring camera [5, 6] mainly uses TOF ranging principle. The system illumination light source emits light beam to the scene, and detection sensor detects reflected light signal and figures out the propagation time of the optical signal. Finally, spatial coordinates of the object are calculated by the relative position of the optical signal propagation velocity and light source and sensor. These sensors measure time delay between transmission of a light pulse and detection of the reflected signal on an entire frame once by using

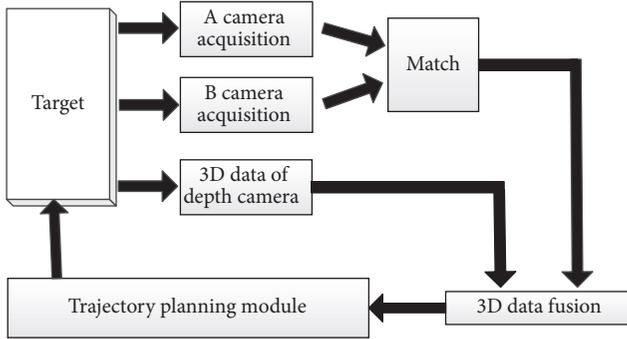


FIGURE 1: The proposed 3D reconstruction system.

extremely faster shutter. Hence TOF sensors can obtain near real-time scene depth. But, in the current generation, TOF sensors are limited in terms of resolution.

In general, binocular stereo vision system has the advantages of simple structure and high vertical resolution, but the stereo matching process is complex and has lots of matching errors in the weak texture. The vertical resolution of depth camera is independent of the distance of the scene in the imaging range, and it is relatively simple to achieve real-time three-dimensional image by using the data of depth camera, but the resolution and stability of TOF depth camera are lower by the limitation of hardware.

In this paper, a novel vision system of mobile robot is proposed, which can achieve more accurate 3D data by fusing stereo matching and direct depth measurement. In the proposed 3D reconstruction system, an image processing machine is used to simulate the human brain, and two parallel cameras and a depth camera are used to simulate human eyes as shown in Figure 1. Specifically, the image processing host utilizes binocular stereo cameras and achieves 3D reconstruction data based on stereo matching. Then image processing host will fuse the 3D reconstruction data of stereo matching with the data of depth camera and achieve 3D reconstructed data with high precision. Finally, it can figure out the physical distance of the scene objects from the camera by using this 3D data and achieve the strategy of trajectory planning [7, 8] to improve the stability of the robot. Moreover, double threads processing method is applied to improve the efficiency of the system. When the robot works, image processing machine obtains 3D reconstruction data in short time; the latest motion route is designed to guide the robot running. The experiments demonstrate that the 3D data fusion between the depth camera and stereo camera can meet the requirement of actual applications, but also it can improve the accuracy of matching. And, in practice, it is proved that the effect is very good and can improve the accuracy of 3D reconstruction.

The rest of this paper is organized as follows: Section 2 introduces the design of the system. In Section 3, we describe the framework of the 3D reconstruction system. In Section 4, experimental results are detailed and justify the superiority of our approach. Finally, conclusions are made in Section 5.

## 2. Design of the System

As shown in Figure 2, the whole system is composed of five parts: two color cameras of the same type, one depth camera, one image processing host, one mobile robot, and one mobile control host. Two color cameras are placed in the top of the mobile robot on the same horizontal line, and depth camera is placed below the left camera in the same vertical line. Three cameras communicate with image processing host through three USB3.0 line; the mobile robot is driven by two brushless DC motors and they are controlled by the mobile robot control host. The mobile robot control host and the image processing host communicate with each other through the serial port. In normal operation, the system is initialized. The three cameras start to take pictures of the same scene synchronously and transmit the image data to the image processing host through USB3.0 in real-time. Image processing host, first of all, utilizes the correlation algorithm [9, 10] to deal with two color images and calculate the disparity data of two images. Second, the image processing host matches the 3D image data of the low-resolution depth camera with the left image data of the color cameras. And, then, the image processing machine projects the spatial point detected by the TOF camera to the matching view. Finally, the 3D reconstruction data of the binocular camera is fused with the 3D data of the depth camera to obtain the 3D reconstruction data in high precision, by using which the trajectory planning is carried out. When the image processing host is in the process, image capturing thread updates and refreshes the scene data ceaselessly to ensure planning in short time. When image processing thread is completed, the trajectory planning data will be sent to the mobile control host through the serial port by the thread; mobile control host will control mobile robot to avoid obstacle.

As shown in Figure 2, the horizontal placement of two color cameras is particularly important for the accuracy of binocular matching. We can obtain relative rotation parameters and relative translation parameters by using a MATLAB toolbox for calibrating [11, 12]. The parameters will be used to correct the images and will make the images in the line alignment, and the overlapping area of the two images will be cut [13, 14]. The purpose of these preprocessing operations is to improve the accuracy of stereo matching and improve the matching time.

From Figure 2, we can also see the relative position of the depth camera and the left color camera. In order to make the scene reconstruction and fusion more accurate, it is necessary to register the data obtained by depth camera and color camera; namely, it is essential to match between the depth camera and the left color camera [15–17]. An effective camera registration method is developed between the depth camera and the left color camera in this paper. Firstly, we calibrate the two images (one is from left color camera; the other is from depth camera) to get the rotation parameters and relative translation parameters. Then, the image processing host uses the parameters to project the point of the depth camera to the image of left color camera. At last, a novel fusion method is proposed to get the high precision depth data.

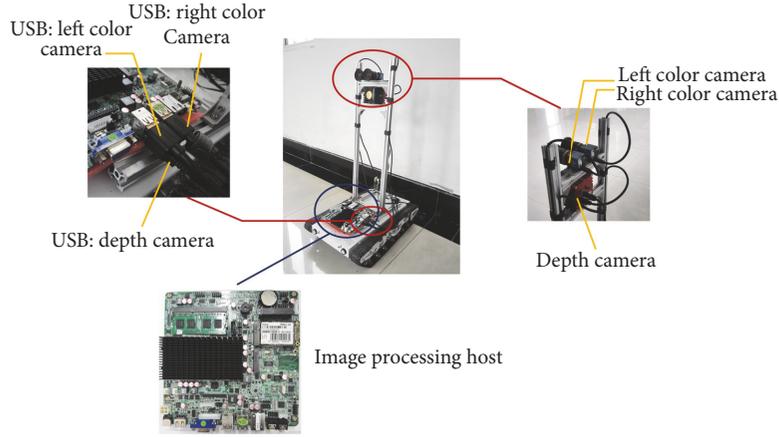


FIGURE 2: The components of the system.

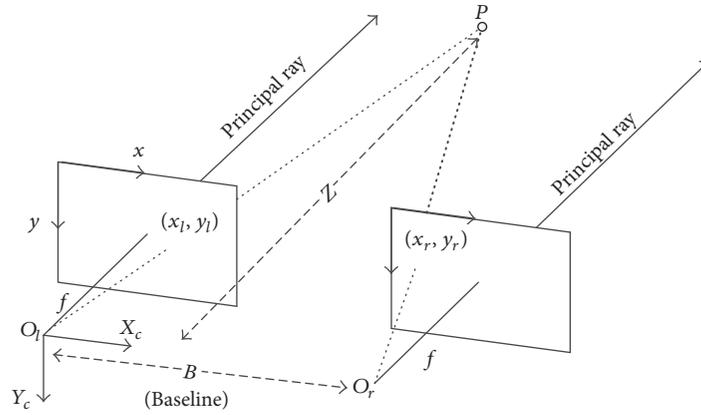


FIGURE 3: Analysis of binocular ranging.

### 3. 3D Fusion Reconstruction System

3.1. *Binocular Ranging and Calibration.* Let  $(u, v)$  be the pixel coordinate of color image;  $(X_c, Y_c, Z_c)$  is the corresponding coordinate in the color camera coordinate system. Based on the principle of small hole imaging,

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 \\ 0 & \frac{1}{dy} \end{bmatrix} \begin{bmatrix} f & 0 \\ 0 & f \end{bmatrix} \begin{bmatrix} \frac{X_c}{Z_c} \\ \frac{Y_c}{Z_c} \\ 1 \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix}. \quad (1)$$

Here,  $f$  is the focal length of color camera;  $(c_x, c_y)$  are the coordinates of the principal point.  $d_x$  and  $d_y$  are physical sizes of pixel in the horizontal and vertical directions, respectively.

Define  $f_x = f/d_x, f_y = f/d_y$ ; we can obtain

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{X_c}{Z_c} \\ \frac{Y_c}{Z_c} \\ 1 \end{bmatrix}. \quad (2)$$

Here,  $f_x, f_y, c_x, c_y$  are the internal parameters of color camera. In this paper, the internal parameters of left and right

cameras are obtained using the MATLAB Zhengyou Zhang calibration toolbox:

$$\begin{aligned} & [f_x, f_y, c_x, c_y]_{\text{left camera}} \\ & = [728.33946, 802.46120, 366.05260, 264.87057]. \\ & [f_x^r, f_y^r, c_x^r, c_y^r]_{\text{right camera}} \\ & = [725.16501, 798.07585, 343.09124, 219.13864]. \end{aligned} \quad (3)$$

As shown in Figure 3, the coordinates of the target point in the left view are  $(x_l, y_l)$ , the coordinates of the target point in the right view are  $(x_r, y_r)$ , binocular ranging mainly utilizes the differences of the target point horizontal coordinates on the direct images of two views (i.e., the disparity [18]:  $d = x_l - x_r$ ), and the disparity has an inverse relationship with distance  $Z$  from the target point to the image plane:

$$\begin{aligned} \frac{B}{Z} &= \frac{B-d}{Z-f} \implies \\ Z &= \frac{f \times B}{d}. \end{aligned} \quad (4)$$

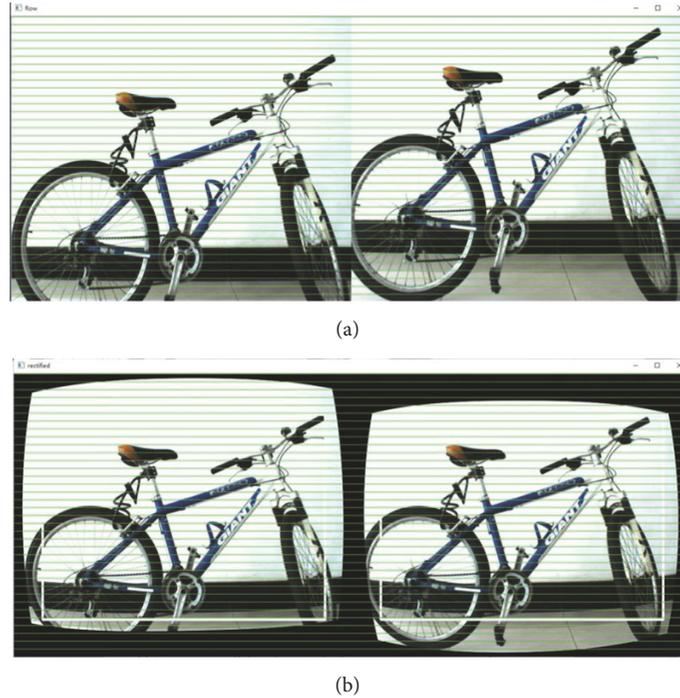


FIGURE 4: Line alignment correction: (a) camera images before the correction; (b) camera images after the correction.

Hence, the key of binocular ranging is to get the disparity map between left and right images. Various binocular stereo matching methods have been proposed to calculate the disparity maps. Before binocular stereo matching, binocular correction should be applied to align the lines of left and right images strictly, making the lines exactly at the same level. Therefore, the points of the left image and the corresponding points of right image have the same number of rows. Because of this, one-dimensional search of rows can find the corresponding point for stereo matching, which will save the cost of computation. Specifically, left and right images can be rectified to achieve line alignment using the relative rotation matrix  $R$  and translation matrix  $T$ , which are obtained using the MATLAB Zhengyou Zhang calibration toolbox in this paper:

$$R = \begin{bmatrix} 0.999957 & 0.003964 & 0.008409 \\ -0.003916 & 0.999976 & -0.005686 \\ -0.008431 & 0.005653 & 0.999948 \end{bmatrix}, \quad (5)$$

$$T = \begin{bmatrix} -47.36289 \\ -0.01925 \\ 0.47293 \end{bmatrix}.$$

As shown in Figure 4, it can be seen that the images after line alignment calibration can well achieve the line alignment. In this paper, then, the BM stereo matching algorithm is applied to obtain the disparity map because of the strict requirement about algorithm time.

**3.2. Depth Registration.** In the system, the model of the TOF depth camera is SR4000 camera, produced by the Swiss company Mesa Imaging. The TOF depth camera uses a modulator to emit infrared light from multiple angles. When the light returns to the sensor, it is possible to obtain the distance information for each pixel by the round trip time in real-time. The TOF depth camera can obtain depth information of the scene in real-time while the resolution of the camera is just  $144 \times 176$ . Hence, in this paper, a fusion system is developed to integrate depth camera and binocular stereo matching and improve the quality of 3D reconstruction.

Before fusion, the data on the depth map need to be aligned pixel by pixel to the color pixel. The registration process depends on the result of camera calibration [19, 20]. The parameters for registration are the conversion matrixes between the two coordinate systems [21, 22]. In this paper, an effective registration method is applied to align the depth image to the color image. Specifically, with the intrinsic parameters of the color camera we can mark the spatial position of each experimental checkerboard in the color camera coordinate system. Then, the depth camera can capture the position of the plate, which is coplanar with the checkerboard in physical space as shown in Figure 5. And, finally, we can obtain the projection matrix between the depth camera and the color camera by gradient descent optimization method.

Assume that the space point  $P$  is  $(X_c, Y_c, Z_c)$  under the color camera coordinate system and  $(X_d, Y_d, Z_d)$  under the depth camera coordinates.

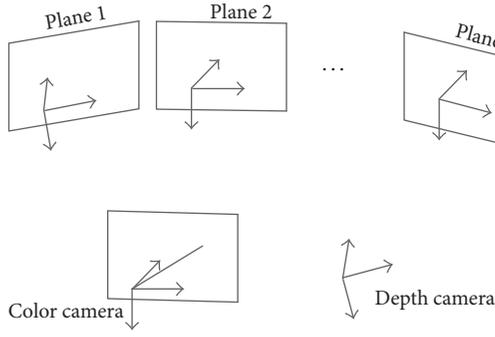


FIGURE 5: The comparison of the depth camera coordinate system and the color camera coordinate system.

Using formula (2), we can obtain

$$\begin{aligned} f_x \times \frac{X_c}{Z_c} + c_x &= u, \\ f_y \times \frac{Y_c}{Z_c} + c_y &= v. \end{aligned} \quad (6)$$

Then we can obtain formulas (7) and (8):

$$\frac{X_c}{Z_c} = \frac{u - c_x}{f_x}, \quad (7)$$

$$\frac{Y_c}{Z_c} = \frac{v - c_y}{f_y}. \quad (8)$$

Conversion formula between color camera and depth camera coordinate system:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R_d \begin{bmatrix} X_d \\ Y_d \\ Z_d \end{bmatrix} + T_d. \quad (9)$$

Here,  $R_d$  and  $T_d$  are the relative rotation and translation matrixes between left color camera and the depth camera.

Expanding formula (9), we can get

$$\begin{aligned} X_c &= r_{11}X_d + r_{12}Y_d + r_{13}Z_d + t_1, \\ Y_c &= r_{21}X_d + r_{22}Y_d + r_{23}Z_d + t_2, \\ Z_c &= r_{31}X_d + r_{32}Y_d + r_{33}Z_d + t_3. \end{aligned} \quad (10)$$

Combining formula (7) and (8), we can obtain

$$\begin{aligned} \frac{X_c}{Z_c} &= \frac{r_{11}X_d + r_{12}Y_d + r_{13}Z_d + t_1}{r_{31}X_d + r_{32}Y_d + r_{33}Z_d + t_3} = \frac{u - c_x}{f_x}, \\ \frac{Y_c}{Z_c} &= \frac{r_{21}X_d + r_{22}Y_d + r_{23}Z_d + t_2}{r_{31}X_d + r_{32}Y_d + r_{33}Z_d + t_3} = \frac{v - c_y}{f_y}. \end{aligned} \quad (11)$$

The extrinsic parameters  $R_d$  and  $T_d$  can be calculated from formulas (11) using gradient descent optimization method. The accuracy of  $R_d$  and  $T_d$  plays an important role in the accuracy of the system, since the depth information of the TOF camera is determined by the phase difference between

the emitting infrared light and reflected infrared light. If the reflected light intensity is higher, the credibility of the depth value is higher. By contraries, if the light is largely absorbed, the obtained depth value may be far from the true value. So the depth value of black block is less precise while the depth value of white block is more realistic. Therefore, we can only extract the depth data of the white checkerboard for subsequent processing. And we can extract depth value of the white block pixel by setting the gray level threshold.

Moreover, due to the noise of TOF camera points on the same plane will have a slight deviation. In order to obtain the one closest to the real value of the depth, we fit the three-dimensional coordinate data of the same template by using equation  $aX + bY + CZ + d = 0$ . The optimal equation coefficients  $a, b, c, d$  are obtained by least squares method [23, 24]. In order to get more accurate extrinsic parameters  $R_d$  and  $T_d$ , we take many groups of pictures to do the same operation above. Figure 6(a) reports some images of the chess board map of left color camera; Figure 6(b) shows gray images of the chess board map of depth camera; Figure 6(c) shows depth images of the chess board map of depth camera. After the calculation process above, the parameters  $R_d$  and  $T_d$  are obtained as follows:

$$\begin{aligned} R_d &= \begin{bmatrix} -1.0353572 & 0.0191776 & 0.0209886 \\ 0.00625004 & -0.8740292 & 0.0792625 \\ -0.0206831 & -0.0268092 & 0.7136297 \end{bmatrix}, \\ T_d &= \begin{bmatrix} -0.033684 \\ -0.021450 \\ 0.450148 \end{bmatrix}. \end{aligned} \quad (12)$$

Using  $R_d$  and  $T_d$ , the depth image can be registered to the color image effectively as shown in Figure 7.

**3.3. Depth Fusion Algorithm.** As shown in Figure 7(c), lots of depth regions need to be filled after the registration. Hence, an effective depth fusion algorithm is proposed in this paper as shown in Figure 8. Firstly, the bilinear interpolation is applied to the registered depth map. After interpolation, there still exist black holes to be filled as shown in the Figure 9(d). Then, a fusion method is proposed based on the result of binocular stereo matching to improve the accuracy. Specifically, we traverse the pixel of depth image after interpolation, when the pixel value is zero, and then look for the corresponding value of the binocular matching result as shown in Figure 9(c). If the value of binocular matching is valid, the result of binocular matching will be filled in the depth image (before filling, the disparity result of binocular stereo matching should be transformed into the actual depth value). Then, an initial fusion depth image is obtained as shown in Figure 9(e).

Moreover, a filling method based on mean-shift segmentation is applied further based on the following prior assumption:

- (1) The pixels with similar colors around a region are likely to have similar depth.
- (2) World surfaces are piecewise smooth.

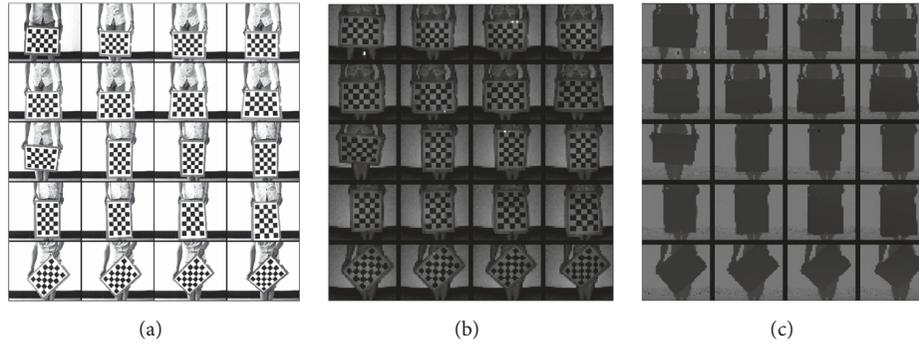


FIGURE 6: The images used to obtain the parameters  $R_d$  and  $T_d$  between left color camera and the depth camera. (a) Gray images of the chessboard map of left color camera; (b) images of the chessboard map of depth camera; (c) depth images of the chessboard map of depth camera.

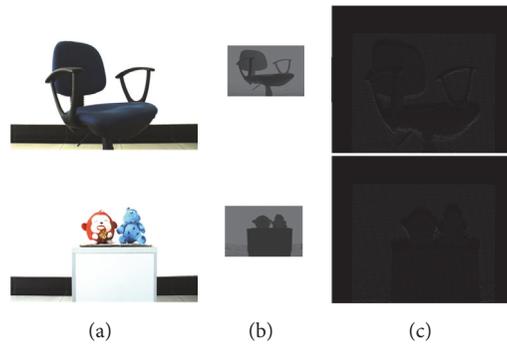


FIGURE 7: Depth registration: (a) the left color image; (b) the low-resolution depth image; (c) depth image registered with color image.

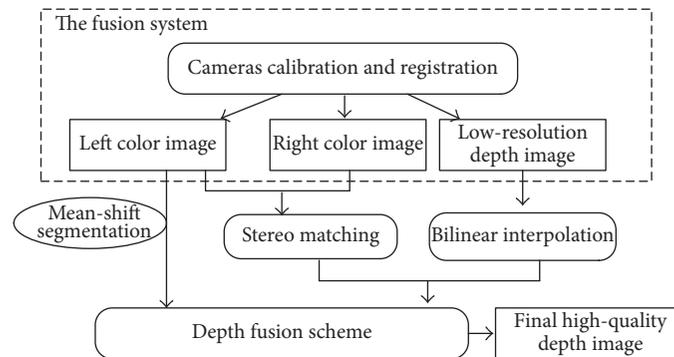


FIGURE 8: The framework of our fusion system.

Specifically, the mean-shift segmentation algorithm is used in the left color image. Then the pixels in the remaining black holes are assigned using the maximum probability value of the nonzero pixels in the same color segmentation region. Finally, the final depth image can be obtained after median filtering as shown in Figure 9(f). It can be seen that the quality of the final depth image is improved obviously.

#### 4. Experimental Results

The system utilizes a mobile robot, two of the same model Mercury (MERCURY) Series-High Speed 30 surveillance cameras made by Daheng, China (<http://www.daheng-image.com/>), and a Swiss Mesa Imaging SR4000 depth camera as

based hardware. Image processing system runs in VS2012 based on OpenCV2.4.9 and the industrial control panel is used as the image processing host. In the basis above, we tested the system proposed in this paper. After the system is initialized, the program starts running. First, the processing host reads the parameters into the memory, and then the system starts to capture the images and display images (only display in test mode). The host will match the images by using the parameters in memory and obtain a new three-dimensional reconstruction data. Finally, the system can complete the trajectory planning via these data to avoid obstacles.

Firstly, experiments are carried out to test the system. As shown in Figures 10–12, the system can obtain satisfying depth images after fusion.

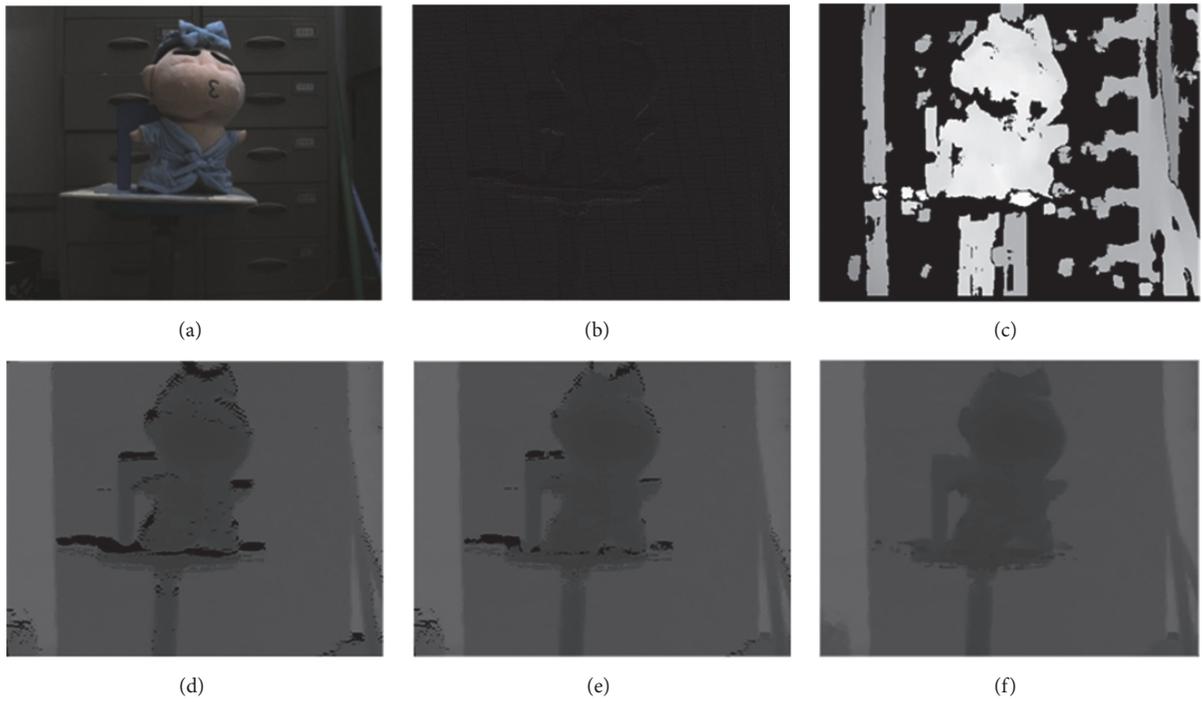


FIGURE 9: The effect of the proposed depth fusion algorithm: (a) the left color image; (b) the registered depth image; (c) disparity image of binocular stereo matching (BM algorithm); (d) depth image of interpolation; (e) the initial fusion depth image; (f) the final depth image.

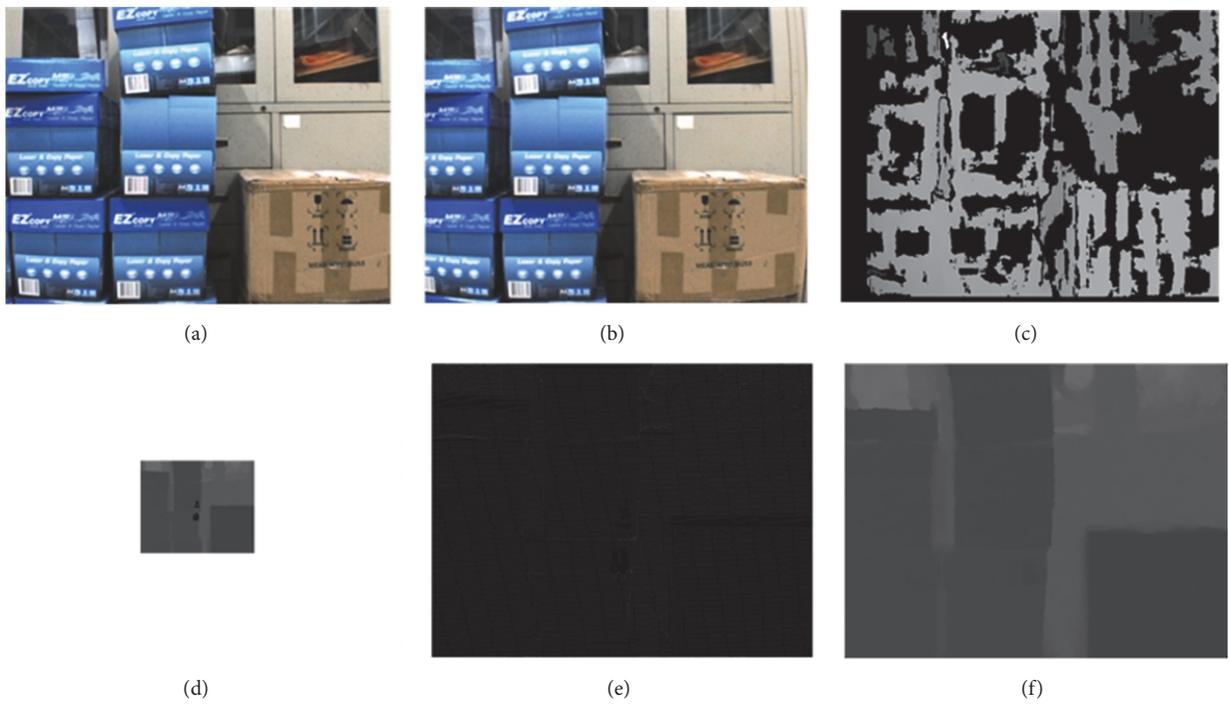


FIGURE 10: Result of the proposed system: (a) left color image; (b) right color image; (c) disparity map; (d) low-resolution depth image from depth camera; (e) depth image registered with the left color image; (f) the final fusion depth result.

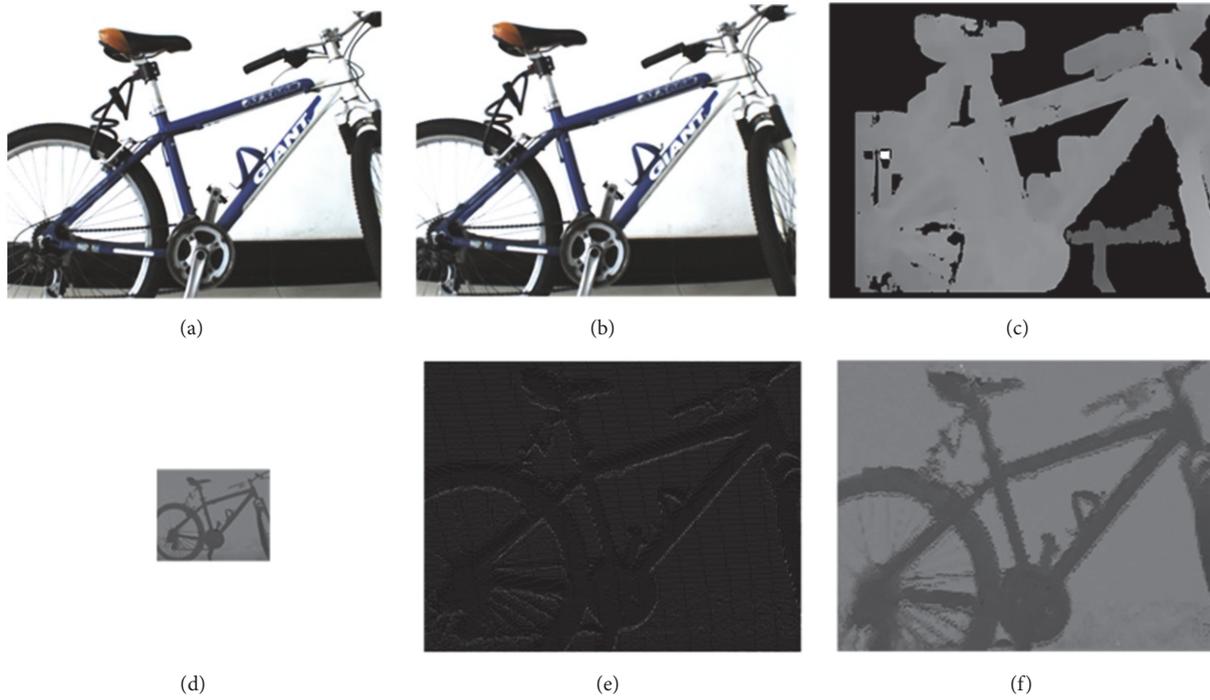


FIGURE 11: Result of the proposed system: (a) left color image; (b) right color image; (c) disparity map; (d) low-resolution depth image from depth camera; (e) depth image registered with the left color image; (f) the final fusion depth result.

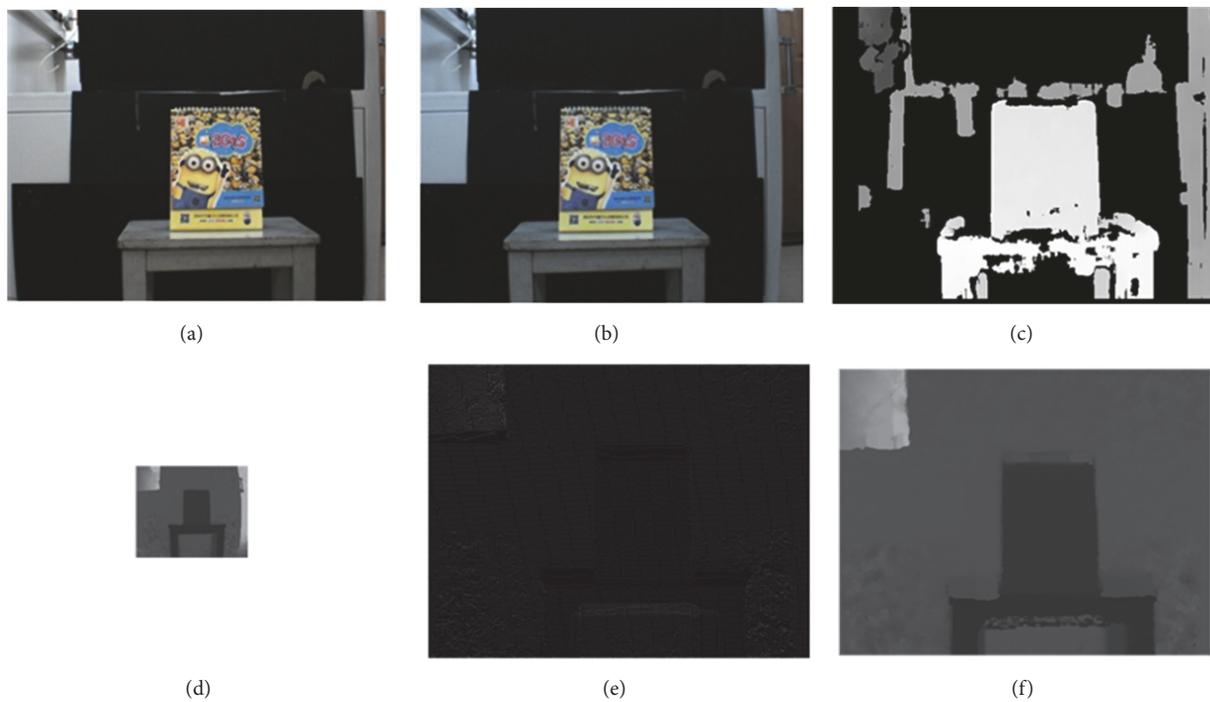


FIGURE 12: Result of the proposed system: (a) left color image; (b) right color image; (c) disparity map; (d) low-resolution depth image from depth camera; (e) depth image registered with the left color image; (f) the final fusion depth result.

TABLE 1: Accuracy experiments for the system.

| Regions | Object coordinates (cm) ( $x, y, z$ ) |                          |                       |
|---------|---------------------------------------|--------------------------|-----------------------|
|         | The coordinate of our system          | Actual coordinate        | Error                 |
| I       | (15.82, 67.48, 237.46)                | (15.71, 67.54, 237.40)   | (0.11, -0.06, 0.06)   |
|         | (13.05, 73.21, 257.63)                | (13.01, 73.09, 257.65)   | (0.04, 0.12, -0.02)   |
|         | (15.87, 66.72, 238.17)                | (15.75, 66.81, 238.09)   | (0.12, -0.09, 0.08)   |
| II      | (-27.11, 20.18, 238.90)               | (-27.23, 20.10, 239.10)  | (0.12, 0.08, -0.20)   |
|         | (-57.24, 20.33, 204.07)               | (-57.36, 20.40, 204.28)  | (0.12, -0.07, -0.21)  |
|         | (-76.99, 77.70, 273.17)               | (-77.15, 77.65, 273.05)  | (0.16, 0.05, 0.12)    |
| III     | (-4.43, -58.88, 206.88)               | (-4.70, -58.90, 206.90)  | (0.27, 0.02, -0.02)   |
|         | (-10.33, -55.01, 193.18)              | (-10.10, -55.00, 193.34) | (-0.23, -0.01, -0.16) |
|         | (-66.33, -58.17, 216.09)              | (-66.21, -58.36, 216.10) | (-0.12, 0.19, -0.01)  |
| IV      | (36.26, -18.20, 207.18)               | (36.30, -18.20, 207.25)  | (-0.04, 0, -0.07)     |
|         | (59.41, -18.18, 207.01)               | (59.40, -18.30, 207.00)  | (0.01, 0.12, 0.01)    |
|         | (73.41, -4.99, 209.19)                | (73.35, -5.00, 209.05)   | (0.06, 0.01, 0.14)    |

TABLE 2: Processing time.

| Serial number | Time/s   |
|---------------|----------|
| 1             | 0.521100 |
| 2             | 0.529700 |
| 3             | 0.596000 |
| 4             | 0.579333 |
| 5             | 0.549250 |
| 6             | 0.553500 |
| 7             | 0.547000 |
| 8             | 0.567750 |
| 9             | 0.514250 |
| 10            | 0.555250 |

Then, the accuracy of the system is compared in Table 1. “The coordinate of our system” is the coordinate value of an object obtained using our system. “Actual coordinate” is obtained from actual measurements (the depth camera coordinate is defined as the real-world coordinate). As shown in Table 1, the MSE of the measurement experiments is 0.019, 0.0076, and 0.0133 for  $X$ ,  $Y$ ,  $Z$  directions, respectively. The accuracy of our system can well meet the requirements of practical applications.

Then, we analyze the time performance of the system. As shown in Table 2, the average processing time is 0.5513133 seconds, which can well meet the requirements of actual applications, such as mobile robot avoidance.

In the system, software (i.e., image processing system) takes VS2012 as the development platform, which is based on OpenCV2.4.9. In order to observe the change of scene and view the effect of fusion, we code the program based on MFC. The program uses multithreaded control mode. The camera part is controlled by a separate thread to facilitate real-time acquisition, which easily refreshes the buffer of the image in real-time. The fusion part is controlled by another separate processing thread. Acquisition thread and processing thread will not affect each other, which can guarantee the high efficiency of processing and avoid delay. The MFC interface

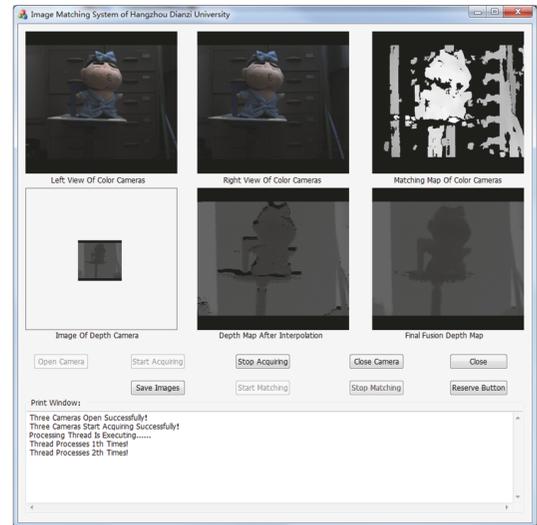


FIGURE 13: The running result of the system.

and running results are shown in Figure 13 (use button “Start Matching” for the choice of whether to start the integration; you can use the button “Stop Matching” to end the integration). The current system uses a multithreaded way to satisfy the strict requirement of time cost; later, we will try to use GPU acceleration method to make time cost lower.

## 5. Summary

In this paper, the visual processing section is used to acquire the three-dimensional scene data of the front of the mobile robot. The three cameras will transmit the image data and the 3D data of the depth camera to the image processing host via USB3.0. Image processing host uses calibration data to calibrate the images and maps the depth data to the left color image after dimensional registration. At the same time, the host will match the two color images by using correlation algorithm. And then the system will fuse the calibrated 3D

data of the depth camera with the matching result of the two color images to achieve the three-dimensional scene data with higher accuracy. Finally, the high precision 3D data is used to guide the mobile robot to avoid obstacles.

The system is proposed based on the following two observations: (1) Real-time and accuracy of binocular matching are a contradiction, and it is difficult to ensure that both real-time and high accuracy meet the requirements. It can be used in the scenes where the accuracy requirements are not very strict. But, it is hard to meet the requirements of high precision mobile robots. (2) The three-dimensional camera data of the TOF depth camera is accurate, and the real-time of the camera is very high too, but due to the structural constraints, the resolution is very low, only  $144 * 176$ , which makes it hard to meet the requirements of mobile robots either. However, the difficulty of this system is to accurately register the depth data to the left color picture. In order to reduce the impact of various objective factors such as noise, this paper calculates the  $R_d$  and  $T_d$  parameters matrix by taking multiple sets of images. The test results show that the system can meet the requirements of mobile robot trajectory planning. The system uses double threads processing method to improve efficiency of the system. And, later, we will try to use GPU acceleration method to ensure real-time.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by Natural Science Foundation of China (U1609216 and 61401129) and Zhejiang Natural Science Foundation of China (LY17F010020).

## References

- [1] C. Cadena, L. Carlone, H. Carrillo et al., "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.
- [2] N. Q. Ann, M. H. Achmad, L. Bayuaji, M. R. Daud, and D. Pebrianti, "Study on 3D scene reconstruction in robot navigation using stereo vision," in *Proceedings of the 2016 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, pp. 72–77, Selangor, Malaysia, October 2016.
- [3] S. A. Waskitho, A. Alfarouq, S. Sukaridhoto, and D. Pramadihanto, "FloW vision: Depth image enhancement by combining stereo RGB-depth sensor," in *Proceedings of the 5th International Conference on Knowledge Creation and Intelligent Computing, KCIC 2016*, pp. 182–187, November 2016.
- [4] A. A. Ahmed, M. K. Elbashir, and A. A. Osman, "Distance alert system using Stereo vision and feature extraction," in *Proceedings of the 2017 International Conference on Communication, Control, Computing and Electronics Engineering, ICCCEE 2017*, January 2017.
- [5] M. K. Singh, K. S. Venkatesh, and A. Dutta, "Accurate rough terrain modeling from fused 3D point cloud data," in *Proceedings of the 12th IEEE International Conference Electronics, Energy, Environment, Communication, Computer, Control, INDICON 2015*, December 2015.
- [6] O. S. Gedik and A. A. Alatan, "3D tracking using visual and range data," in *Proceedings of the 2012 20th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, Mugla, Turkey, April 2012.
- [7] W. Liu, Z. Li, L. Li, and F. Wang, "Parking Like a Human: A Direct Trajectory Planning Solution," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10.
- [8] P. Pharpatarara, B. Herisse, and Y. Bestaoui, "3-D trajectory planning of aerial vehicles using RRT\*," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 3, pp. 1116–1123, 2017.
- [9] J. Cheng, C. Leng, J. Wu, H. Cui, and H. Lu, "Fast and accurate image matching with cascade hashing for 3D reconstruction," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014*, pp. 1–8, June 2014.
- [10] K. Lu, X. Wang, Z. Wang, and X. Li, "Image-based 3d models reconstruction," in *Proceedings of the IET International Communication Conference on Wireless Mobile and Computing (CCWMC 2011)*, pp. 482–485, Shanghai, China, 2011.
- [11] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [12] A. Fetić, D. Jurić, and D. Osmanković, "The procedure of a camera calibration using Camera Calibration Toolbox for MATLAB," in *Proceedings of the 35th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2012*, pp. 1752–1757, hrv, May 2012.
- [13] X. Cao and H. Foroosh, "Camera calibration using symmetric objects," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3614–3619, 2006.
- [14] M. Shimizu and M. Okutomi, "Calibration and rectification for reflection stereo," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, June 2008.
- [15] V. Gandhi, J. Čech, and R. Horaud, "High-resolution depth maps based on TOF-stereo fusion," pp. 4742–4749.
- [16] Y.-S. Kang and Y.-S. Ho, "Disparity map generation for color image using TOF depth camera," in *Proceedings of the 5th 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 3DTV-CON 2011*, May 2011.
- [17] Y.-S. Kang and Y.-S. Ho, "High-quality multi-view depth generation using multiple color and depth cameras," in *Proceedings of the 2010 IEEE International Conference on Multimedia and Expo, ICME 2010*, pp. 1405–1408, July 2010.
- [18] P. An, Z. Zhang, and L. Shi, "Theory and experiment analysis of disparity for stereoscopic image pairs," in *Proceedings of the 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing, ISIMP 2001*, pp. 68–71, May 2001.
- [19] J. Jung, J.-Y. Lee, Y. Jeong, and I. S. Kweon, "Time-of-Flight Sensor Calibration for a Color and Depth Camera Pair," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 7, pp. 1501–1513, 2015.
- [20] F. Basso, A. Pretto, and E. Menegatti, "Unsupervised intrinsic and extrinsic calibration of a camera-depth sensor couple," in *Proceedings of the 2014 IEEE International Conference on Robotics and Automation, ICRA 2014*, pp. 6244–6249, June 2014.
- [21] H. Yu, K. Zhao, Y. Wang, L. Kan, M. Sun, and W. Jia, "Registration and fusion for ToF camera and 2D camera reading," in *Proceedings of the 2013 Chinese Automation Congress, CAC 2013*, pp. 679–684, November 2013.

- [22] J. Kim, S. Park, S. Kim, and S. Lee, "Registration method between ToF and color cameras for face recognition," in *Proceedings of the 2011 6th IEEE Conference on Industrial Electronics and Applications, ICIEA 2011*, pp. 1977–1980, June 2011.
- [23] Y. Zheng and S. Peng, "A practical roadside camera calibration method based on least squares optimization," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 2, pp. 831–843, 2014.
- [24] D. Buchczik and W. Ilewicz, "Evaluation of calibration results using the least median of squares method in the case of linear multivariate models," in *Proceedings of the 21st International Conference on Methods and Models in Automation and Robotics, MMAR 2016*, pp. 800–805, September 2016.



**Hindawi**

Submit your manuscripts at  
<https://www.hindawi.com>

