UDC 621.391

IMPLEMENTATION PRINCIPLE OF MONOCULAR RANGING

QICHENG GUO

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received April 8, 2024

Abstract. As a commonly used measurement technology, monocular ranging is widely used in various fields due to its advantages such as simplicity, ease of use, low cost, and real-time measurement. This article will explain the principles and implementation methods of monocular ranging, application fields, and its potential challenges and development directions.

Keywords: monocular ranging, principles and methods, application areas, development directions.

Introduction

Monocular ranging uses image information obtained by a single camera or sensor to infer the distance between an object and the observer through a series of algorithms and methods [1]. In the field of computer vision and machine perception, monocular ranging technology plays an important role. It is a key component to realize autonomous navigation, environment perception and distance measurement of intelligent systems. Common monocular distance measurement methods include: Stereo Vision, Structured Light, Motion-based Methods, Focus Variation, and Texture Features.

Monocular ranging principle

1. Camera imaging model.

In the monocular distance measurement method, to obtain distance information, you need to obtain a point in the three-dimensional real world. Since the processed object is a two-dimensional plane image captured by a camera, therefore, how to convert a point on a two-dimensional image into a point in the three-dimensional world is a problem that must be considered. Converting points on the image to points in the real world requires mutual conversion between the pixel coordinate system, the image coordinate system, the camera coordinate system and the world coordinate system [2]. The interrelationships between the four coordinate systems are shown in Figure 1. The coordinate system is described as follows:

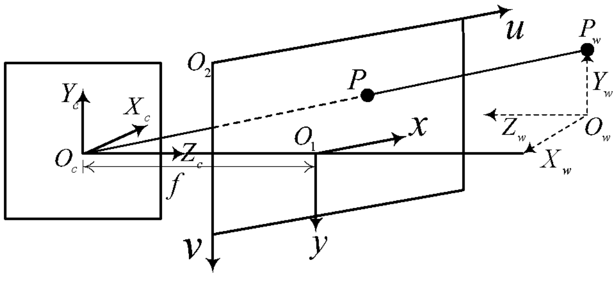


Figure 1. The relationship between the four coordinate systems

Pixel coordinate system. Digital images are generally three-dimensional images and are composed of many pixels. The origin of the pixel coordinate system is O_2, with the width direction as the u-axis and the height direction as the v-axis.

Image coordinate system. The origin of the image coordinates is O_1, and the pixel coordinate system and the image coordinate system are parallel, with the image width direction as the *x*-axis, the height direction as the *y*-axis, and the length unit is *mm*.

Camera coordinate system. The origin O_c of the camera coordinate system, the X_c axis and the Y_c axis are parallel to the *x*-axis and *y*-axis of the image coordinate system respectively, and the camera Z_c axis coincides with the camera optical axis.

World coordinate system. The environment we are in is under the world coordinate system, which is the plane X_w Y_w Z_w in Figure 1. P_w completes the conversion from world coordinates to coordinates on the image from a point in the real world to point *P* on the image.

2. Coordinate system conversion.

2.1. Convert pixel coordinate system to image coordinate system.

The pixel coordinate system represents the position information of each pixel in pixels, but it cannot express the physical size of the object in the image, so conversion between coordinate systems is required.

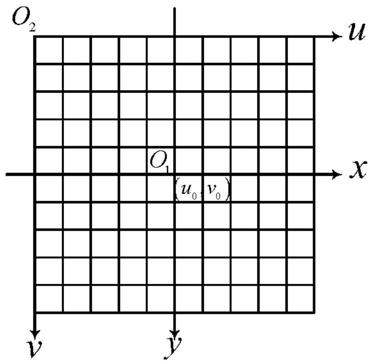


Figure 2. Convert pixel coordinate system to image coordinate system

In Figure 2, the relationship between the coordinates (*x*, *y*) of the image coordinate system and the coordinates (*u*, *v*) of the pixel coordinate system can be expressed as formula (1)

$$\begin{cases} x = udx - u_0 dx \\ y = vdy - v_0 dy \end{cases}, \tag{1}$$

where (u_0, v_0) are the pixel coordinates of the image center, *dx* and *dy* are the unit physical lengths of the horizontal and vertical pixels on the photosensitive plate respectively. The form written as a homogeneous coordinate matrix is formula (2)

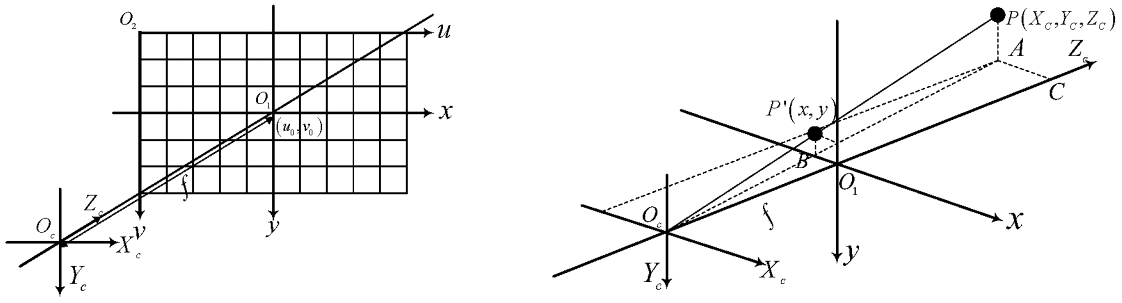
$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} dx & 0 & 0 \\ 0 & dy & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \\ 0 \end{bmatrix} + \begin{bmatrix} -u_0 dx \\ -v_0 dy \\ 1 \end{bmatrix} = \begin{bmatrix} dx & 0 & -u_0 dx \\ 0 & dy & -v_0 dy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}. \tag{2}$$

2.2. Transform the image coordinate system to the camera coordinate system

Figure 3 shows the process of imaging the object into the image coordinate system. The distance between $O_c O_1$ is the focal length *f*. Point *P* and point *P'* are the coordinates in the camera coordinate system and image coordinate system respectively.

It is easy to see from the above figure that the triangle $O_c O_1 B$ is similar to the triangle $O_c CA$, and the triangle $O_c BP'$ is similar to the triangle $O_c AP$. According to the principle of similar triangles, there is formula (3)

$$\frac{O_c O_1}{O_c C} = \frac{O_1 B}{CA} = \frac{O_c B}{O_c A} = \frac{P' B}{PA}. \tag{3}$$



a *b*

Figure 3. The process of imaging an object into the image coordinate system: *a* – camera coordinate system; *b* – similar triangle model

And the distance of O_cO_1 is the focal length f. Combined with $P(X_C, Y_C, Z_C)$ and $P'(x, y)$ point coordinates, the above formula can be written as (4)

$$\frac{f}{Z_c} = \frac{x}{X_c} = \frac{y}{Y_c}. \tag{4}$$

By further pushing it down, we can get formula (5)

$$\begin{cases} X_c = \dfrac{xZ_c}{f} \\ Y_c = \dfrac{yZ_c}{f} \end{cases}. \tag{5}$$

The form of writing it as a homogeneous coordinate matrix is formula (6)

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{Z_c}{f} & 0 & 0 \\ 0 & \dfrac{Z_c}{f} & 0 \\ 0 & 0 & Z_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{6}$$

2.3. Transform the camera coordinate system to the world coordinate system.

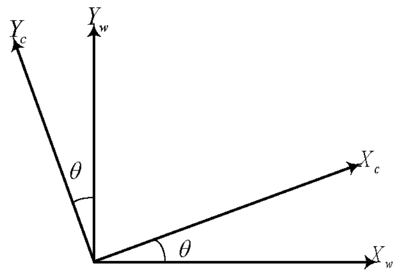


Figure 4. Convert pixel coordinate system to image coordinate system

As shown in Figure 4, the transformation from the camera coordinate system to the world coordinate system can be described as a process of rotation and translation. Adding up the components of rotation and translation respectively is the entire process of coordinate system transformation.

For the rotation process, assuming that the camera rotates around the *z*-axis of the coordinate system, there is formula (7)

$$\begin{cases} X_w = X_c \cos\theta - Y_c \sin\theta \\ Y_w = X_c \sin\theta + Y_c \cos\theta \\ Z_w = Z_c \end{cases}. \tag{7}$$

In the same way, rotating around the x-axis will yield formula (8)

$$\begin{cases} X_w = Z_c \\ Y_w = Y_c \cos\alpha - Z_c \sin\alpha \\ Z_w = Y_c \sin\alpha - Z_c \cos\alpha \end{cases}. \tag{8}$$

Rotating around the y-axis will yield formula (9)

$$\begin{cases} X_w = Z_c \sin\beta + X_c \cos\beta \\ Y_w = Y_c \\ Z_w = Z_c \cos\beta - X_c \sin\beta \end{cases}. \tag{9}$$

For the translation component, it can be expressed as formula (10)

$$\begin{cases} X_w = X_c + T_x \\ Y_w = Y_c + T_y \\ Z_w = Z_c + T_z \end{cases}. \tag{10}$$

After obtaining the translation vector and rotation matrix, the formula for transforming from the camera coordinate system to the world coordinate system can be completely written as formula (11)

$$\begin{bmatrix} X_w & Y_w & Z_w & 1 \end{bmatrix} = \begin{bmatrix} X_c & Y_c & Z_c & 1 \end{bmatrix} \begin{bmatrix} R & 0 \\ T & 1 \end{bmatrix}. \tag{11}$$

The rotation matrix R is expressed as formula (12)

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha & 0 \\ 0 & -\sin\alpha & \cos\alpha & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & -\sin\beta & 0 \\ 0 & 1 & 0 & 0 \\ \sin\beta & 0 & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta & 0 & 0 \\ -\sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{12}$$

The translation matrix T is expressed as: $\begin{bmatrix} T_x & T_y & T_z \end{bmatrix}$. The combined formula completes the conversion from the pixel coordinate system to the world coordinate system, and is expressed as formula (13)

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{Z_c dx}{f} & 0 & \frac{-Z_c u_0 dx}{f} \\ 0 & \frac{Z_c dy}{f} & \frac{-Z_c v_0 dy}{f} \\ 0 & 0 & Z_c \\ 0 & 0 & 1 \end{bmatrix}. \tag{13}$$

In this way, for a point on the image, the specific distance value can be obtained using the above formula (13) combined with the internal and external parameters of the camera.

Application fields of monocular ranging

Monocular ranging technology is widely used in the following fields [3]:

1. Robot technology: In the field of robots, monocular ranging is the key to achieving autonomous navigation and obstacle avoidance. Through monocular ranging, the robot can sense the distance of objects in the surrounding environment to plan the best path and avoid obstacles.
2. Computer vision: In computer vision, monocular ranging can help the system understand the position and size of objects in the image, thereby achieving tasks such as target detection, tracking, and three-dimensional reconstruction.
3. Virtual and augmented reality: In virtual and augmented reality applications, monocular ranging can help the system accurately superimpose virtual objects into the real world to achieve a realistic virtual experience.
4. Self-driving cars: For self-driving cars, accurate distance measurement is an important factor in ensuring safe driving. Monocular ranging can help the vehicle sense the distance of other vehicles, pedestrians and obstacles in the surrounding environment, thereby making timely driving decisions.
5. Medical imaging: In the medical field, monocular ranging can help doctors accurately measure the location and size of organs, tumors, etc. in the patient's body, thereby guiding the diagnosis and treatment process.
6. Military and security: Monocular ranging also plays an important role in military reconnaissance, target tracking, security monitoring and other fields. By accurately measuring the distance to a target, military personnel can be helped to make appropriate tactical decisions and safety measures.

Future challenges and development directions

The challenges and future development directions of monocular ranging technology are as follows:

1. Accuracy and robustness: Improve the accuracy and robustness of monocular ranging methods to adapt to complex and changeable environments.
2. Real-time and efficiency: Optimize the algorithm and hardware to improve the real-time and efficiency of the monocular ranging algorithm.
3. Cross-field integration: Integrate monocular ranging technology with other sensing technologies and artificial intelligence algorithms to expand its applications in different fields.

Conclusion

Monocular ranging is an important measurement technology. It is widely used in many fields due to its fast speed, simple calculation, and low price [4]. However, it still has shortcomings such as low accuracy and poor robustness [5]. With the continuous development of computer vision and artificial intelligence technology, monocular ranging technology still has great room for development and potential.

References

1. Chen S, Mei S, Jia G. // Journal of Internet Technology. 2021. Vol. 22(5). P. 1131–1142.
2. Lu W, Wang T T, Chu J H. // Advanced Materials Research. 2012. Vol. 403. P. 1451–1454.
3. Witus G, Hunt S. // Unmanned Systems Technology X. SPIE. 2008. Vol. 6962. P. 38–44.
4. Chwa D, Dani A P, Dixon W E. // IEEE Transactions on Control Systems Technology. 2015. Vol. 24(4). P. 1174–1183.
5. Nakamura K, Ishigaki K, Ogata T. // 2013 IEEE Intelligent Vehicles Symposium (IV). IEEE. 2013. P 1368–1373.