Research on location method based on monocular vision

Hao Chen1,a, Meian lin2,b*, Lixia Xue3,c, Tian Gao 4,d, Haojie Zhu5,e

1 College of computer and information engineering, Inner Mongolia Agricultural University, Inner Mongolia, China
2 College of computer and information engineering, Inner Mongolia Agricultural University, Inner Mongolia, China
3 College of computer and information engineering, Inner Mongolia Agricultural University, Inner Mongolia, China
4 College of computer and information engineering, Inner Mongolia Agricultural University, Inner Mongolia, China
5 College of computer and information engineering, Inner Mongolia Agricultural University, Inner Mongolia, China

aemail: chenhaomeng@emails.imau.edu.cn
bCorresponding author: bemail: 20071036@imau.edu.cn
cemail: xue2019@emails.imau.edu.cn,demail: 2281560360@qq.com,eemail: 614321916@qq.com

Abstract: Monocular vision based positioning method is easy to operate and has a wide application prospect. The monocular vision ranging method proposed in this paper can get the position of the target object from the camera and the angle away from the camera, so as to achieve accurate positioning. In addition, the world coordinate system can be constructed, so that the distance between two or more target objects can be obtained, and the accuracy is high.

1. Introduction

In recent years, with the rapid development of computer vision technology and the rapid improvement of hardware performance, vision sensor is widely used because it can collect rich environmental information, low cost and easy to use. The positioning method based on vision sensor has become the focus of research. Compared with radar or laser equipment, camera and other equipment have the advantages of low cost, wide application range, strong portability, and can collect rich information content. According to the number of applied vision sensors, the vision positioning methods are divided into monocular vision positioning, binocular vision positioning and multicular vision positioning [1]. Among them, the most simple and convenient one is monocular vision positioning, which only uses one camera to complete the positioning work. Monocular vision positioning operation is simple, less calibration steps, wide applicability, does not need to solve the problem of feature point matching in binocular vision, consider the distortion problem caused by multi vision positioning, and avoids the defects of small field of vision, stereo matching and so on. The vigorous development of artificial intelligence and the rise of small UAV and autopilot technology, but also to the development of machine vision into a steady stream of power, because the monocular vision positioning method is simple to
operate, strong practical, it has been the focus of academic and engineering circles. Monocular vision positioning methods are divided into single frame based positioning method and double frame or multi frame based positioning method[2]. The positioning based on single frame image is to complete the positioning work according to the information of one frame image. In this way, we need to manually set some feature points of the object in advance, and use the model of light propagation along the line and small hole imaging to obtain the camera parameters to calibrate the camera. This method has the characteristics of simple form, small amount of calculation, easy algorithm implementation and low hardware requirements, but it has the disadvantages of poor robustness, real-time performance and dependence on manual marking. The location method based on two or more images is to get two or more images at different times in the same scene, and complete the target location according to the position deviation of the target feature points in different images. The camera calibrated in advance is used to capture the target in motion. The projection points of the target in two adjacent frames are used to form multiple matching point pairs to estimate the basic matrix. The essential matrix is further obtained from the relationship between the basic matrix and the essential matrix, and the external motion parameters (rotation and translation) of the monocular camera are obtained through decomposition, Three dimensional information of the target is obtained by coordinate system transformation [3]. This method is complex and has matching problems, but it does not rely on manual marking.

Most of the current monocular vision positioning methods are based on the ranging model, which only calculates the relative distance between the camera and the target object, but does not know the actual position of the target object relative to the camera. In addition, some positioning models using deep learning have a large number, cumbersome calculation, high hardware requirements, and need strong computing power to complete, which are not suitable for on-board computers, small UAVs and some machines with low computing power.

The monocular vision positioning algorithm proposed in this paper is based on the vision positioning method of single frame image, which is obtained by using the light propagation along a straight line and the small hole imaging model. This method can not only get the distance between the object and the camera, but also get the world coordinates of the object and the orientation of the camera. Compared with other algorithms, the accuracy of this algorithm is also improved. It can be applied to the construction of map in vehicle computer and UAV, the obstacle avoidance of robot walking and the object grasping of robot arm.

2.Materials and Methods

2.1. Building models

This model is based on the monocular vision ranging system of pinhole imaging [4]. The camera and imaging principle are shown in Figure 1.

Two coordinate systems are constructed, which are pixel coordinate system and world coordinate system. H is the height of the camera from the horizontal ground, α It is the angle between the optical axis and the horizontal line. P is the characteristic point of the target on the ground. Let the world coordinates of point p be p (x, y), and the coordinates of the imaging plane of point p be p (x, y). M and N are the intersection points of the extended line of p'p'x and the extended line of o'p'y with the horizontal plane of the optical center O, respectively. In the pinhole imaging model, we consider the distance between the optical center O and the origin o ' of the image plane as the focal length f of the camera. According to the principle of camera projection imaging and the spatial geometric relationship, the horizontal distance from the feature point P to the camera can be deduced:

\[
    d = \frac{H(\sqrt{x^2+f^2}+y+f\tan\alpha)}{\sqrt{x^2+f^2}(f+y)} \tag{1}
\]
3

In the world coordinate system, the Y axis is parallel to the main optical axis of the camera, and the X axis is parallel to Mn. Because Mo and o’p are the projection of p’p on the horizontal plane and the horizontal ground where the camera is located, Mo is parallel to o’p. So we can get the following results \( \Delta OMN \sim \Delta O'PP' \).

It can be seen from the geometric model that:

\[
\begin{align*}
OOM' &= f \\
MN &= O'P'x = x \\
ON &= \frac{f}{\cos \alpha} \\
OM &= \frac{f^2}{\sqrt{\cos \alpha^2}} + x^2
\end{align*}
\]

Simultaneous (2), (3), (4), (5) get (6), (7), (8):

\[
\begin{align*}
\cos \beta &= \frac{f}{\sqrt{f^2 + x^2 \cos \alpha^2}} \\
\sin \beta &= \frac{x \cos \alpha}{\sqrt{f^2 + x^2 \cos \alpha^2}} \\
\beta &= \arccos \left( \frac{f}{\sqrt{f^2 + x^2 \cos \alpha^2}} \right)
\end{align*}
\]

Simultaneous (1), (6), (7) get (9), (10):

\[
\begin{align*}
X &= \frac{H(x^2 + f^2 - y^2 \tan \alpha) \sin \beta}{\sqrt{x^2 + f^2 - (f \tan \alpha + y)^2}} \\
Y &= \frac{H(x^2 + f^2 - y^2 \tan \alpha) \cos \beta}{\sqrt{x^2 + f^2 - (f \tan \alpha + y)^2}}
\end{align*}
\]

The distance between the target object and the camera is obtained by formula (1), the world coordinates of the target object are obtained by formula (10), (11), and the deflection angle from the target object to the camera is obtained by formula (9).

2.2. Experimental steps and methods

(1) Camera calibration

Equipment used in the experiment: a hand-held laser rangefinder, a camera (including bracket), a laptop computer, a gyroscope, a 50m tape measure, an angle measuring instrument, several bookshelves as the target object.

Firstly, the laser rangefinder is used to measure the distance between the camera and the horizontal ground 645mm, and the gyroscope is used to measure the left and right inclination angle of the camera 0°, the up and down dip angle is 66.56°.

Because the camera is shooting with a fixed attitude, the height h and the inclination angle of the camera \( \alpha \) is a fixed value. In order to solve the camera internal reference focal length \( f \), we need to
measure the distance \( d \) from the 12 feature points to the camera and the pixel coordinates \((x, y)\) of the feature points, and use the formula (1) to calculate the focal length \( f \) corresponding to the 12 feature points. The formula is as follows:

\[
\begin{align*}
(d^2 \cdot \tan(a) - h^2) \cdot f^4 + (2 \cdot yf \cdot \tan(a) \cdot (d^2 + h^2)) \cdot f^3 + (d^2 \cdot yf^2 + d^2 \cdot \tan(a) \cdot 2 \cdot h^2 \cdot xf^2) \cdot f^2 + (2 \cdot yf \cdot xf^2 \cdot \tan(a) \cdot (d^2 + h^2)) \cdot f + xf^2 \cdot (d^2 \cdot yf^2 - h^2 \cdot xf^2) &= 0
\end{align*}
\]

The four roots of \( F \) can be solved by solving function of MATLAB. Because the physical meaning of focal length determines that \( f > 0 \) is a real number, the value of \( F \) can be determined from it.

According to the relationship between the pixel coordinates \((x, y)\) and the real focal length \( f' \), the fitting focal length \( f' \) is obtained by using MATLAB.

\[
f'(x, y) = p00 + p10 \cdot x + p01 \cdot y + p20 \cdot x^2 + p11 \cdot x \cdot y + p02 \cdot y^2 + p30 \cdot x^3 + p21 \cdot x^2 \cdot y
\]

The principle of \( F(x, y) \) relation fitting is shown in figure:

Figure 2 The principle of \( F(x, y) \) relation fitting

The fitting focal length of some data is shown in Table 1 (unit: mm):

| \( H \) | \( f \) | \( d \) | \( f' \) |
|-------|-------|-------|-------|
| 645   | 2.373 | 1000  | 2.308 |
| 645   | 2.561 | 1100  | 2.438 |
| 645   | 2.486 | 1400  | 2.807 |
| 645   | 2.467 | 900   | 1.846 |
| 645   | 2.006 | 1200  | 1.504 |
| 645   | 1.604 | 1300  | 1.662 |
| 645   | 1.142 | 1500  | 1.049 |
| 645   | 2.418 | 1600  | 3.00  |
| 645   | 4.320 | 1700  | 3.024 |
| 645   | 3.924 | 1800  | 3.094 |
| 645   | 3.245 | 1900  | 3.442 |

(2) Measurement data

After the camera is calibrated, the downward inclination angle of the camera is arranged according to the angle diagram 4 of the characteristic points of the target object, and the characteristic points are deviated by 10 degrees on the left and right of the main optical axis of the camera with a tape along the angle division 10°, 20°, 30°, 40°. The target objects are arranged at 1m, 1.1m, 1.2m, 1.3m, 1.4m, 1.5m, 1.6m, 1.7m, 1.8m and 1.9m away from the camera. In order to avoid the problem of mutual occlusion of target objects in the future, only four target objects are arranged at a time to take a picture.

The obtained image is processed with Python to obtain the pixel coordinates of each target object feature point, as shown in the figure. Because the construction of the pixel coordinate system is different from the actual pixel coordinate system, the obtained pixel coordinates need to be processed. The processing formula is as follows:
Finally, the obtained pixel coordinates (x, y) and fitting focal length \( f' \) of the target object feature points are brought into formula (1), and the distance \( d \) between the target object and the camera can be obtained. Then, the obtained distance \( D \), pixel coordinates (x, y) and fitting focal length \( f' \) are put into formula (6), formula (7), formula (8), formula (9) and formula (3) respectively Formula (10) can get the world coordinates of the target object and the orientation of the target object in the camera, which can realize the accurate positioning of the target object.

3. Results & Discussion

3.1. experimental result

The calculation results are shown in Table 2:

| x     | y     | \( \beta_1 \) | X      | Y      | \( \beta \) |
|-------|-------|---------------|--------|--------|------------|
| 1.248 | 0.6578| 30°           | 448.9314 | 893.1662 | 28.79°      |
| 1.2636| 0.4992| 30            | 477.1935 | 991.1035 | 27.59°      |
| -1.6614| 0.5434| -30           | -530.7620| 963.4789 | -31.56°     |
| 0.2496| 0.6370| 10            | 112.1934 | 892.9796 | 7.20°       |
| 0.2600| 0.3276| 10            | 999.8260 | 18.6490  | 9.19°       |
| 0.2626| 0.2106| 10            | 174.2290 | 1086.1141| 8.41°       |
| 0.2678| 0.1066| 10            | 174.3175 | 1187.2714| 13.55°      |
| -0.7410| -0.2626| 10            | 29.9792  | 1246.9160| 13.12°      |
| -0.754 | -0.3198| -10           | -428.8629| 1850.9664| -13.27°     |
| -0.7566| -0.3692| -10           | -443.1889| 1950.2778| -13.02°     |

Because the field of view of the camera is 160°. The maximum deviation of the target object from the camera is 80°. So, we use \( \Delta \beta \), the angle error can be expressed by formula (14) \( \beta \). The accuracy of the test.

\[
\Delta \mu = \left( 1 - \frac{\Delta \beta}{80} \right) \times 100
\]

Table 3 Partial deflection angle \( \beta \) Accuracy results

| x     | y     | \( \beta_1 \) | \( \beta \) | \( \Delta \beta \) | \( \Delta \mu \) |
|-------|-------|---------------|------------|--------------|------------|
| 1.248 | 0.6578| 30°           | 28.79°     | 1.21°        | 98.5       |
| 1.2636| 0.4992| 30            | 27.59°     | 2.41°        | 97.0       |
| -1.6614| 0.5434| -30           | -31.56°    | 1.56°        | 98.1       |
| 0.2496| 0.6370| 10            | 7.20°      | 2.8°         | 96.5       |
| 0.2600| 0.3276| 10            | 9.19°      | 0.91°        | 98.9       |
| 0.2626| 0.2106| 10            | 8.41°      | 1.59°        | 98.0       |
| 0.2678| 0.1066| 10            | 13.55°     | 3.55°        | 95.6       |
| -0.7410| -0.2626| 10            | 13.12°     | 3.12°        | 96.1       |
| -0.754 | -0.3198| -10           | -13.27°    | 3.27°        | 95.9       |
| -0.7566| -0.3692| -10           | -13.02°    | 3.02°        | 96.2       |

The accuracy of boundary coordinates is expressed by formula (15) and formula (16):

\[
d_1 = \sqrt{X^2 + Y^2}
\]

\[
\Delta D = \frac{d_1}{d} \times 100
\]

Table 4 Partial world coordinate accuracy

| d     | X      | Y      | d1     | \( \Delta D \) |
|-------|--------|--------|--------|--------------|
| 1000  | 448.9314 | 893.1662 | 999.6426 | 99.96       |

\( x = (x_1 - 960) \times 0.0026 \)  \( y = (y_1 - 540) \times 0.0026 \)
3.2. analysis of experimental results
It can be seen from table 4 that the deflection angle is obtained $\beta$, the accuracy is more than 95%. By calculating the distance between the target object and the camera, and then calculating the angle of the target object from the camera, accurate positioning can be achieved.

It can be seen from table 5 that the world coordinates of the target object are accurate, and the accuracy rate is more than 99.99%. It can be used to build a real-time map with camera as the center and get the distance between two target objects.

3.3. error analysis and improvement
Deflection angle required $\beta$, the main reason for the large error is that the main optical axis of the camera is not completely parallel to the Y axis of the world coordinate system, and the tape is prone to offset when measuring the deflection angle of the target object. Later, we are going to change the experimental equipment, use the laser rangefinder to determine the position of the object to be measured, and make sure that the main optical axis of the camera is parallel to the Y axis of the world coordinate system.

4. Conclusions
The method proposed in this paper can calculate the distance between the target object and the camera, at the same time, it can get the deflection angle of the target object relative to the camera, so as to achieve accurate positioning. Moreover, the world coordinate system is constructed, which can get the world coordinate of the target object, so as to calculate the distance between two or more target objects. It can be used for unmanned driving to construct self centered world coordinate system, judge whether the width of the road ahead can pass, and for intelligent robots and UAVs to travel to avoid obstacles, determine the location information of the target object, etc.

Acknowledgments
This research was funded by the National Natural Science Foundation of China "Research on accurate ranging method based on vehicle monocular vision" (granted No.61741207) and the Inner Mongolia natural Science Foundation "Agricultural Machinery Obstacle Avoidance Algorithms Based on CNN and Vehicle monocular Vision Dynamic Ranging Model" (granted No.2019MS06016).

References
[1] Na Zhou. (2007)Research on camera positioning technology based on monocular vision.d. Nanjing: Nanjing University of Aeronautics.
[2] Rongming Li, Libin Lu, Guodong Jin.(2011)Review of monocular vision positioning methods. J. Modern computer (Professional Edition)., 11: 9-12.
[3] Guoguo Zhang. (2009)Research on positioning system based on monocular vision .d. Wuhan: Huazhong University of science and technology.
[4] Meian Li, Haiping Zhao, Baozhu Shi, Run Tian, Yang Liu. A monocular vision ranging method.P. Inner Mongolia Autonomous Region: cn109489620b, 2020-12-25.