J Inf Process Syst, Vol.16, No.1, pp.155~170, February 2020
https://doi.org/10.3745/JIPS.03.0132

ISSN 1976-913X (Print)
ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Passive Ranging Based on Planar Homography in a Monocular Vision System

Xin-mei Wu*, Fang-li Guan**, and Ai-jun Xu*

Abstract

Passive ranging is a critical part of machine vision measurement. Most of passive ranging methods based on machine vision use binocular technology which need strict hardware conditions and lack of universality. To measure the distance of an object placed on horizontal plane, we present a passive ranging method based on monocular vision system by smartphone. Experimental results show that given the same abscissas, the ordinatesis of the image points linearly related to their actual imaging angles. According to this principle, we first establish a depth extraction model by assuming a linear function and substituting the actual imaging angles and ordinates of the special conjugate points into the linear function. The vertical distance of the target object to the optical axis is then calculated according to imaging principle of camera, and the passive ranging can be derived by depth and vertical distance to the optical axis of target object. Experimental results show that ranging by this method has a higher accuracy compare with others based on binocular vision system. The mean relative error of the depth measurement is 0.937% when the distance is within 3 m. When it is 3–10 m, the mean relative error is 1.71%. Compared with other methods based on monocular vision system, the method does not need to calibrate before ranging and avoids the error caused by data fitting.

Keywords

Corner Detection, Depth Extraction Model, Monocular Vision, Passive Ranging, Planar Homography

1. Introduction

As a key parameter in object positioning, distance measuring has been widely studied in many areas, like 3D reconstruction, new military technology for high technique weapons, and so on [1-4]. Traditional ranging method, such as tape and total station, are time and labor intensive and inefficient. With the development of laser radar and machine vision, non-contact measurement methods have emerged [5,6]. These methods are mainly divided into active and passive ranging [7-9]. Laser scanning is one of the mainstream active ranging methods [10,11]. It has a higher measurement accuracy and can be used to describe the 3D structure of one object [12]. However, for general public who are not expert in this field, this kind of active measurement instrument is limited. It requires expert knowledge, which limits its' use in daily practice. Passive ranging can also be realized based on machine vision. It estimates distance and

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received July 30, 2018; first revision December 11, 2018; accepted December 30, 2018.

Corresponding Author: Ai-jun Xu (ajxu_zafu@163.com)

* School of Information Engineering, Zhejiang Agriculture and Forestry University, Hangzhou, China; Zhejiang Provincial Key Laboratory of Forestry Intelligent Monitoring and Information Technology, Hangzhou, China; Key Laboratory of State Forestry and Grassland Administration on Forestry Sensing Technology and Intelligent Equipment, Hangzhou, China (956617584@qq.com, ajxu_zafu@163.com)

** State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China (372157342@qq.com)

Copyright© 2020 KIPS

obtains object size through image pixel information and camera imaging principles [13,14]. It has the advantages of rich image information and low cost. Machine vision measurement includes both monocular vision and binocular vision measurements [15,16]. The early image information extraction methods were mostly based on the binocular stereo vision principle or camera motion information, and required multiple images to extract the depth [17-19]. In contrast, monocular vision method does not require strict hardware conditions during image acquisition and allows for device integration. Recently, researches on information extraction based on monocular vision have been gradually progressed. Liu et al. [20] designed a method to extract depth map from video based on non-parametric fusion of multiple cues. This method combined with clues, including the image contour, geometrical perspective and space-time correlation among contours to estimate a more accurate video depth. The depth information of the whole image could be obtained by monocular depth clues, and the algorithm did have a simple structure. However, its application might be limited because it needs prior information such as the scene structure of the image.

The calibration methods based on monocular vision system, which involve the camera’s intrinsic and extrinsic parameters, can also be used to obtain the depth information [21-23]. When combined with a camera projection model, camera calibration can be used to study the conversion relationship between the image coordinate system and the world coordinate system. This method requires more than three checkerboard images in different orientations, and records the corresponding coordinate of each point in the world coordinate system and image coordinate system. Thus, calibration has a great influence on measurement accuracy. Wu et al. [24] established a mathematical model to fit the mapping relationship between the object distance and pixel, and used this relationship to extract depth. The accuracy of this method may be affected by long-distance measurement and data fitting. Huang et al. [25] proposed a method to obtain the depth information by detecting the corner point of the vertical checkerboard image and establishing the mapping relationship between ordinate pixel and actual imaging angle. Because different cameras have different intrinsic parameters, the model established by this method had poor applicability and could not calculate the target distance in any direction.

Based on the above analysis of depth extraction and distance measurement methods, given the target contour, we present a method for depth estimating and passive ranging. To investigate the mapping relationship between the ordinate pixel of the image point and actual imaging angle of its corresponding object point, we do the following works: first, we combine the corner detection method proposed by Andreas Geiger and the cornerSubPix() function provided by OpenCV to extract sub-pixel corners. Then, the Pearson correlation analysis is used to verify the relationship between the actual imaging angle of the object point and ordinate pixel of the corresponding image point for different models of cameras and rotation angles. Experiment results show that given the same abscissas, the ordinatesis of the image points linearly related to their actual imaging angles. So, according to this principle, the actual imaging angles and ordinate pixels of the special conjugate points are substituted into the assumed linear function. Then we can calculate the constant coefficients. And a depth extraction model suitable for different models of smartphones is established. Furthermore, by substituting the intrinsic parameters of the camera and the ordinate pixel of the target point into the camera calibration model, we can calculate the depth of any image point. Finally, the vertical distance from target object to optical axis of the camera is calculated by the principle of camera stereo imaging system, and the distance from target object to the camera is calculated according to Pythagorean theorem. The main difference between our study and the others mentioned above can be summarized as: it can measure depth by a smartphone which is portable and flexible. It leads this method to be practicable in daily works. The research is also of great significance for the autonomous obstacle avoidance and path planning of unmanned vehicles in horizontal roads, remote monitoring of unmanned sweeping vehicles, and automatic measurement of tree factors in forestry resource survey.

2. Principle of Passive Ranging based on Monocular Vision System

The image is collected by camera of smartphone. To calculate the distance from any point on the horizontal ground to the camera, we first use Pearson correlation analysis method to prove that given the same abscissas, the ordinatesis of the image points linearly related to their actual imaging angles. Then, we establish a depth extraction model by assuming a linear function and substituting the actual imaging angles and ordinate pixels of the special conjugate points into the function. Furthermore, by substituting the intrinsic parameters of camera obtained from camera calibration and the ordinate pixel into that model, we can calculate the depth of target object. According to the camera imaging principle, the vertical distance from any point to the optical axis can be calculated. Finally, the Pythagorean theorem is used to derive the distance from any points to the camera plane.

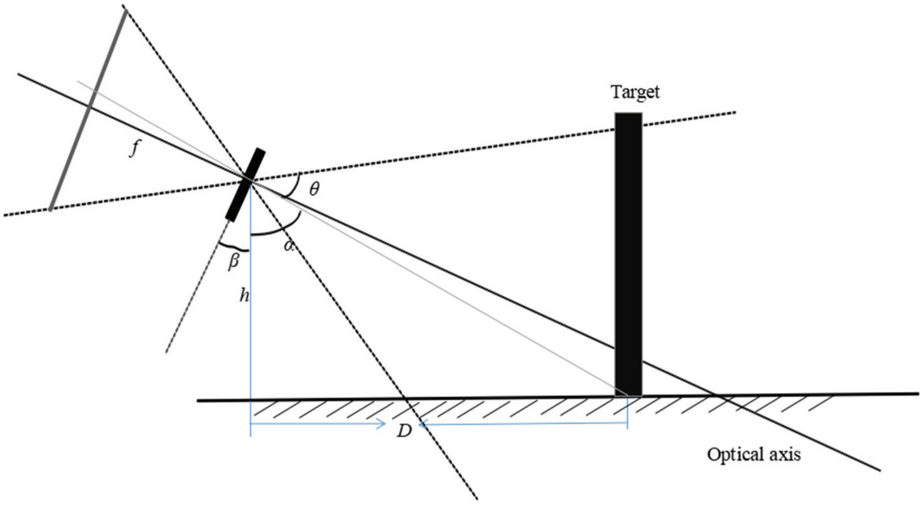


Fig. 1. Projection geometry model of shoot.

The projection geometry model of image acquisition is shown in Fig. 1. Where θ denotes half of the camera's field of view, f denotes the focal length, and h denotes the height of camera. The camera rotation angle β derived from the gravity sensor of the smartphone. α denotes the object's actual imaging angle. D denotes the depth of a target object. As is shown in Fig. 1, ignoring the nonlinear distortion, the depth of any target object can be derived:

$$D = h \cdot \tan \alpha \tag{1}$$

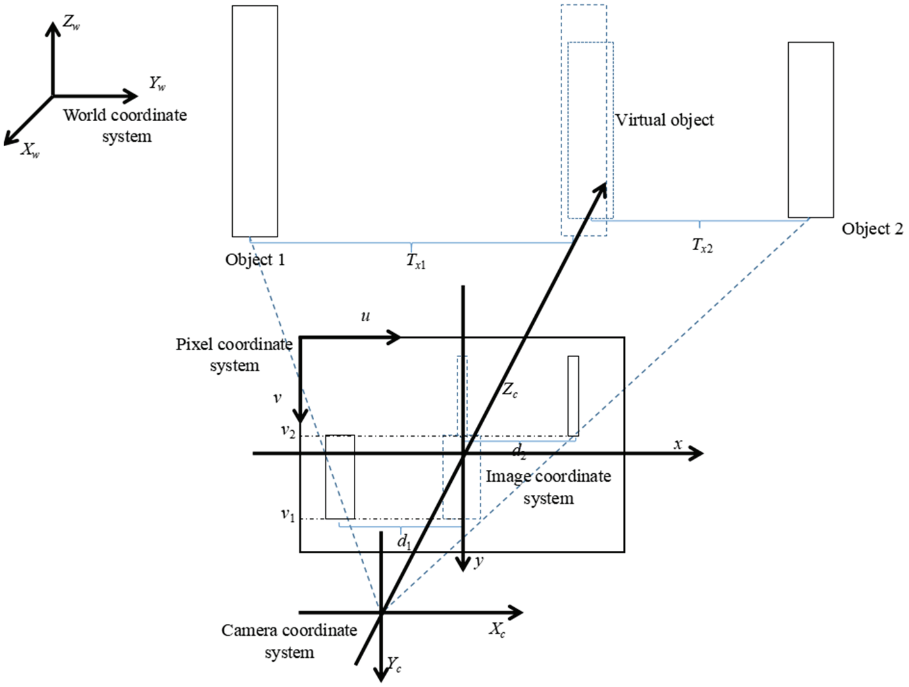


Fig. 2. The diagram of each coordinate system in the Pinhole model.

Fig. 2. shows the relationship of each coordinate system in the pinhole model. To calculate the distance from any point to the camera plane, according to the depth that has been calculated above, we also need to calculate the vertical distance T_x from the target object to optical axis (in Fig. 2, T_x denotes the distance from the target object to its corresponding virtual object placed on the optical axis). Then T_x can be expressed as:

$$T_x = \frac{d * D}{f} \tag{2}$$

where d denotes the parallax between the target object and its corresponding virtual object in the image plane. According to the Pythagorean theorem, we can calculate the distance from the target object to camera L:

$$L = \sqrt{T_x^2 + D^2} \tag{3}$$

3. Checkerboard Design and Corner Detection

3.1 Design of Checkerboard

While detecting and extracting corners, a perspective transformation may lead to an inaccurate corner detection and extraction results. To counter this problem, based on a traditional checkerboard [26], we improve it with a fixed cell width and increased length. Empirically, the checkerboard is found to be sufficient for corners detection over a wide range of perspective transformations.

We extract the corner of the traditional checkerboard tiled on the ground horizontally which has equal length and width, and analyze the relationship between the distances of each two adjacent rows and ordinate pixels of the corners in the same line. Then we can calculate the increment cell length according to the relationship. The two adjacent row of the new checkerboard has an equal pixel difference in the image (the image is acquired when the camera rotation angle equals 0). This checkerboard can improve the accuracy of long-distance corner extraction with a wide range of perspective transformation.

In order to calculate the length increment between two adjacent rows, we design six groups of experiments and extract the traditional checkerboard that contains 45×45 mm cells. Then we calculate the actual distance of the unit pixel between two adjacent rows in the world coordinate system. To make sure the same ordinate pixel differences of two adjacent rows, the length of each cells y_i in new checkerboard is shown in Table 1. Let x_i be the distance from the ith row of corners in traditional checkerboard to the camera, and the length difference Δd_i of two adjacent rows can be expressed as:

$$\Delta d_i = \mathrm{y}_{i+1} - y_i = \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \cdot \Delta x_i \tag{4}$$

Suppose the relationship between y_i and x_i is $f(x)$, according to formula (4), we can get:

$$\Delta \hat{d} = \Delta x \cdot f'(x) \tag{5}$$

Table 1. Computing length of each grid

Numbers	I_1	I_2	I_3	I_4	I_5	I_6
y_{1}	45.00	45.00	45.00	45.00	45.00	45.00
y_{2}	50.35	51.05	51.59	49.78	48.74	47.91
y_{3}	56.40	55.36	64.33	60.08	61.59	60.99
y_{4}	66.34	72.48	61.24	65.99	69.91	73.62
y_{5}	70.03	65.47	81.26	67.28	82.31	77.22
y_{6}	94.91	78.90	90.26	83.54	81.74	99.12
y_{7}	79.09	99.68	97.33	100.45	104.27	103.76
y_{8}	114.10	107.99	106.91	100.23	127.06	109.91
y_{9}	122.64	119.63	116.49	120.33	124.32	118.93
y_{10}	117.24	137.79	129.57	130.12	136.04	110.40
y_{11}	137.94	119.36	123.13	139.19	123.29	137.63
y_{12}	153.43	154.18	155.12	162.88	151.05	152.64
y_{13}	147.57	157.29	194.69	150.60	163.18	183.68
y_{14}	154.58	171.54	181.73	172.29	162.24	195.40
y_{15}	172.33	192.72	189.07	192.20	190.60	192.52
y_{16}	226.15	207.61	193.33	212.60	219.43	207.77
y_{17}	239.35	194.01	218.59	193.19	225.88	199.82
y_{18}	231.79	237.05	215.06	230.12	236.40	210.73
y_{19}	255.18	255.28	239.22	245.36	246.67	258.52
y_{20}	276.53	225.38	251.38	269.27	283.88	278.13
y_{21}	264.24	242.15	276.81	278.70	294.44	282.98
y_{22}	252.99	269.33	297.21	312.88	316.30	287.61

According to Pearson correlation analysis, there is a highly significant linear correlation between the length of each cells in new checkerboard and the distance from the *i*th row of corners in traditional checkerboard to the camera (p<0.01), and the correlation coefficient r is 0.975. The least squares method is used to calculate the derivative of $f(x)$, $f'(x)$=0.262.

Therefore, when the checkerboard's first row contains $d*d$ mm cells, the remaining rows are fixed in width, and the length difference Δd between two adjacent rows is 0.262×d mm. The new checkerboard is shown in Fig. 3. Corners of this checkerboard can be accurately extracted. Furthermore, the influence of the perspective transformation can be reasonably avoided.

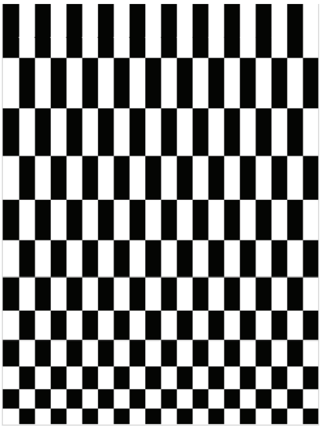


Fig. 3. New checkerboard.

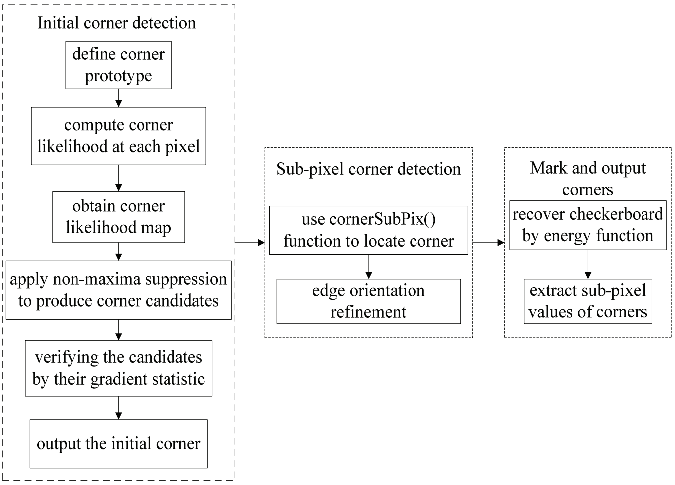

Initial corner detection
define corner prototype
compute corner likelihood at each pixel
obtain corner likelihood map
apply non-maxima suppression to produce corner candidates
verifying the candidates by their gradient statistic
output the initial corner
Sub-pixel corner detection
use cornerSubPix() function to locate corner
edge orientation refinement
Mark and output corners
recover checkerboard by energy function
extract sub-pixel values of corners

Fig. 4. Implementation process of corner detection algorithm.

3.2 Corner Detection Algorithm

While taking photos on the horizontal ground, due to the perspective transformations, traditional corner detection algorithms, such as Harris and Stephens [27] and Shi [28], are poor in robustness. Additionally, these methods also fail to detect corners when the smartphone rotates at a large angle. Therefore, we optimize Geiger's corner detection method [29] to extract sub-pixel corners. The corner detection algorithm implementation process is shown in Fig. 4.

The algorithm does not require the size of cells and checkerboards when detect corners, and it is robust enough when extract corners from images with high distortion. The corner extraction results are shown in Table 2.

Table 2. Result of sub-pixel corner point detection

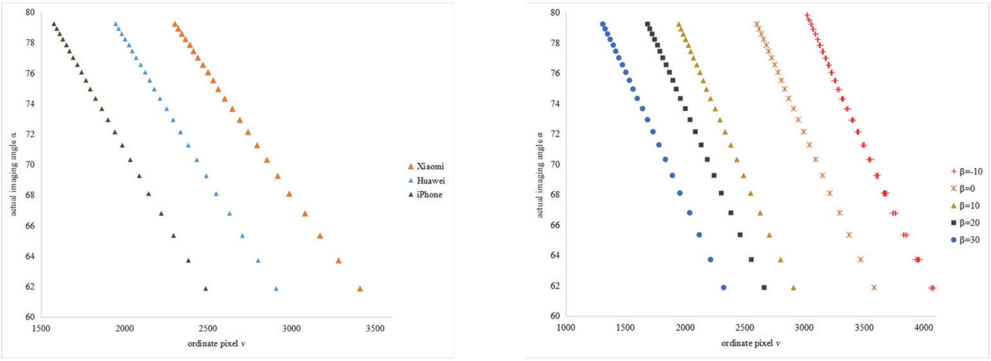
ID	Initial corner	Sub-pixel corner	Distance (mm)
1	(158, 3454)	(158.429, 3453.30)	895
2	(242, 3377)	(242.818, 3377.01)	958
3	(332, 3295)	(331.134, 3294.45)	1034
4	(418, 3216)	(418.893, 3215.49)	1123
5	(506, 3135)	(506.277, 3135.20)	1186
6	(590, 3057)	(589.468, 3057.80)	1288
7	(668, 2986)	(667.982, 2985.12)	1403

4. Passive Ranging Model based on Monocular Vision

4.1 Correlation Analysis

Three smartphones, like Xiaomi, Huawei, and iPhone are selected to analyze the relationship among actual imaging angle of object point α, the ordinate of image point v, and rotation angle of camera β. The camera rotation angle β are set as -10°, 0°, 10°, 20°, 30°, respectively. The corner detection algorithm mentioned in Section 3 is used to extract pixels, and we use SPSS version 22 do regression analysis based on these data. The results are shown in Fig. 5: Fig. 5(a) shows the relationship of the ordinate pixels and actual imaging angles for three different models of smartphones when β = 10°; Fig. 5(b) shows the relationship between ordinate pixel values and imaging angles for different camera rotation angles.

As can be seen from Fig. 5, the actual imaging angle of the object point decreases as the ordinate pixel of the corresponding image point increase. And for varying rotation angles and smartphones, the relationship between ordinate pixel and actual imaging angle are different. Additionally, given the same abscissas, the ordinatesis of the image points linearly related to their actual imaging angles, where p<0.01 and the correlation coefficient $r \geqslant 0.99$.



(a)

(b)

Fig. 5. Relationships of image ordinate pixels and actual imaging angles: (a) for three different models of smartphones and (b) for different camera rotation angles.

4.2 Passive Ranging Method

4.2.1 Camera intrinsic parameters acquisition

In photogrammetry, to determine the projection transition between the coordinate systems in the pinhole model, it is necessary to use camera parameters to construct a projection geometric model. We combine Zhang's calibration method [30] and a camera calibration model with nonlinear distortion term to calibrate camera of smartphone. It can correct nonlinear distortions and acquire camera intrinsic parameters.

According to the pinhole camera imaging principle, image points have the following relationship in the image plane coordinate system and the pixel coordinate system:

$$\begin{cases} u = x / d_x + u_0 \\ v = y / d_y + v_0 \end{cases} \tag{6}$$

where d_x, d_y (unit: mm) denotes the length and width of pixel on the image plane, respectively. Since a pixel projected on the image plane is a rectangle, the length and width of each physical pixel cannot be kept consistent, d_x is not equal to d_y. (u_0, v_0) denote the origin o of the image plane coordinate system in the pixel coordinate system. In the camera coordinate system, point P_c (X_c, Y_c, Z_c) is projected on the image coordinate system (x, y, f). The image plane is perpendicular to the optical axis, and the distance from the origin to the image plane is f. According to the principle of similar triangles, we get:

$$\begin{cases} x = f \cdot X_c / Z_c \\ y = f \cdot Y_c / Z_c \end{cases} \tag{7}$$

The transformation from the world coordinate system P_W (X_W, Y_W, Z_W) to camera coordinate system P_c is a rigid body motion, including translation and rotation. So from world coordinate system to camera coordinate system:

$$P_c = \boldsymbol{R} \cdot (P_W - \boldsymbol{C}) = \boldsymbol{R} \cdot P_W + \boldsymbol{T} \tag{8}$$

Combining Eqs. (6) to (8), the relationship of the coordinate system can be expressed by homogeneous coordinates and matrix as:

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} {}^{1}\!/_{d_x} & 0 & u_0 \\ 0 & {}^{1}\!/_{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \boldsymbol{R} & \boldsymbol{T} \\ 0^T & 1 \end{pmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \boldsymbol{R} & \boldsymbol{T} \\ 0^T & 1 \end{pmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = \boldsymbol{M}_{\text{int}} \boldsymbol{M}_{\text{ext}} \boldsymbol{P}_{\boldsymbol{W}} \tag{9}$$

where $\boldsymbol{M}_{\text{int}}$ denotes the camera intrinsic parameters and $\boldsymbol{M}_{\text{ext}}$ denotes the extrinsic parameters. Camera external parameters include rotation matrix $\boldsymbol{R}$ and translation matrix $\boldsymbol{T}$.

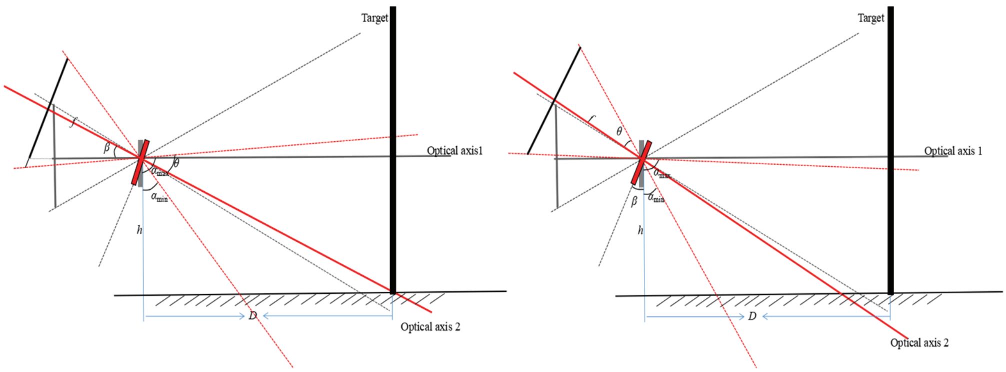
4.2.2 Depth extraction model

For different models of smartphone and camera rotation angles, the image points' ordinates and the actual imaging angles of the corresponding object points are extremely negatively linearly related. Thus, we get:

$$\alpha = F(v, \beta) = k \cdot v + b \tag{10}$$

The constant coefficients k and b are related to the camera rotation angle β. The camera projection geometric model is shown in Fig. 6. As can be seen from Fig. 6, when an object point is projected on the bottom of image, its α takes the minimum value 90–θ–β, while v takes the effective number of pixels in column coordinates of the image sensor. Then, we have:

$$90 - \beta - \theta = k \cdot v_{\max} + b \tag{11}$$



(a) (b)

Fig. 6. Projection geometry model of shoot: (a) projection geometry model of shoot with FOV above horizontal line and (b) projection geometry model of shoot with FOV under horizontal line.

When $\alpha_{\min}$ + 2θ > 90°, the field of view (FOV) of the camera is above horizontal line—projection geometry model of shoot is shown in Fig. 6(a), α takes the maximum value 90°, v infinitely close to v_0-tanβ*f_y. If the camera rotates counterclockwise, α and v takes the same values. Additionally, when $\alpha_{\min}$ + 2θ < 90°, the FOV is lower than the horizon—projection geometry model of shoot is shown in Fig. 6(b), the maximum of actual imaging angle $\alpha_{\max}$ = 90–β+θ, and v = 0. Therefore, substituting into formula (10) results in:

$$\begin{cases} 90 = k \cdot \left(v_0 - \tan\beta \cdot f_y\right) + b, & \theta > \beta \\ 90 - \beta + \theta = b\,, & \theta < \beta \end{cases} \tag{12}$$

According to the construction principle of the pinhole camera, the tangent value of θ is equal to half the length of the camera CMOS or CCD image sensor L_{CMOS} divided by the camera focal length f. The physical unit is converted into pixel units to calculate θ:

$$\theta = \arctan \frac{L_{CMOS}}{2 \cdot f} = \arctan \frac{v_{\max}}{2 \cdot f_y} \tag{13}$$

Therefore, combining (5)~(8), $F(\alpha, \beta)$ can be obtained:

$$F(v,\beta) = \begin{cases} \alpha = -\dfrac{\theta+\beta}{v_{\max} - v_0 + \tan\beta \cdot f_y} v + 90 + \dfrac{(v_0 - \tan\beta \cdot f_y)\cdot(\theta+\beta)}{v_{\max} - v_0 + \tan\beta \cdot f_y} \pm \delta & \theta > \beta \\ \alpha = -\dfrac{2\theta}{v_{\max}} v + 90 + \theta - \beta \pm \delta & \theta < \beta \end{cases} \tag{14}$$

The imaging principle of smartphone’s camera lens is pinhole imaging whose object point, image point as well as camera optical center are in one line. However, because of the manufacturing error, it is actually not an ideal linear model that lead to nonlinear distortion of the image and δ in Eq. (14) is its distortion parameter.

Then, the depth extraction model can be established by combining formula (14) and (1):

$$D=\begin{cases} h\cdot\tan(-\dfrac{\arctan\dfrac{v_{\max}}{2f_y}+\beta}{v_{\max}-v_0+\tan\beta\cdot f_y}v+90+\dfrac{(v_0-\tan\beta\cdot f_y)\cdot(\arctan\dfrac{v_{\max}}{2f_y}+\beta)}{v_{\max}-v_0+\tan\beta\cdot f_y})\pm\delta' & \theta>\beta \\ h\cdot\tan(-\dfrac{2\arctan\dfrac{v_{\max}}{2f_y}}{v_{\max}}v+90+\arctan\dfrac{v_{\max}}{2f_y}-\beta)\pm\delta' & \theta<\beta \end{cases} \tag{15}$$

4.2.3 Distance measurement

Based on the depth of the target object derived from above, we also need to calculate the vertical distance T_x from the target object to the optical axis. Fig. 7 is a schematic diagram of a camera stereo imaging system, where point P denotes the camera position, and line segment AB is parallel to the image plane. Let coordinates of A be (X, Y, Z) in the camera coordinate system. And the coordinates of point B are (X+T_x, Y, Z) in the camera coordinate system. Points A and B are projected on the image plane, where A' (x_l, y_l), B' (x_r, y_r). According to formula (7):

$$\begin{cases} x_l = f\cdot X/Z, & y_l = f\cdot Y/Z \\ x_r = f\cdot(X+T_x)/Z, & y_r = f\cdot Y/Z \end{cases} \tag{16}$$

Combining Eqs. (6) and (16), The horizontal parallax d of the two points A' and B'—with the same Y and depths—can be expressed as:

$$\begin{aligned} d &= x_r - x_l \\ &= f\cdot(X+T_x)/Z - f\cdot X/Z = f\cdot T_x/Z \\ &= (u_r-u_0)\cdot d_x - (u_l-u_0)\cdot d_x = (u_r-u_l)\cdot d_x \end{aligned} \tag{17}$$

Therefore, given camera focal length f, image center point (u_0, v_0), the physical size d_x of each pixel in the x-axis on the image plane and depth of the target object, the vertical distance T_x from the target object to the optical axis can be calculated:

$$T_x = \frac{|u-u_0|\cdot d_x\cdot D}{f} \tag{18}$$

According to formula (3), we can obtain the distance L from the target object to the projection point of the camera:

$$L=\sqrt{T_x^{\,2}+D^2}=\sqrt{\left(\left.|u-u_0|\cdot D\middle/f_x\right.\right)^2+(h\cdot\tan\alpha)^2} \tag{19}$$

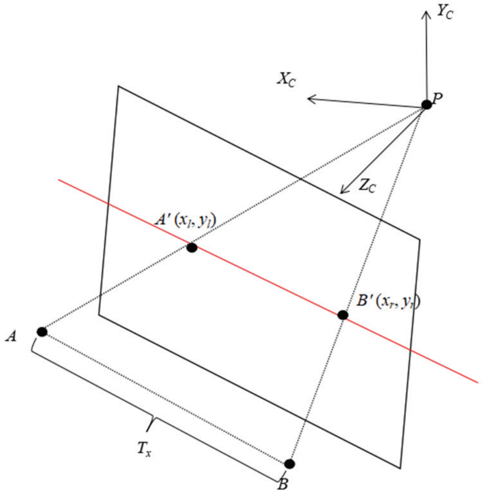


Fig. 7. Principle of camera stereoscopic imaging system.

5. Experiment Result and Discussion

To verify the feasibility and accuracy of the passive ranging method, we conducted experiments used Xiaomi 3 (MI 3) smartphone. Java combined with C++ were used to write a passive ranging application for smartphones. After the application was written and debugged according to the above method, the accuracy of the depth extraction model and passive ranging were verified separately in the laboratory and natural environment.

The intrinsic parameters of the camera are: f_x = 3486.5637, u_0 = 1569.0383, f_y = 3497.4652, v_0 = 2107.98988, and the image resolution is 3120 × 4208. Substituted the parameters into the model to get the specific depth extraction model of the camera:

$$D = \begin{cases} h \cdot \tan\left(\dfrac{31.03 + \beta}{2100.01 + 3497.47\tan\beta}(2107.99 - 3497.47\tan\beta - v) + 90\right) \pm \delta' & \theta > \beta \\ h \cdot \tan(-0.0147v + 121.03 - \beta) \pm \delta' & \theta < \beta \end{cases} \tag{20}$$

Table 4. Data of depth measurement

Group	Ordinate pixel v	Actual imaging angle α (°)	Calculated imaging angle α' (°)	Angle error (°)	True depth D (mm)	Calculated depth D' (mm)	Depth error (mm)	Relative error (%)
I_1	4075.79	60.81	60.92	0.11	546	548.50	2.50	0.458
	3840.65	64.38	64.39	0.01	636	636.52	0.52	0.082
	3646.92	67.21	67.26	0.05	726	727.71	1.71	0.236
	3490	69.51	69.58	0.07	816	819.21	3.21	0.393
	3364.75	71.39	71.43	0.04	906	907.85	1.85	0.205
	3257.5	72.97	73.01	0.04	996	998.52	2.52	0.253
I_2	3144.6	74.61	74.68	0.07	1035	1040.36	5.36	0.517
	3079.58	75.58	75.64	0.06	1108	1113.23	5.23	0.472
	3013.13	76.58	76.62	0.04	1194	1198.16	4.16	0.348
	2948.42	77.57	77.581	0.011	1293	1294.21	1.21	0.093

Table 4. (Continued)

Group	Ordinate pixel *v*	Actual imaging angle α (°)	Calculated imaging angle α' (°)	Angle error (°)	True depth *D* (mm)	Calculated depth D' (mm)	Depth error (mm)	Relative error (%)
	2885.83	78.22	78.50	0.28	1366	1400.82	34.82	2.549
	2827.4	79.09	79.36	0.27	1478	1517.03	39.03	2.640
	2772.96	79.92	80.17	0.25	1603	1644.84	41.84	2.610
	2722.87	80.71	80.91	0.2	1741	1781.31	40.31	2.315
	2676.69	81.44	81.59	0.15	1892	1927.69	35.69	1.886
	2635.39	82.11	82.20	0.09	2056	2080.55	24.55	1.194
	2597.41	82.73	82.76	0.03	2233	2243.41	10.41	0.466
	2562.52	83.30	83.28	0.02	2423	2418.80	4.20	0.173
	2530.99	83.80	83.75	0.05	2626	2602.32	23.68	0.902

5.1 Ranging in Laboratory

In experiment 1, the camera rotation angle β was 0°. In group I_1, the height of camera h_1 was 305 mm. In group I_2, the height of camera h_2 was 285 mm. The corners pixels were extracted, and their actual imaging angles and depths were calculated based on the depth extraction model and ordinate pixels. The experimental data are shown in Table 4. The true depth was measured by a tape. The actual imaging angle of the corner can be calculated according to the cosine value of it, which equal to the actual depth divided by height. And the relative error was obtained by dividing the absolute error (the difference between the calculated depth and the true depth) by the true depth.

From Table 4, we can conclude that the relative error of the depth calculated by depth extraction model does not exceed 3%. The average relative error of depth is 0.93% when the distance is from 0.5 to 2.6 meters. The errors of the depth extracted by depth extraction model may related to many factors, such as the accuracy of the image processing algorithm, different light conditions or some other factors. In addition, due to the nonlinear distortion of camera lens, the closer the target object is to the optical axis of the camera, the smaller the image distortion error and the more accurate the measurement, and vice versa. However, from Table 4 we can conclude that in a certain rang, the measurement error is random, and is acceptable in our next tree DBH and height measurement works.

In experiment 2, the camera rotation angles β of experimental groups I_1, I_2, I_3, I_4, I_5 were -10°, 0°, 10°, 20° and 30°, respectively, the height of camera h_1 was 408 mm. We also calculated the relative error root mean square (rRMS) of depth D, vertical distance T_x and distance L under different camera rotation angles. Experimental data is shown in Table 5.

Table 5. Root mean square of the relative error of the measured values with different camera rotation angles

β	Root mean square of relative error		
	D	T_x	L
-10°	0.0319	0.0362	0.0323
0°	0.0179	0.0207	0.0176
10°	0.0199	0.0280	0.0205
20°	0.0280	0.0331	0.0291
30°	0.1241	0.1351	0.1253

The experimental results show that when the camera rotates counterclockwise, the relative error RMS of the depth D, the vertical distance T_x and the distance L were relatively larger. Otherwise, the relative errors RMS were smaller. It is because that once the smartphone camera is rotated clockwise, the range of imaged ground will be far away from the centre of the image and more closer to the bottom of the image, where the nonlinear distortion is larger. It is beneficial to improve the measurement accuracy when we collect image by rotated the smartphone clockwise. The measurement error was also affected by the height of camera, camera intrinsic parameter accuracy and so on.

5.2 Ranging in Nature Environment

To verify the accuracy of the passive ranging method in nature scene, we took five images by smartphone camera and each image contained three target objects. In experiment 3, the camera rotation angle β was 0°, the height of camera h was 1285 mm. Experimental data are shown in Table 6. The experimental results showed that the relative errors of this method were no more than 6%, while its average relative error was 1.71% within a range of 3–10 m. Sheng et al. [31] developed an underwater binocular vision ranging system with an average relative error of 2.34%, Zou and Yuan [32] achieved a relative error less than 10% of the passive ranging based on monocular vision. Therefore, compare with other passive ranging methods based on machine vision, this method had a relative higher measurement accuracy. In addition, our method is not as accurate as it reported by Huang et al. [25] (a relative error less than 3%). However, compare with this method, we should not to simulate linear relation for all kind of cameras, different camera rotation angles or heights of camera.

The accuracy of this passive ranging method may directly determined by the accuracy of depth extraction model and T_x measurement result.

Table 6. Measurement accuracy of target object distance

Group	True distance L (mm)	Pixel	Calculated distance L' (mm)	Absolute error (mm)	Relative error (%)
I_1	2609	(3921.23, 1338)	2576.28	32.72	1.25
	4977	(3116.51, 2215.3)	4949.51	27.49	0.55
	6000	(2947.6, 1008.67)	5956.71	43.29	0.72
I_2	3000	(3735.74, 904.7)	2961.45	38.55	1.28
	4010	(3339.2, 1472.43)	3946.42	63.58	1.59
	10320	(2584.7, 580.25)	10425.44	516.58	5.01
I_3	5002	(3112, 1097.92)	4933.8	68.20	1.36
	7720	(2761.7, 1502.4)	7590.05	129.95	1.68
	8000	(3762.32, 789.55)	7768.54	231.46	2.89
I_4	3617	(3477.88, 1049.93)	3556.47	60.53	1.67
	5214	(3056, 1473.02)	5191.46	22.54	0.43
	7000	(2849.52, 692.23)	6885.23	114.78	1.64
I_5	3215	(3614.7, 591.6)	3292.41	77.41	2.41
	4500	(3214.54, 1489.1)	4417.75	82.25	1.83
	5207	(3057.1, 896)	5278.95	71.95	1.38

6. Conclusion and Future Work

In this paper, we present a depth extraction model and passive ranging method based on monocular vision system using smartphone. First, we use an optimized corner extraction algorithm to detect and extract the sub-pixel corners of a checkerboard with a fixed width and increased length, and investigate the linear relationship of the actual imaging angle of the object point and the ordinate pixel of the corresponding image point with different camera rotation angles. It is verified that given the same abscissas, the ordinatesis of the image points linearly related to their actual imaging angles ($p<0.01$, $r\geq0.99$). Therefore, by assuming a linear function and substituting the actual imaging angles and the ordinate pixels of the special conjugate points (maximum and minimum values) into the linear relationship function, we establish a depth extraction model suitable for various of smartphones. What's more, an improved camera calibration model with a nonlinear distortion term is used to obtain the distortion parameters and intrinsic parameters of camera, and the intrinsic parameters are used to calculate the depth of the target object. According to the principle of camera stereo imaging system, we calculate the vertical distance from the target object to the camera optical axis, and range the distance by Pythagorean theorem. To verify the accuracy of the model, we conduct two sets of experiments in both close and long-distance ranging in the laboratory and nature environment. The experimental results show that the average relative error of the depth measurement is 0.937% when the distance is within 0.5–2.6 m. What's more, the relative error of measurement is 1.71% when the distance is 3–10 m. Therefore, using this method to measure distance has a high measurement accuracy.

Compared with other passive ranging methods, this method uses a smartphone to measure distance and extract depth which is convenient, portable and easy to be used in daily practice. It does not require a large scene calibration site and avoids errors caused by data fitting. In addition, we only need to obtain the intrinsic parameters by camera calibration at the first time, and then we can calculate the distance from the target object to the camera by a single image. It does not require any calibrations or known dimension objects to be placed in the measuring scene. However, when the target object to be photographed is far away from the camera, due to the perspective transformations, the detection accuracy of its contour is reduced and the measurement accuracy may also be affected. To solve this problem, in the next step we will devote to use the deep learning method to detect and extract a more specific object contour. Moreover, this technique can further be used as the basic of an object's height and width measurement. Therefore, in the future of our work, we will also engage to use this method to measure the 3D information of an object.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No. 31670641, The research of tree's height and DBH measurement method based on the intelligent mobile terminal) and Zhejiang Science and Technology Key R&D Program Funded Project (No.2018C02013).

References

[1] B. Hou, B. Khanal, A. Alansary, S. McDonagh, A. Davidson, M. Rutherford, J. V. Hajnal, D. Rueckert, B. Glocker, and B. Kainz, "3-D reconstruction in canonical co-ordinate space from arbitrarily oriented 2-D images," *IEEE Transactions on Medical Imaging*, vol. 37, no. 8, pp. 1737-1750, 2018.

[2] M. Waechter, M. Beljan, S. Fuhrmann, N. Moehrle, J. Kopf, and M. Goesele, "Virtual rephotography: novel view prediction error for 3D reconstruction," *ACM Transactions on Graphics*, vol. 36, no. 1, article no. 8, 2017.

[3] Y. I. Abdel-Aziz, H. M. Karara, and M. Hauck, "Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry," *Photogrammetric Engineering & Remote Sensing*, vol. 81, no. 2, pp. 103-107, 2015.

[4] D. L. McKay, M. R. Wohlers, C. K. Chuang, J. S. Draper, and J. Walker, "Airborne validation of an IR passive TBM ranging sensor," in *Proceedings of SPIE 3698: Infrared Technology and Applications XXV*. Bellingham, WA: International Society for Optics and Photonics, 1999, pp. 491-500.

[5] R. C. Bradshaw, D. P. Schmidt, J. R. Rogers, K. F. Kelton, and R. W. Hyers, "Machine vision for high-precision volume measurement applied to levitated containerless material processing," *Review of Scientific Instruments*, vol. 76, no. 12, article no. 125108, 2015.

[6] M. Aki, T. Rojanaarpa, K. Nakano, Y. Suda, N. Takasuka, T. Isogai, and T. Kawai, "Road surface recognition using laser radar for automatic platooning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 10, pp. 2800-2810, 2016.

[7] R. Bajsy, "Active perception vs. passive perception," in *Proceedings of the 3rd Workshop on Computer Vision: Representation and Control,* Bellaire, MI, 1985, pp. 55-59.

[8] H. Zhang, H. Wei, H. Yang, and Y. Li, "Active laser ranging with frequency transfer using frequency comb," *Applied Physics Letters*, vol. 108, no. 18, article no. 181101, 2016.

[9] J. Yang, C. Yang, J. Liu, N. Zhu, L. Yu, and Y. Liu, "Visual passive ranging system based on target feature size," *Optics and Precision Engineering*, vol. 26, no. 1, pp. 245-252, 2018.

[10] F. Lin, X. Dong, B. M. Chen, K. Y. Lum, and T. H. Lee, "A robust real-time embedded vision system on an unmanned rotorcraft for ground target following," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 2, pp. 1038-1049, 2011.

[11] J. Sun, G. Sun, P. Ma, T. Dong, and Y. Yang, "Laser target localization based on symmetrical wavelet denoising and asymmetric Gauss fitting," *Chinese Journal of Lasers*, vol. 44, no. 6, article no. 604001, 2017.

[12] R. Y. Takimoto, M. S. G. Tsuzuki, R. Vogelaar, T. de Castro Martins, A. K. Sato, Y. Iwao, T. Gotoh, and S. Kagei, "3D reconstruction and multiple point cloud registration using a low precision RGB-D sensor," *Mechatronics*, vol. 35, pp. 11-22, 2016.

[13] J. Shi, Y. Li, G. Qi, and A. Sheng, "Machine vision based passive tracking algorithm with intermittent observations," *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, vol. 45, no. 6, pp. 33-37, 2017.

[14] C. Xu, D. Huang, and F. Kong, "Small UAV passive target localization approach and accuracy analysis," *Chinese Journal of Scientific Instrument*, vol. 36, no. 5, pp. 1115-1122, 2015.

[15] J. Mei, D. Zhang, and Y. Ding, "Monocular vision for pose estimation in space based on cone projection," *Optical Engineering*, vol. 56, no. 10, article no. 103108, 2017.

[16] R. Szeliski, *Computer Vision: Algorithms and Applications*. New York, NY: Springer, 2010.

[17] A. Ming, T. Wu, J. Ma, F. Sun, and Y. Zhou, "Monocular depth-ordering reasoning with occlusion edge detection and couple layers inference," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 54-65, 2015.

[18] E. Alexander, Q. Guo, S. Koppal, S. J. Gortler, and T. Zickler, "Focal flow: Velocity and depth from differential defocus through motion," *International Journal of Computer Vision*, vol. 126, pp. 1062-1083, 2018.

[19] C. S. Royden, D. Parsons, and J. Travatello, "The effect of monocular depth cues on the detection of moving objects by moving observers," *Vision Research*, vol. 124, pp. 7-14, 2016.

[20] T. Liu, Y. Mo, G. Xu, X. Dai, X. Zhu, and J. Lu, "Depth estimation of monocular video using non-parametric fusion of multiple cues," *Journal of Southeast University (Natural Science Edition)*, vol. 45, no. 5, pp. 834-839, 2015.

[21] Y. Seo, A. Heyden, and R. Cipolla, "A linear iterative method for auto-calibration using the DAC equation," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Kauai, HI, 2001.

[22] G. Wu and Z. Tang, "Distance measurement in visual navigation of monocular autonomous robots," *Jiqiren (Robot)*, vol. 32, no. 6, pp. 828-832, 2010.

[23] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, 1997, pp. 1106-1112.

[24] C. Wu, C. Lin, and C. Lee, “Applying a functional neurofuzzy network to real-time lane detection and front-vehicle distance measurement,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 577-589, 2012.

[25] X. Huang, F. Gao, G. Xu, N. Ding, and L. Xing, “Depth information extraction of on-board monocular vision based on a single vertical target image,” *Journal of Beijing University of Aeronautics and Astronautics*, vol. 41, no. 4, pp. 649-655, 2015.

[26] D. Scaramuzza, A. Martinelli, and R. Siegwart, “A toolbox for easily calibrating omnidirectional cameras,” in *Proceedings of 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China, 2016, pp. 5695-5701.

[27] C. Harris and M. Stephens, “A combined corner and edge detector,” in *Proceedings of the British Machine Vision Conference*, Manchester, UK, 1988, pp. 147-151.

[28] J. Shi, “Good features to track,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, 1994, pp. 593-600.

[29] A. Geiger, F. Moosmann, O. Car, and B. Schuster, “Automatic camera and range sensor calibration using a single shot,” in *Proceedings of 2012 IEEE International Conference on Robotics and Automation*, Saint Paul, MN, 2012, pp. 3936-3943.

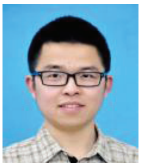
[30] Z. Zhang, “Flexible camera calibration by viewing a plane from unknown orientations,” in *Proceedings of the 7th IEEE International Conference on Computer Vision*, Kerkyra, Greece, 1999, pp. 666-673.

[31] M. Sheng, H. Zhou, H. Huang, and H. Qin, “Study on an underwater binocular vision ranging method,” *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, vol. 46, no. 8, pp. 93-98, 2018.

[32] B. Zou and Y. Yuan, “High precision distance measurement based on monocular vision for intelligent traffic,” *Journal of Transportation on Systems Engineering and Information Technology*, vol. 18, no. 4, pp. 46-53,60, 2018.

Xin-mei Wu **https://orcid.org/0000-0001-7983-454X**

She received B.S. degree in GIS from Anhui Science and Technology University in 2015. She is currently working as master’s degree candidate of Zhejiang Agriculture and Forestry University in Zhejiang, China. Her main research covers machine vision and close-range photogrammetry.



Fang-li Guan **https://orcid.org/0000-0001-7409-2129**

He was born in 1992 in Zhejiang Province, China. He received master’s degree from Zhejiang Agriculture and Forestry University in 2018. He is currently working as Ph.D. degree Candidate of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University in Wuhan, China. His main research covers computer vision and smart navigation for pedestrians.

Ai-jun Xu **https://orcid.org/0000-0001-6789-6938**

He was born in 1976 in Anhui Province, China. He received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University in 2007. He is currently working as a Professor in School of Information Engineering-Zhejiang Agriculture and Forestry University, Hangzhou, China. His current research interest includes computer application technology and the application of GIS in the direction of agricultural informatization, etc.