Research Article

Monocular Vision Ranging and Camera Focal Length Calibration

Lixia Xue ⓘ**,**[1,2] **Meian Li** ⓘ**,**[1] **Liang Fan** ⓘ**,**[3] **Aixia Sun** ⓘ**,**[1] **and Tian Gao** ⓘ[1]

[1]*Inner Mongolia Agricultural University, Computer and Information Engineering College, Hohhot 010000, China*
[2]*Inner Mongolia Autonomous Region Key Laboratory of Big Data Research and Application of Agriculture and Animal Husbandry, Hohhot 010018, China*
[3]*School of Computer Science and Informatics, Cardiff University, Cardiff CF243AA, UK*

Correspondence should be addressed to Meian Li; limeian1973@126.com

The camera calibration in monocular vision represents the relationship between the pixels' units which is obtained from a camera and the object in the real world. As an essential procedure, camera calibration calculates the three-dimensional geometric information from the captured two-dimensional images. Therefore, a modified camera calibration method based on polynomial regression is proposed to simplify. In this method, a parameter vector is obtained by pixel coordinates of obstacles and corresponding distance values using polynomial regression. The set of parameter's vectors can measure the distance between the camera and the ground object in the field of vision under the camera's posture and position. The experimental results show that the lowest accuracy of this focal length calibration method for measurement is 97.09%, and the average accuracy was 99.02%.

1. Introduction

Measuring the distance between self and the obstacle is a crucial part of many fields. The type of method to measuring distance by cameras is called the vision-based ranging method, which may promote the development of automatic measurement and has a great research value [1, 2]. The vision-based ranging method includes the monocular vision-based ranging method and stereo vision-based ranging method. Stereo vision-based ranging methods use the parallax of cameras to measure, which needs to match multiple images taken by multiple cameras. Monocular vision-based ranging method has a higher performance than the stereo vision-based ranging method because it does not need match images in the data preprocessing stage [3].

Monocular vision-based ranging methods can be divided into three categories, including proportion-based methods, machine learning-based methods, and coordinate transformation-based methods. Proportion-based methods are according to the principle that the distance is inversely proportional to the image's size from the target in the image plane [4, 5]. Taking the ranging model proposed by Bao and

Wang [5], this model first assumes the width of all vehicles as a fixed value and then uses this value to learn the model parameters from the images of vehicles at different distances. The ranging accuracy of this model is unsteady because if the measured vehicle is not directly in front of the camera, the value of width will no longer be correct. Machine learning-based methods try to let the computer learn the parameters of the ranging model. For example, Meng et al. [6] proposed the ranging model based on the R-CNN [7] and nonlinear regression. This model first utilizes the R-CNN to detect the position of the preceding vehicle, and then, the distance is obtained by a nonlinear regression model. The accuracy of the objective decision is more than 90%. Coordinate transformation-based methods have strong interpretability. This type of method utilizes the relationship between world coordinates, camera coordinates, pixel coordinates, and camera parameter matrix to calculate the distance between the camera and obstacles [8, 9]. Taking the method proposed by Lu et al. [9], this model increased the ranging accuracy to 95.43%, which derives the ranging model based on the principle of visual imaging. Although the ranging accuracy of this model is higher than previous methods, camera

calibration as a key technology in this model needs to be improved.

The high-precision camera calibration aims to determine a set of geometric parameters. It is a prerequisite to extract 3D information from the captured images using the object's projections in the image plane. Camera calibration methods can be divided into two categories, including traditional calibration and self-calibration. The traditional calibration utilizes multiple reference points on the object to establish the relationship between the 2D pixel coordinates and the 3D world coordinates. Yakimovsky and Cunningham [10] in 1978 proposed a camera calibration method to calculate the transformation matrix of stereo cameras. They used a highly linear lens and ignored the distortion to improve the accuracy of 5 mm at a distance of 2 m. Due to the narrow field of view and not considering lens distortion, this method may cause more errors in a wide field. The unknown parameters computed by linear equations in the traditional camera calibration method may not be linearly independent, which increases the error of calibration. Martins et al. [11] in 1981 proposed the biplane calibration method using the points on the double calibration plane. This method avoids any restrictions on the extrinsic camera parameters. On the premise that there is no deflection in the coordinate system, the average error is about 4 mils with a distance of 25 inches. However, the nonlinear lens distortion is not corrected. Tsai [12] proposed a novel camera calibration method which considered the camera distortion. First, the camera parameters are solved using the direct linear transformation method or perspective projection transformation matrix. Then, the obtained parameters are taken as the initial values. The nonlinear optimization method is used to improve the precision of the calibration. This method requires high-precision calibration targets and a similar size field of vision. Chen et al. [13] proposed a calibration method based on a bundle adjustment system, which reduces the requirement of calibration targets. The world coordinate system is relatively stationary with the steam hammer to decrease the error of calibration. Experiments show that the calibration system measures the ram speed of a steam hammer accurately.

Traditional camera calibration has a strong dependence on calibration targets, which has high requirements for the precision of calibration targets. This method is difficult to transfer the unknown scene because it is usually limited to a specific field of vision and distance. Therefore, the self-calibration was proposed because of its higher operability, which obtains the image sequence by controlling the camera motion and then calculates the parameters by matching the image sequence. Luong and Faugeras [14] proposed a self-calibration method based on point correspondences and fundamental matrices. They used point correspondences between three images to estimate the perspective projection matrices and parameters of the camera. In contrast to traditional camera calibration methods, this method washes calibration targets with a known 3D shape. This method has the disadvantages of the high cost of polynomial calculation and sensitivity to noise since the continuation work in the complex plane. On this basis, Zhang [15] proposed a calibration method between the traditional calibration method and the self-calibration method. This method utilizes the two-dimensional measurement information for the camera's calibration, which reduces the requirements for the equipment compared with the traditional calibration. In this method, many different angle images of a two-dimensional plane calibration target (checkerboard) are taken to detect the feature points and calculate the parameter matrix of the camera. Nonlinear refinement is based on the maximum likelihood criterion to optimize the calibration results. Although this method has simplified the calibration process compared with previous methods, it still has some requirements for the standardization of operation. For example, the image taken should avoid noise. Otherwise, the inaccurate corner extraction will increase the calibration error. Dong and Isler [16] proposed a method for external parameters of camera calibration that obtained the point-to-to plane constraints by two noncoplanar triangles. This method reduces the dependence on the camera's initial state estimation. At the same time, it reduces the number of observations without reducing the accuracy of the calibration. Xu [17] proposed a camera calibration method based on mirrored. This method reflects two groups of orthogonal phase-shifting sinusoidal figures by a mirror to calculate the relationship between cameras by constraints between phases.

To sum up, monocular vision-based ranging methods based on coordinate transformation have higher accuracy than other methods. This type of method [18, 19] has not only a model efficiency but also interpretability and stability. They are generally based on the linear imaging model constructed by the pinhole imaging principle to simplify the derivation of the ranging model because the pinhole imaging principle can directly construct the geometric correspondence between the world coordinates and the pixel coordinates. At the same time, considering that most of the vision-based ranging methods are used in the automatic measurement of robots or intelligent vehicles, where the camera is fixed in a certain attitude, a lot of measurement targets are ground objects. Therefore, this study proposes a ranging model different from the previous vision-based ranging methods, which measures the distance between the camera and the ground object when the camera has an inclination angle in three dimensions. Because the world coordinate system is two-dimensional, the computational complexity of the model is lower and there are not many requirements for the three-dimensional structure of the object.

As the only internal parameter of the camera needs to be calibrated in the model, the accuracy of focal length calibration directly affects the range's accuracy. In the linear imaging model, the object is presented to the image plane through a small hole depending on the principle of straight-line propagation. But in fact, the camera is convex lens imaging, which has distortion and defocusing phenomenon. Distortion will lead to the inconsistency between the theoretical imaging position and the actual imaging position, and defocusing will lead to the image in some positions not presented on the image plane. Previous references [20, 21]

proposed the calibration of a nonlinear camera, which calibrated the focal length and distortion parameters and then restored the theoretical imaging position through the distortion parameters and the actual imaging position. Although radial distortion, centrifugal distortion, and thin prism distortion are considered in these methods, defocusing or other possible factors are not well considered. Therefore, the accuracy of the ranging model can be improved.

In this study, all the imaging points are considered to be obtained based on pinhole imaging, so there is no need to restore the pixel coordinates. The effects of distortion and defocusing are reflected in the value of focal length. That is to say, each point at each position has its corresponding focal length. This focal length includes various factors that may affect the imaging position. This study uses the actual distance and pixel coordinates to calculate the focal length corresponding to different pixel positions and then analyses the distribution of focal length value relative to pixel coordinates. It is found that their distribution is not a simple linear distribution. Therefore, to improve the accuracy of range, this study attempts to learn the distribution of focal length corresponding to different pixel positions by nonlinear regression (polynomial regression). The results show that the method can effectively improve the accuracy of the ranging model.

In a word, this study focuses on proposing a simple and high-accuracy camera's focal length calibration method to improve the accuracy of the monocular vision-based ranging model. The major method is to use a simple linear imaging model to deduce the complexity of the ranging model and then combine the distortion and defocusing phenomenon caused by the nonlinear imaging of the camera into the focal length calibration process. The main innovations of this study are as follows:

(1) When the camera has an inclination angle in three dimensions, the ranging model for ground object based on the linear imaging model and geometric coordinate transformation is proposed

(2) The distortion caused by convex lens imaging and the influence caused by defocusing is reflected in the focal length of the linear imaging model

(3) The calibration process does not require the calibration target, and the camera does not need to move.

(4) The focal length values containing the effect by distortion and defocusing are calculated by the ranging model, and the nonlinear distribution is learned by polynomial regression.

2. Proposed Methods

2.1. Monocular Vision-Based Model. In the real scene, most of the obstacles are on the ground, such as pedestrians or other vehicles. The model in this study focuses on the ground feature points, that is, the 3D world coordinate system is reduced to the 2D coordinate system. The posture of the installed camera depends on the installed location. The camera does not have an inclination angle when the camera is facing the front and the optical axis is parallel to the ground. The optical axis is defined as a line perpendicular to the image plane and passing through the optical center. Based on the pinhole camera model and the principle of straight-line propagation of light [22], the ranging model for the ground objects without the inclination angle of the camera is shown in Figure 1.

As shown in Figure 1, O' is the projection of the camera optical center O on the ground plane. The optical axis (OO'') passes through the camera optical center O and is parallel to the ground plane; the optical axis (OO'') is perpendicular to the image plane. The world coordinate system is established with O' as the origin, the line passing through O' and parallel to the optical axis as Y-axis, and the line passing through O' and perpendicular to Y-axis as X-axis. P is the measured point on the ground plane, P_Y is the projection of P on the Y-axis, P_X is the projection of P on the X-axis, O'' is the projection of the optical axis on the image plane, the height from the optical center O to the ground is H (the length of OO'), the image on the image plane of the point P is the point P', and the physical coordinates of P' on the image plane are $P'(x, y)$. The length of OO'' is recorded as the focal length f. The length of $O'P$ (d) is the distance which we will calculate. The camera will have an inclination angle with three different dimensions in the real scene. The details of the three dimensions are shown in Figure 2.

As shown in Figure 2(a), the field of vision of the camera is changed after the camera rotates. The ranging model is consistent with the noninclination angle. As shown in Figure 2(b), the position of the optical center did not change, and the optical axis is still parallel to the ground. The difference with the noninclination angle is that the image plane rotates around its center in the two-dimensional plane. The image plane can be restored to the state without inclination by reverse rotation in the two-dimensional plane. The inclination angle obtained by this way of rotation is called the left or right inclination angle. Since it is not involved in reconstructing the ranging model, the detail of restoring the left-right inclination angle to a noninclination angle is described in the discussion. As shown in Figure 2(c), the optical axis of the camera is not parallel to the ground after rotation. The ranging model needs to reconstruct. The inclination angle obtained by this way of rotation is called the up or down inclination angle. Taking down the inclination angle for an ample, the illustrated of the ranging model is shown in Figure 3.

The meanings of P, P', O, O', O'', P_X, P_Y, P_X', and P_Y' in Figure 3 are the same as is shown in Figure 1. The difference between Figures 1 and 3 is that OO'' is the optical axis with an inclination angle α, that is, the angle between the horizontal line and the optical axis is called α. When the camera has a down inclination angle, regulation $\alpha > 0$, and when the camera has an up inclination angle, regulation $\alpha < 0$. M is the intersection point of $P'P_X'$ and the horizontal plane (plane OMN), N is the intersection point of $P_Y'O''$ and the horizontal plane (plane OMN), and the plane ONM is parallel to the ground plane; the optical axis (OO'') is
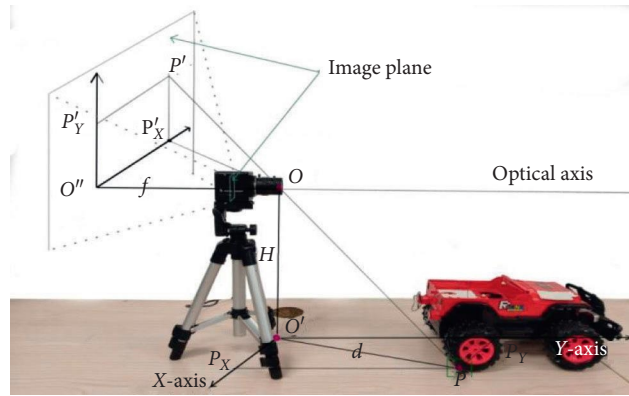
Figure 1: Ranging model without an inclination angle.
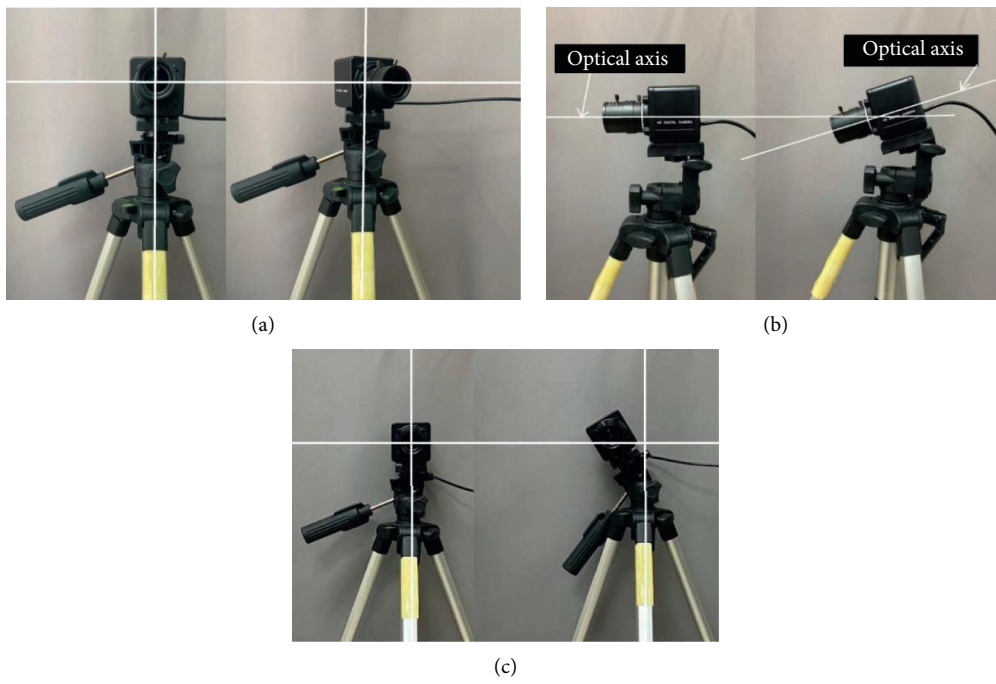

(a)

(b)

(c)

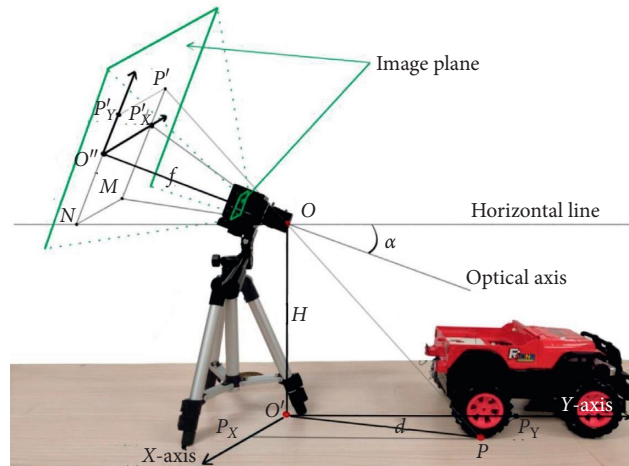Figure 2: Camera inclination angle decomposition.

Figure 3: Ranging model with an inclination angle.

perpendicular to the image plane. The length of OO'' is defined as the focal length f, and the length of $O'P$ (d) is the distance which we will calculate.

MO is the projection of $P'O$ on the plane OMN, $O'P$ is the projection of OP on the ground plane, and the three points of P, O, and P are colinear. Because the plane OMN is parallel to the ground plane, according to the property of parallel planes, the angle between a line and its projection on two parallel planes is equal. Therefore,

$$\angle OPO' = \angle P'OM = \angle P'OP'_X + \angle P'_X OM. \tag{1}$$

The optical axis OO'' is perpendicular to the image plane and $O''N$ is in the image plane. According to the vertical property of line and planes, a straight-line perpendicular to a plane is perpendicular to any straight line in the plane. In $\Delta NO''O$, $NO''\perp OO''$, and $\angle O''ON = \alpha$, $OO'' = f$, $O''N$ is

$$O''N = f \cdot \tan \alpha. \tag{2}$$

Because the optical axis OO'' is perpendicular to the image plane and $O''P'_X$ is in the image plane, $OO''\perp O''P'_X$. Following Pythagoras' theorem, OP'_X is

$$OP'_X = \sqrt{x^2 + f^2}. \tag{3}$$

OO'' is perpendicular to the image plane and in the plane $OO''P'_X$. According to the judgment theorem of plane perpendicularity, OO'' is perpendicular to the image plane. The plane $OO''P'_X$ passing through OO'' is also perpendicular to the image plane. OP'_X is in the plane $OO''P'_X$ and

$P'_X M$ is in the image plane. In $\Delta P'_X OM$, $OP'_X \perp P'_X M$, $O''N \| P'_X M$, and $O''P'_X = NM$. According to the properties of a parallelogram, the opposite sides are parallel and equal, $P'_X M = O''N$, and $\angle P'_X OM$ is

$$\angle P'_X OM = \tan^{-1}\frac{P'_X M}{OP'_X} = \tan^{-1}\frac{f \cdot \tan \alpha}{\sqrt{x^2 + f^2}}. \tag{4}$$

The plane $OO''P'_X$ is perpendicular to the image plane, and OP'_X is in the plane $OO''P_X$ and $P'P'_X$ is in the image plane. In $\Delta P' OP'_X$, $OP'_X \perp P'P'_X$, $\angle P'OP'_X$ is

$$\angle P'OP'_X = \tan^{-1}\frac{P'P'_X}{OP'_X} = \tan^{-1}\frac{y}{\sqrt{x^2 + f^2}}. \tag{5}$$

OO' is perpendicular to the ground plane, in $\Delta OPO'$, $PO'\perp OO'$, $OO' = H$, $O'P = d$, and $\angle OPO'$ is

$$\angle OPO' = \tan^{-1}\frac{OO'}{O'P} = \tan^{-1}\frac{H}{d}. \tag{6}$$

The simultaneous equations (8) and (11)–(13) can be obtained:

$$\tan^{-1}\frac{f \bullet \tan \alpha}{\sqrt{x^2 + f^2}} + \tan^{-1}\frac{y}{\sqrt{x^2 + f^2}} = \tan^{-1}\frac{H}{d}. \tag{7}$$

According to the arctangent addition theorem, we get the following results:

$$\tan^{-1}\frac{f \bullet \tan \alpha}{\sqrt{x^2 + f^2}} + \tan^{-1}\frac{y}{\sqrt{x^2 + f^2}} = \tan^{-1}\frac{\left(f \bullet \tan \alpha/\sqrt{x^2 + f^2}\right) + \left(y/\sqrt{x^2 + f^2}\right)}{1 - \left(f \bullet \tan \alpha/\sqrt{x^2 + f^2}\right) \cdot \left(y/\sqrt{x^2 + f^2}\right)} = \tan^{-1}\frac{H}{d}. \tag{8}$$

Namely,

$$\frac{\sqrt{x^2 + f^2} \cdot (f \cdot \tan \alpha + y)}{x^2 + f^2 - y \cdot f \cdot \tan \alpha} = \frac{H}{d}. \tag{9}$$

From equation (9), it concludes that

$$d = \frac{H \cdot \left(x^2 + f^2 - y \cdot f \cdot \tan \alpha\right)}{\sqrt{x^2 + f^2} \cdot (f \cdot \tan \alpha + y)}. \tag{10}$$

From equation (10), the coordinate (x, y) of the measured point P' on the image plane is the physical coordinate. Its unit should be a millimeter. However, the coordinate $(x0, y0)$ of the point P' directly extracted from the image plane is the pixel coordinate. This pixel coordinate system is established with the upper left corner of the picture as the original center, and the unit is pixels. It is necessary to translate the pixel coordinate $(x0, y0)$ into the physical

coordinate (x, y) in equation (10). The transformation formulas are shown in equation (11).

$$x = \left(x0 - \frac{CR}{2}\right) \cdot d_x,$$
$$y = \left(y0 - \frac{RR}{2}\right) \cdot d_y. \tag{11}$$

In equation (11), x and y denote the physical coordinates of point P', $x0$ and $y0$ denote the pixel coordinates of point P', and d_X and d_y denote the size of a pixel unit in x and y directions. RR and CR denote the row resolution and the column resolution of the image plane.

2.2. The Proposed Camera Calibration Model. According to equation (10), the constraint equation of focal length is shown in equation (12), where the distance d in equation (10) is replaced by the accurate distance D in equation (12) measured by the rangefinder.

$$
\begin{aligned}
&\left(D^2 (\tan \alpha)^2 - H^2\right) \cdot f^4 + 2^* y \cdot \tan \alpha \cdot \left(D^2 + H^2\right) \cdot f^3 \\
&+ \left(\left[D^2 \cdot y^2 + D^2 (\tan \alpha)^2 x^2 - H^2 y^2 (\tan \alpha)^2 - 2^* H^2 x^2\right] \cdot f^2\right) \\
&+ \left(2^* y \cdot x^2 \cdot \tan \alpha \cdot \left(D^2 + H^2\right) \cdot f + x^2 \left(D^2 \cdot y^2 - H^2 x^2\right)\right) = 0.
\end{aligned} \tag{12}
$$

When the camera's height from the ground and the camera's inclination angle remain unchanged, the pixel coordinates of the targets at different positions are different. According to the experimental data, the focal length which is calculated by equation (12) is different. The reason for this phenomenon can be explained by the defocus phenomenon [23] of convex lens imaging. For a camera, the image plane is fixed according to the clearest image in the center. Therefore, the image outside the center of the image plane does not exactly fit the image's position. The points with different object distances correspond to different focal lengths. To obtain the focal length corresponding to different physical coordinates (x, y), this study utilizes polynomial regression to learn the relationship between focal length and the physical coordinates. The physical coordinates are the independent variables, and focal lengths are the dependent variables. The polynomial in the form of equation (13) and the coefficient vector V in equation (13) are obtained by polynomial regression.

$$
\begin{aligned}
f(x, y) &= v00 + v10^* x + v01^* y \\
&\quad + v11^* x^* y + v02^* y^2 \ldots \ldots.
\end{aligned} \tag{13}
$$

2.3. Error Compensation Algorithm. To improve the accuracy of the ranging, we propose an algorithm to compensate for the ranging error which is produced from the coordinate extraction of the feature point and the measurement of the camera inclination angle by adjusting the pixel coordinates $(x0, y0)$. According to the experimental data, the generalization ability of the model is the strongest and the most stable when the ranging error threshold is set at 0.5%. The procedure is outlined in Algorithm 1. $E(d)$ is the difference between the calculated distance d and the accurate distance D.

3. Experiments

The equipment in our experiment is shown in Figure 4. The camera in this experiment is 1920 * 1080 pixels. 100 targets are placed in the range from 2 to 20 meters. The spatial location of the camera remains the same. The accurate distances D between the optical center projected on the ground and the targets are measured by a laser range finder.

As is shown in Figure 5, the pixel coordinates of the contact point between each target and the ground are extracted using MATLAB. From the 100 pieces of data, 14 pieces were randomly selected as the training dataset and 6 pieces as the test dataset. The focal length of the camera in the training dataset is calculated by equation (11). Utilizing the polynomial regression function in the fitting curve toolbox of MATLAB obtained the regression vector V in equation (13). The focal length of the test

dataset is calculated by equations (11) and (12). Equation (10) is used to calculate the distance d of the test dataset.

4. Results

The values of focal length calculation results of a training set are given in Table 1. For H (the height of the camera), α is the inclination angle of the camera. D is the distance between a target and the camera, which is measured by the laser rangefinder. (x, y) are the physical coordinates calculated according to equation (11), where $CR = 1920$, $RR = 1080$, and $d_x = d_y = 0.0026$ millimeters. Then, the focal length f is calculated according to equation (12). All variables in Table 1 are in millimeters (mm), except the variable a.

Taking (x, y) and f in Table 1 to polynomial regress, the number of parameters in the regression vector is called the length of the vector. Before regression, the convergence value of the length of the regression vector V needs to be determined. The length of the vector V depends on the regression accuracy and the size of the dataset. When the regression accuracy is constant and the size of the dataset is large enough, the value of vector V and the length of the coefficient vector V should be convergent. When the regression accuracy is fixed above 99.5%, Figure 6 shows the length of the vector V for different sizes of the dataset. As is shown in Figure 6, when the size of the dataset is less than 70, the length of the vector V gradually increases with the increase in the size of the dataset. When the size of the dataset reaches 70, the length of the vector V is stable at 12. Therefore, the length of the vector V is converged to 12. The calibration equation of focal length is shown in equation (14), and there are 12 regression parameters in equation (14). After the length convergence value of the regression vector V is determined, Figure 7 shows the parameters in the regression vector obtained by polynomial regression of the data in Table 1. Figure 8 is the regression model.

$$
\begin{aligned}
f(x, y) &= v00 + v10 * x + v10 * y + v20 * x^2 \\
&\quad + v11 * x * y + v02 * y^2 + v21 * x^2 \\
&\quad * y + v12 * x * y^2 + v03 * y^3 + v22 * x^2 \\
&\quad * y^2 + v13 * x * y^3 + v04 * y^4.
\end{aligned} \tag{14}
$$

f and $E(d)$ are recalculated based on equation (14) and the values in Table 1, and the results are given in Table 2. f' is the regression focal length calculated from Figure 7. Before $E(d)$ and after $E(d)$, respectively, represent the ranging error before and after using Algorithm 1. To verify the effectiveness of the proposed method, a number of experiments under different camera heights and inclination angles were made. The result of the maximum and average ranging error is given in Table 3. The maximum error of ranging is 2.91%, and the maximum average error of ranging is 0.98%.

Set $t = 0$, $a1 = a2 = 1$, $b1 = b2 = 0$, $E(d) = (abs(d - D)/D)$,
Repeat
$t =$ the number of $E(d)$ greater than the training threshold in the training set,
Repeat
    $y0' = y0 + 1$ or $y0' = y0 - 1$, $x0' = x0 + 1$ or $x0' = x0 - 1$.
Until $E(d)$ of all data in the training set is less than the threshold,
Utilizing the first-order linear regression method to obtain the regression expression like:
    $y0' = a1 * y0 + b1$, $x0 = a2 * x0 + b2$.
Until $t = 0$

ALGORITHM 1: Error compensation.
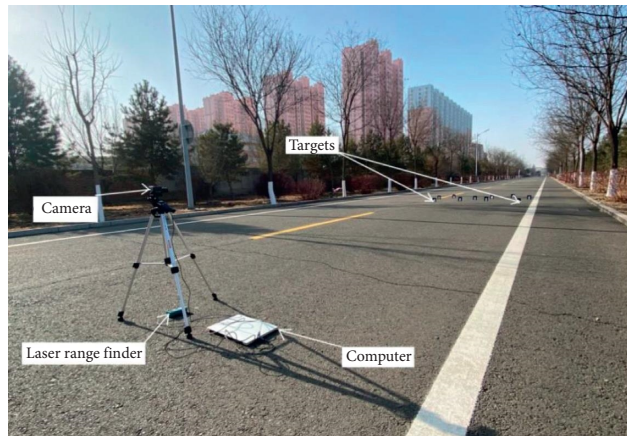
FIGURE 4: Experimental equipment and target placement.
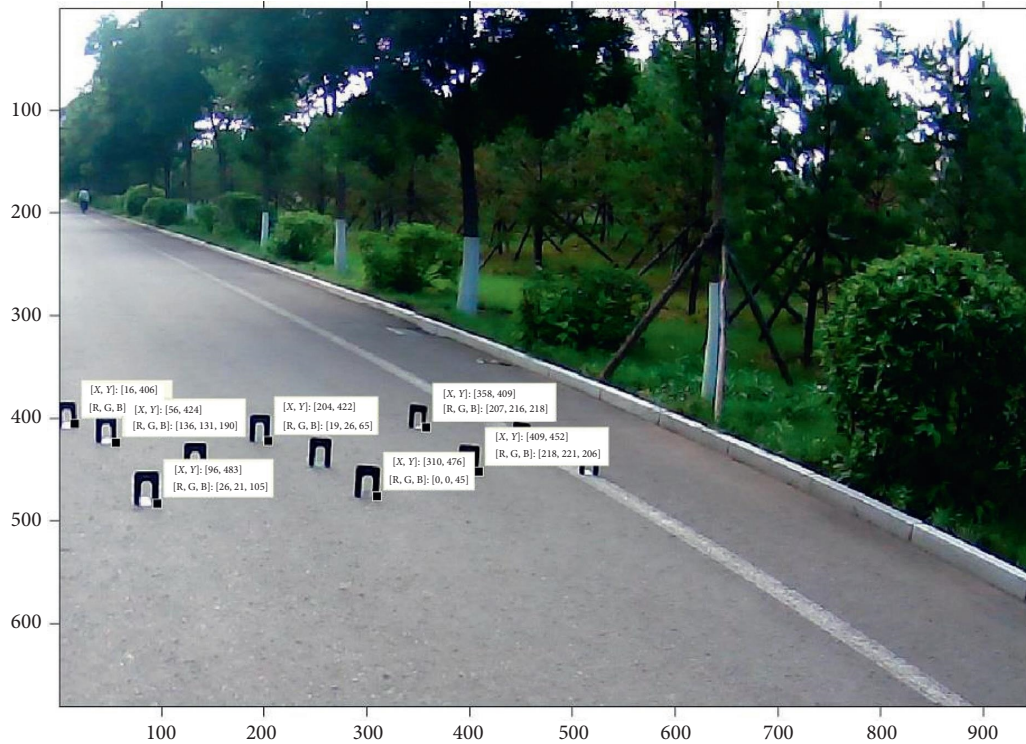
FIGURE 5: Pixel coordinates of target extraction.

TABLE 1: Focal length calculation results of a training set.

H/mm	D/mm	x/mm	y/mm	f/mm	E (d)/%
1451	10009	0.0832	−0.4316	4.608727	0.003032
1451	10273	0.156	−0.442	4.543764	0.005428
1451	11290	0.5382	−0.4654	4.268784	0.014776
1451	11825	0.7436	−0.4706	4.131007	0.013635
1451	13507	0.026	−0.5512	4.20472	0.041678
1451	12744	0.273	−0.5252	4.221513	0.065757
1451	13571	0.5018	−0.5382	4.112991	0.100842
1451	16760	0.104	−0.6032	3.964895	0.623398
1451	16939	0.0442	−0.611	3.991179	0.709382
1451	17367	0.2574	−0.611	3.940146	0.117892
1451	15414	0.3952	−0.5798	4.024374	0.233182
1451	18511	0.0026	−0.637	3.968945	0.242895
1451	19034	0.1716	−0.6422	3.948959	0.209105
1451	19839	0.3536	−0.6292	3.934795	0.098061
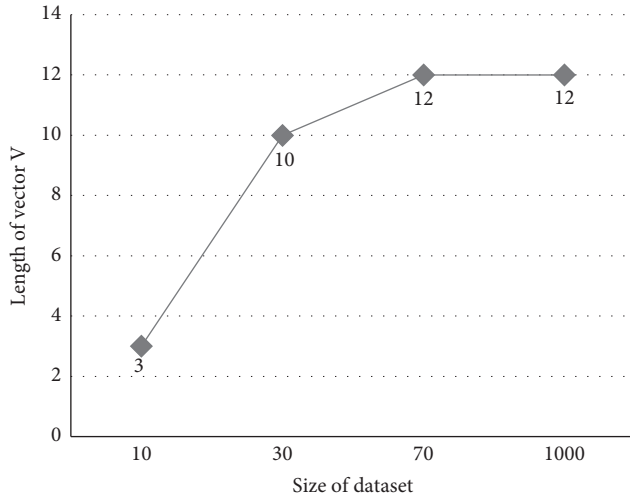
FIGURE 6: Comparison of the length of vector V under different sizes of the dataset.

$f(x, y) = p00 + p10^*x + p01^*y + p20^*x^2 + p11^*x^*y + p02^*y^2 + p21^*x^{2*}y + p12^*x^*y^2 + p03^*y^3 + p22^*x^{2*}y^2 + p13^*x^*y^3 + p04^*y^4$

$p00 = 4.126 \ (4.049, 4.203)$

$p10 = −0.05472 \ (−0.09939, −0.01004)$

$p01 = 0.2758 \ (0.1467, 0.4049)$

$p20 = 0.01416 \ (−0.06318, 0.0915)$

$p11 = −0.07034 \ (−0.2391, 0.0984)$

$p02 = −0.01169 \ (−0.3247, 0.3013)$

$p21 = −0.02118 \ (−0.0954, 0.05303)$

$p12 = 0.007137 \ (−0.111, 0.1253)$

$p03 = −0.05335 \ (−0.131, 0.02431)$

$p22 = −0.0002216 \ (−0.07504, 0.0746)$

$p13 = 0.01673 \ (−0.09963, 0.1331)$

$p04 = 0.03578 \ (−0.1355, 0.207)$

FIGURE 7: The values of the parameters.

5. Discussion

5.1. Ranging Model for the Camera with a Left or Right Inclination Angle.
Compared with no inclination angle, when the camera has a left or right inclination angle as shown in Figure 2(b), it is equivalent to that the pixel coordinate system rotates the same angle in the same direction. It is also equivalent to that in the same pixel coordinate system, and the imaging points rotate in the opposite direction with the same inclination. The diagram of coordinate transformation is shown in Figure 9.

$P1$ is the image of the fourth quadrant when the camera has a right inclination angle, and P is the image point without an inclination angle restored to the left, where $|OP1| = |OP|$. In $\Delta P_1 OP1_y$, according to the arctangent function,

$$\angle OP_1 P1_y = \tan^{-1}\frac{x1}{y1}. \tag{15}$$

Clearly,

$$\angle OPP_x = \angle POP_y = \angle OP_1 \ P1_y - \beta = \tan^{-1}\frac{x1}{y1} - \beta. \tag{16}$$

In ΔPOW,

$$
\begin{aligned}
x0 &= OP_y = OP \cdot \sin \angle OPP_x = \sqrt{x1^2 + y1^2}. \\
&\sin\left(\tan^{-1}\frac{X}{Y} - \beta\right) \cdot y0 = OP_x \\
&= OP \cdot \cos \angle OPP_x = \sqrt{x1^2 + y1^2}. \\
&\cos\left(\tan^{-1}\frac{X}{Y} - \beta\right),
\end{aligned}
\tag{17}
$$

$P(x0, \ y0)$ in equation (17) is the pixel coordinates of the measured point without a left inclination angle and right inclination angle in equation (11). When the pixel of the target is in other quadrants, the formula of coordinate transformation is different from the fourth quadrant, but the principle is the same.

5.2. Robustness of the Vector V.
To study the robustness of vector V, the height and selection of the camera were analyzed in this study. Table 4 provides the value of vector V when the height of the camera is 229 millimeters and 239 millimeters.

TABLE 2: Focal length calibration results.

f'/mm	d/mm	Before E (d)/%	After E (d)/%
4.60894033416042	10008.6964887034	0.003032	0.22004
4.54340989429504	10273.5575694457	0.005428	0.218179
4.26951768094114	11288.3317364075	0.014776	0.169679
4.13041136577056	11826.6123375319	0.013635	0.181646
4.20616526188184	13501.3705794102	0.041678	0.033825
4.21897294335602	12752.3800233027	0.065757	0.136777
4.116390711837	13557.3146909214	0.100842	0.064484
3.97931880894754	16655.5185071925	0.623398	0.239009
3.97525839762631	17059.1622131422	0.709382	0.497604
3.94269561616738	17346.5256637998	0.117892	0.30663
4.01820166494699	15449.9426832249	0.233182	0.148455
3.9737624793774	18466.0377275484	0.242895	0.427274
3.9450206644398	19073.8011347716	0.209105	0.045672
3.93673296545302	18425.9136618791	0.098061	0.29088

TABLE 3: Results of maximum generalization error and average generalization error.

Range	Inclination angle	Camera height (mm)	Number of groups	Max (generalization E (d))	Average (generalization E (d)) (%)
			1	2.80	0.83
2–20 m	13.6°	1451	2	2.49	0.94
			3	2.16	0.98
			1	2.89	0.83
2–20 m	0°	864	2	2.75	0.68
			3	2.91	0.97
			1	2.90	0.98
5–20 m	−10.5°	1491	2	2.01	0.78
			3	2.46	0.81

Table 5 provides the value of the vector V under camera A and camera B at the same height. As given in Tables 4 and 5, the height and selection of the camera will affect the value of the regression vector V. Therefore, in the application, the same regression vector is only suitable for the same height and the same camera.

5.3. Compare with Other Methods. To verify that the calibration method proposed in this study is suitable for the monocular vision ranging model and can effectively compensate for the distortion and defocus phenomenon, this study is compared with the Zhang [15] and Li [20] method calibrate the same camera, and the calibration results are given in Table 6.

Table 6 provides the intrinsic matrix calibrated by Zhang and Li methods. This matrix contains the focal length information of the camera. The radial distortion and the tangential distortion in Table 6 represent the distortion information of the camera, and the pixel coordinates can be corrected by these distortion parameters. The focal length (regression model) in Table 6 is the regression model of focal length obtained in this study, where x and y are the pixel coordinates.

In order to verify the performance of the three calibration methods in the monocular vision ranging model, a large number of ground measured points are randomly distributed on the ground. The distribution of the measured points on the image plane is shown in Figure 10. These measured points are randomly distributed in all areas of a picture and randomly selected four groups of test sets, 30 feature points in each group. Zhang and Li's methods use their distortion parameters to correct the pixel coordinates before ranging. The focal length obtained by the three calibration methods and the pixel coordinates after correction is substituted into the ranging model to calculate the distance. The ranging error results are shown in Figure 11.

As shown in Figure 11, the proposed calibration method is more stable and accurate than Zhang and Li's method for the monocular vision-based distance measurement model. Table 7 provides the accuracy comparison of the ranging model in this study compared with other ranging models.

Combined with Figure 11 and Table 7, it can be seen that the proposed calibration method in this study is more suitable for the ranging model in this study than other calibration methods. The monocular vision-based ranging model in this study has higher accuracy than other ranging methods.
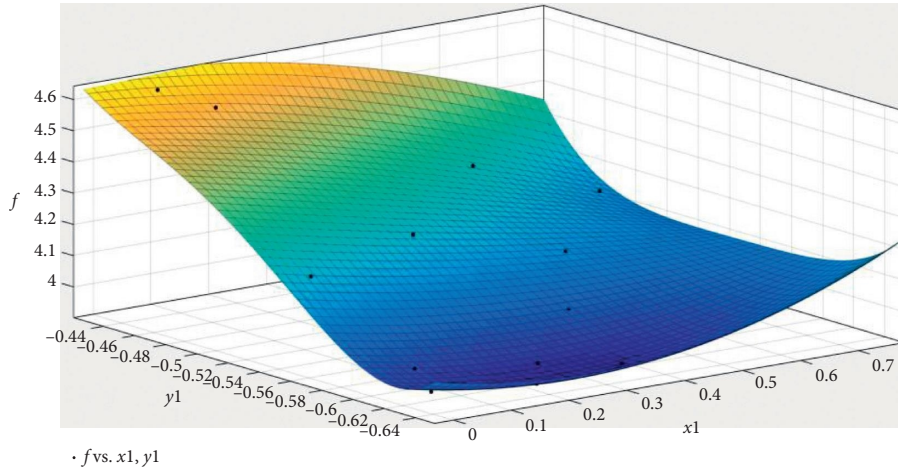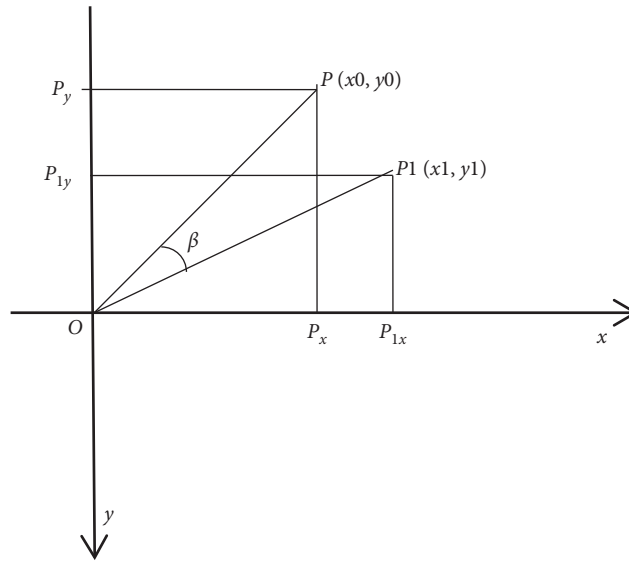
FIGURE 8: Regression model.

FIGURE 9: Diagram of coordinate transformation with a right inclination angle.

TABLE 4: Comparison of P value vectors of the same camera with different heights.

H/mm	$v00$	$v01$	$v02$	$v03$	$v04$	$v10$	$v11$	$v12$	$v13$	$v20$	$v21$	$v22$
229	$5.6e+02$	$1.5e+01$	$-2.7e-01$	$2.03e-03$	$-5.4e-06$	$5.3e-01$	$-1.4e-02$	$1.3e-04$	$-4.7e-07$	$-2.0e-04$	$7.78e-06$	$-4.9-08$
239	-1.744	$3.1e+01$	$-6.1e-01$	$5.73e-03$	$-1.8e-05$	$3.99e-02$	$3.1e-03$	$-9.3e-05$	$3.8e-07$	$-2.8e-03$	$9.79e-05$	$-8.3e-07$

Table 5: Comparison of P value vectors of different cameras with the same height.

Camera	v00	v01	v02	v03	v04	v10	v11	v12	v13	v20	v21	v22
A	$-7.9e-04$	$4.4e-02$	-0.9042	$8.27e-04$	$-2.8e-07$	5.585	$-1.3e+02$	$1.75e-05$	$5.81e+09$	$-5.6e-03$	$3.6e-06$	$-1.3e-08$
B	$3.9e-04$	$-1.8e-02$	0.3128	$-2.5e-04$	$7.38e-08$	6.412	$-2.6e-02$	$-4.8e-06$	$-1.5e-08$	$-2.0e-03$	$3.5e-05$	$3.0e-09$

TABLE 6: Calibration results.

Method	Intrinsic matrix	Radial distortion	Tangential distortion
Zhang [15]	$\begin{bmatrix} 739.72 & 0 & 0 \\ 0 & 729.25 & 0 \\ 590.2 & 618.45 & 1 \end{bmatrix}$ pixel	$[\,0.01 \quad -0.088\,]$ pixel	$[\,0\ 0\,]$
Li [20]	$\begin{bmatrix} 729.84 & 0 & 0 \\ 0 & 725.53 & 0 \\ 606.92 & 605.99 & 1 \end{bmatrix}$ pixel	$[\,0.01 \quad -0.088\,]$ pixel	$[\,0\ 0\,]$

Method	Focal length (regression model)
Proposed	$10.37 - 9.465 * x + 1.478 * y + 2.871 * x\char`^2 - 6.653 * x * y + 4.285 * x * y\char`^2 - 2.029 * y\char`^3 + 1.354 * x$ $* y\char`^2 - 6.76 * y\char`^2 + 3.935 * x\char`^2 * y + 3.348 * x * y\char`^3 + 1.006 * y\char`^4$
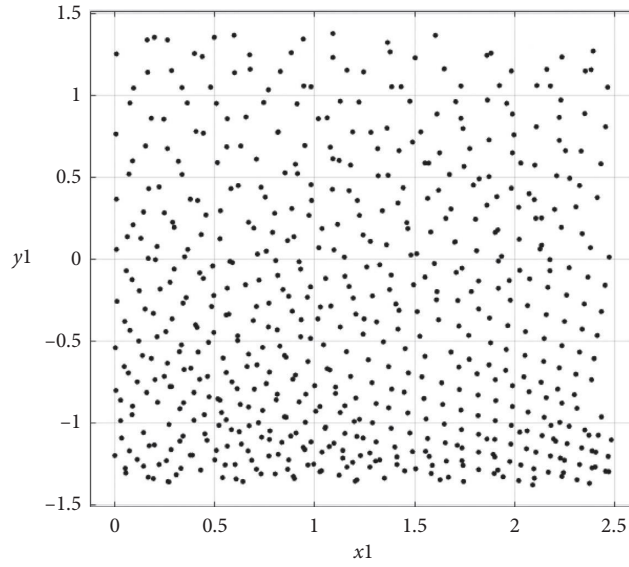
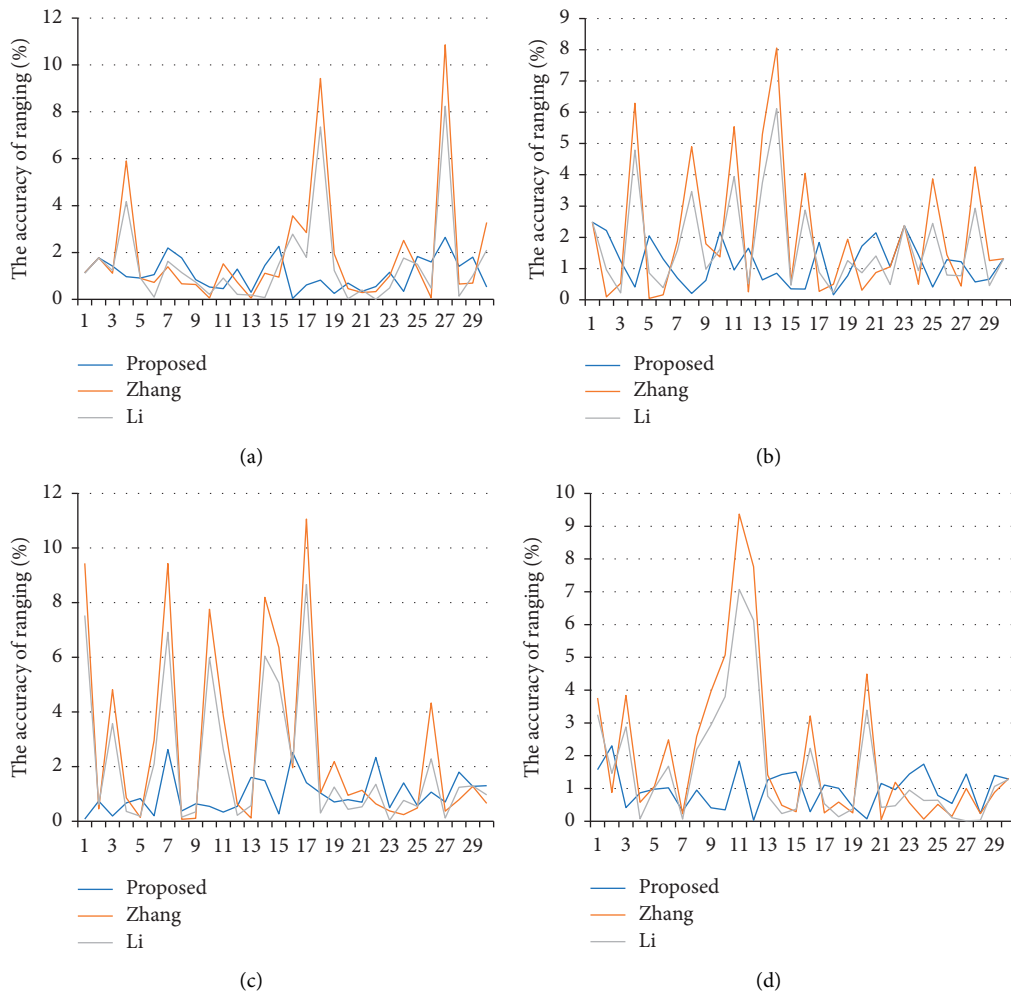FIGURE 10: Pixel distribution of measured points.

(a)

(b)

(c)

(d)

FIGURE 11: Camera inclination angle decomposition.

TABLE 7: The accuracy of the ranging model.

Method	Ranging accuracy
Bao's [5]	97%
Meng's [6]	More than 90%
Zhe's [24]	Approximate 98%
Proposed	Average 97.09%, maximum 99.02%

6. Conclusions

In this study, a camera's calibration for a monocular vision-based ranging model based on polynomial regression was proposed to establish the relationship between world coordinates and pixel coordinates. For the possible distortion of the captured images, we utilize an error compensation algorithm to revise the pixel coordinates. Compared with other methods, our method has a high score for matching. The process of regression and error compensation can effectively compensate for the errors caused by distortion without the distortion parameters of the convex lens. The experimental results show that the accuracy of ranging in this study is more than 97%. In the future, we will combine our method with image recognition to improve safety in fully autonomous driving.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

References

[1] D. S. Varuna, R. Jamie, and K. Ahmet, "Robust fusion of LiDAR and wide-angle camera data for autonomous mobile robots," *Sensors*, vol. 18, no. 2730, Article ID 18082730, 2018.

[2] B. J. Tippetts, D. J. Lee, and J. K. Archibald, "An on-board vision sensor system for small unmanned vehicle applications," *Machine Vision and Applications*, vol. 23, no. 3, pp. 403–415, 2012.

[3] P. Ferrara, A. Piva, and F. Argenti, "Wide-angle and long-range real time pose estimation: a comparison between monocular and stereo vision systems," *Journal of Visual Communication and Image Representation*, vol. 48, pp. 159–168, 2017.

[4] R. A. Vincent and M. R. Hawks, "Passive ranging of dynamic rocket plumes using infrared and visible oxygen attenuation," *Proceedings of the SPIE-the International Society for Optical Engineering*, vol. 8052, no. 3, Article ID 883470, 2011.

[5] D. Bao and P. Wang, "Vehicle distance detection based on monocular vision," in *Proceedings of the 2016 International Conference On Progress In Informatics And Computing (PIC)*, IEEE, Shanghai, China, December 2016.

[6] C. Meng, H. Bao, and Y. Ma, "Visual meterstick: preceding vehicle ranging using monocular vision based on the fitting method," *Symmetry*, vol. 11, no. 9, Article ID 11091081, 2019.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster RCNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[8] D. Chwa, A. Dani, and W. Dixon, "Range and motion estimation of a monocular camera using static and moving objects," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 4, pp. 1–10, 2015.

[9] W. Lu, T. T. Wang, and J. H. Chu, "The method of real-time distance measurement based on monocular vision," *Advanced Materials Research*, vol. 403–408, pp. 1451–1454, 2012.

[10] Y. Yakimovsky and R. Cunningham, "A system for extracting three-dimensional measurements from a stereo pair of TV cameras," *Computer Graphics and Image Processing*, vol. 7, no. 2, pp. 195–210, 1978.

[11] H. A. Martins, J. R. Brik, and R. B. Kelley, "Camera models based on data from two calibration planes," *Computer Graphics and Image Processing*, vol. 17, no. 2, pp. 173–180, 1981.

[12] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 2003.

[13] R. Chen, Z. Li, and K. Zhong, "A stereo-vision system for measuring the ram speed of steam hammers in an environment with a large field of view and strong vibrations," *Sensors*, vol. 19, Article ID 19050996, 2019.

[14] Q. T. Luong and O. D. Faugeras, "Self-calibration of a moving camera from point correspondences and fundamental matrices," *International Journal of Computer Vision*, vol. 22, no. 3, pp. 261–289, 1997.

[15] Z. Y. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[16] W. Dong and V. Isler, "A novel method for the extrinsic calibration of a 2D laser rangefinder and a camera," *IEEE Sensors Journal*, vol. 1, p. 99, 2016.

[17] Y. Xu, "A calibration method for non-overlapping cameras based on mirrored absolute phase target," *International Journal of Advanced Manufacturing Technology*, vol. 104, no. 1, pp. 9–15, 2019.

[18] X. M. Wu, F.-L. Guan, and A.-J. Xu, "Passive ranging based on planar homography in a monocular vision system," *Journal of Information Processing Systems*, vol. 16, no. 1, 2020.

[19] X. Han and CEPREI, "Research on UAV ranging algorithm based on monocular vision," *Electronic Product Reliability and Environmental Testing*, vol. 37, no. 3, 2019.

[20] B. Li, L. Heng, and K. Koser, "A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots & Systems IEEE*, Chicago, IL, USA, September 2014.

[21] Z. Zhang, R. Zhao, and E. Liu, "A convenient calibration method for LRF-camera combination systems based on a checkerboard," *Sensors*, vol. 19, no. 6, p. 1315, 2019.

[22] P. Zhang, "Calibration error simulating analysis of binocular vision based on pinhole camera model," *Electronics Optics and Control*, vol. 16, no. 12, pp. 65–69, 2009.

[23] T. Tianyang, Q. Chen, S. Feng, Y. Hu, and C. Zuo, "Active depth estimation from defocus using a camera array," *Applied Optics*, vol. 57, no. 18, Article ID 004960, 2018.

[24] T. Zhe, L. Q. Huang, and Q. Wu, "Inter-vehicle distance estimation method based on monocular vision using 3D detection," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, pp. 4907–4919, 2020.