

Research on Pose Measurement Based on Monocular Vision

Haoyi Li¹, Chunting Ma¹, Huiyong Deng¹ and Hao Chu²

¹Army Engineering University, Shijiazhuang 050003, China

²95857 Army, Xiaogan 432000, China

*Corresponding author e-mail: 455996796@qq.com

Abstract. In recent years, pose measurement based on vision system has become a research focus. As pose measurement based on monocular vision is simple and stable with low cost and high calculation speed, it has been widely applied in reality. This paper summarizes the pose measurement based on monocular vision, and emphasizes on pose measurement based on characteristic points, lines and advanced geometry features, with analysing each advantages and disadvantages. Ultimately, it makes a brief analysis on the future of pose measurement based on monocular vision in order to provide references on research on pose measurement.

1. Introduction

Pose measurement becomes more important in modern control, navigation and many other fields [1-5]. It requires the measurement system with high accuracy and simple structure. Computer vision position and orientation measurement system has been widely paid attention in many applications such as spacecraft docking at short range and field installation of large objects. So it is significant to research on pose measurement in computer vision.

In recent years, with the development of vision sensor, the method using vision sensors to get target characteristic so as to position has been a focus because the required experimental environment and hardware conditions are more relaxed with low cost and effective in reality.

The vision systems fall into two major categories: monocular vision [1-3] and binocular vision [4, 5]. Monocular vision system applies one camera to screen one image to acquire the location and position, while binocular vision system uses two images to acquire 3D object information by two different positions camera screening the same object, and then extracts characteristic points on the object resulting in spatial location information, and calculates the parameter by the 3D match point pair on that two images. In which, the monocular vision system has many advantages such as simple structure and large field of view (FOV) for measurement, but its measurement accuracy is lower than that of the binocular vision system, especially for measurement of depth, where a large error may exist. At the same time, there is accuracy disparity between vision systems and traditional methods, such as laser tracker [6].

The rest of paper is organized as follows: Section 2 describes the pose measurement based on characteristic points, lines and advanced geometry features and the measurement model, including the existing methods applied in reality. Section 3 makes a conclusion.



2. Pose Measurement Based on Monocular Vision

According to the number of positioning images, pose measurement based on monocular vision can be divided into the measurement based on single image and two or more images.

2.1. Pose Measurement Based on Single Image

2.1.1. Measurement Model. Based on single image, it mainly use the information of an image to acquire the posture parameters. Target localization model based on single image is mentioned as in Figure 1.

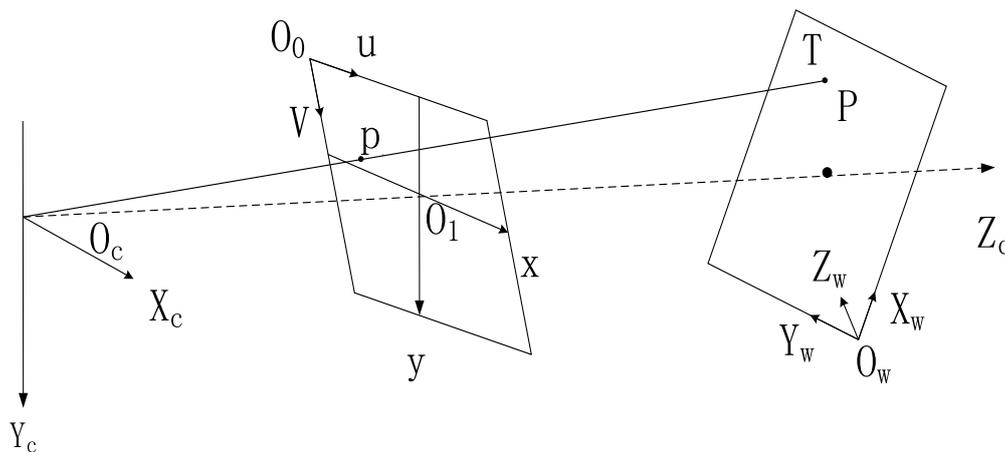


Figure 1. Target localization model based on single image model

In Figure1, O_c is the optical center of the camera. (O_c, X_c, Y_c, Z_c) is the camera coordinate system. X_c, Y_c are parallel to the x and y axis of image plane. Z_c is the camera axis. O_1 is intersection of optical axis and image plane, and is also the center of image plane. O_cO_1 is the focal length of a camera. World coordinate system (O_w, X_w, Y_w, Z_w) sets up in the plane T . For a point on the plane (X, Y, Z) . Projection p on the image plane is intersection of line O_cP , which is point P and O_c connect into and the image plane.

Suppose any point p has the coordinate (x, y) in image plane coordinate system (O_1, x, y) , and has the coordinate (u, v) in image pixel coordinate system (O_0, u, v) . The two have the following relationship:

$$u = \frac{x}{dx} + u_0$$

$$v = \frac{y}{dy} + v_0 \tag{1}$$

Where (u_0, v_0) is the O_1 coordinate in the u, v coordinate system, dx, dy are the physical dimensions of each pixel in the X axis and Y axis direction. The two are expressed in homogeneous coordinates as:

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

Suppose any point P in the space has the coordinate $(X_c, Y_c, Z_c, 1)^T$ in the camera coordinate system and the coordinate $(X_w, Y_w, Z_w, 1)^T$ in the world coordinate system. The two have the following relationship:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M_1 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (3)$$

Where R is a 3×3 orthogonal unit matrix, representing the rotation matrix from world coordinate to camera coordinate; T is a 3×1 vector, representing the translation vectors from world coordinate to camera coordinate; M_1 is a 4×4 matrix.

Any point P in the space has the following relationship with its projection in the image:

$$\begin{aligned} x &= f \frac{X_c}{Z_c} \\ y &= f \frac{Y_c}{Z_c} \end{aligned} \quad (4)$$

Where (x, y) is the image coordinate of point P . (X_c, Y_c, Z_c) is the coordinate on the camera coordinate system. f is the focal of the camera. The projection relationship can be express in homogeneous coordinates as:

$$Z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (5)$$

From (2), (3), (5), the relationship between P coordinate on the world coordinate and the projection $p(u, v)$ is as follows:

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (6)$$

Where $f_x = \frac{f}{dx}$, $f_y = \frac{f}{dy}$, f_x and f_y axis are called the scale factor of u axis and v axis, also called normalized focal length; u_0, v_0 are image plane center.

Make:

$$\begin{aligned} M_1 &= \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \\ M_2 &= \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \end{aligned}$$

And (6) is

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = M_1 M_2 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (7)$$

Where $M = M_1 M_2$, M is a 3×3 matrix, called projection matrix. M_1 is camera intrinsic parameters and M_2 is camera extrinsic parameters, which can be acquired by camera calibration. Three dimensional depth information of space point means the point in Z axis coordinate in camera coordinate system, and that is Z_c .

From (7), it is hard to get the position information in figure only from above message by the methods based on single figure to position a target with fewer target's information. Therefore, the methods based on single figure mainly utilize the geometrical characteristics of target, using these characteristics' space information and the mapping relation between detected image features to get the solution.

According to the geometric characteristics of the target model, it can be divided into 3 types: points, lines and advanced geometry features.

2.1.2. Measurement Based on Point. Pose measurement based on point is also called PnP problem (Perspective-n-Point Problem). The PnP question was proposed by Fisher et al in 1981. Then, the position coordinates of the 3D space of these feature points are determined by the projection relation between the N feature points of the image and their corresponding spatial positions. At present, researchers have two main ways to solve the PnP problem: One is the direct use of geometry method to solve distances from space points to the camera optical center; The other is to attain the rotation matrix R and the shift vector T between the target coordinate system and the camera coordinate system, as shown in the (3). If we know the coordinates of a space point in an object coordinate system, we can obtain its coordinates in the camera coordinate system. The difficulty of the first solution is that the square term of the unknown equation in the set of equations is difficult to obtain an analytic solution. In the second approach, since the rotation matrix R has only 3 independent variables, it is difficult to realize the inter element constraints. Over the years, researchers have conducted in-depth studies of P3P, P4P and P5P, and have made many valuable conclusions [7, 8]. Setting point features on the target has the advantages of high positioning accuracy and high speed.

2.1.3. Measurement Based On Line. As a typical image feature, line feature has strong anti-occlusion ability, and it is easy to extract, so the use of linear features for single image object localization has attracted many scholars' interest. Liu[9,10] establishes the geometric model of localization by the geometric characteristics of its linear orthogonal method using vector and target projection plane formation and straight line features in images of the camera optical center. This method of location has special requirements for the spatial location of 3 straight lines: 3 lines cannot be parallel at the same time and now they can use photo centric coplanar, construct 3 nonlinear equations to calculate the pose parameters of space. But solving nonlinear equations is difficult, and it is prone to error. When other targets are predictable, a method for locating aircraft targets using seeker images is proposed by Yu Yong [11], which uses the main axis of aircraft and its wingspan as a linear feature. With the characteristic of the size and angle of linear images to estimate the change of target imaging attitude, the influence of high positioning accuracy is not affected by the target attitude.

Han Liwei [12] proposed the use of two intersecting lines and their intersection points as features, and estimated the orientation of the robot by using the direction of the line, and then calculated the position of the robot by the change of intersection pixels.

2.1.4. Measurement Based On Advanced Geometry Features. The location method based on advanced geometric features requires solving complex nonlinear equations. Qin Lijuan [13] uses the non-coplanar curve to carry on the spatial target localization research, through constructs the high degree equation group to carry on the computation, but the computation quantity is big and the solution is difficult. When the two space curves in the same plane, it can obtain the pose parameters of nonlinear equations in closed form. A circle is a two curve, commonly used in general, through projection transformation, and will form a circle ellipse in the image. It is a function of certain related geometric properties between the two phases, the pose relationship can be obtained by solving the relationship between camera coordinate system and body coordinate system. Zhang [14] proposed an improved algorithm of monocular vision position and orientation measurement based on circle feature. By adding laser ranger information, the judgment of two sets of solutions was solved. When the radius of the circle is unknown, only one image can be used to obtain the pose information. Zhang Guangjun

proposed the landing plane target with double circular image model. The motion and projection model of airborne camera complete UAV with land Weizi parameter estimation. The method has strong anti-noise ability and can meet the requirement of real-time.

2.2. Pose Measurement Based on Two or More Images

2.2.1. Measurement Model. Pose measurement based on two or more image changes the position of the camera to screen two or more images, and utilizes image matching feature points with projective geometric transformation relation in order to acquire the posture parameters. The model of pose measurement based on two images is shown in Fig.3. Where O_1 and O_2 is the two shooting camera optical center position. I_1 and I_2 is the two shooting image plane. P is a characteristic point in the target. p_1 and p_2 are the projection point, p_1 and p_2 also is the corresponding matching points. Supposed that \mathbf{R} is the rotating matrix that the camera move from A to B, and \mathbf{T} is the translation matrix. The coordinate of P in camera coordinate in position A is $X_{c1} (x_{c1}, y_{c1}, z_{c1})^T$, and the corresponding pixel coordinates is (u_1, v_1) , while in position B is $X_{c2} (x_{c2}, y_{c2}, z_{c2})^T$, and the corresponding is (u_2, v_2) . Target localization model based on two images is mentioned as in Figure 2.

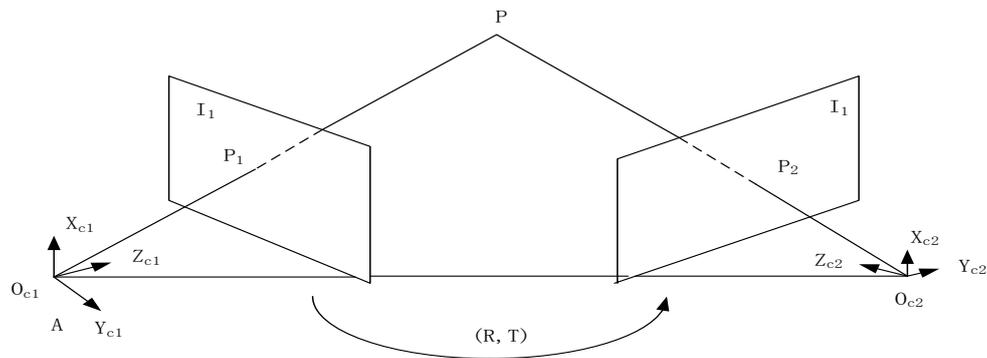


Figure 2. Target localization model based on two images model

The relationship determined by camera motion parameters between position A and B is:

$$X_{c2} = R X_{c1} + T \tag{8}$$

Make two coordinates substituted in (8):

$$\begin{bmatrix} X_{c2} \\ Y_{c2} \\ Z_{c2} \end{bmatrix} = R \begin{bmatrix} X_{c1} \\ Y_{c1} \\ Z_{c1} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{9}$$

The transformation from image pixel coordinate system to camera coordinate system can be gain from (2) and (5):

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = M_1 \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} \tag{10}$$

From (11) and (12):

$$Z_{c2}M_1^{-1} = \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = Z_{c1}RM_1^{-1} \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (11)$$

Translated into:

$$\begin{bmatrix} -RM_1^{-1} \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} & M_1^{-1} \begin{pmatrix} u_2 \\ v_2 \\ 1 \end{pmatrix} \end{bmatrix} \begin{bmatrix} Z_{c1} \\ Z_{c2} \end{bmatrix} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (12)$$

Make $C = \begin{bmatrix} -RM_1^{-1} \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} & M_1^{-1} \begin{pmatrix} u_2 \\ v_2 \\ 1 \end{pmatrix} \end{bmatrix}$, (12) can be expressed:

$$C[Z_{c1} \quad Z_{c2}]^T = T \quad (13)$$

Where M_1^{-1} , R, T can be calculated, the point coordinate in image plane (u_1, v_1) and (u_2, v_2) can be calculated by the extraction and matching of the characteristic points. The deepness Z_{c1} and Z_{c2} can be calculated by the least square method. Z_{c2} is the distance from the camera optical center at present to the target position in Z axis in the camera coordinate.

In the processing of the pose measurement based on two images. Calculating the position relationship between the camera coordinates under two angles of view is a key point. Paper estimates that relationship by essential matrix. It calculates the basic matrix by the matching points, and calculates the essential matrix with the camera internal parameter matrix, and then decomposes it to acquire the relationship, and ultimately achieves the pose measurement.

3. Conclusion

A comprehensive review on pose measurement based on monocular vision is presented in this paper. By reading relevant literature about an important visual target location based on the analysis of the current method and based on computer vision target pose measurement, it can be concluded that the measurement of pose parameters is developing in the following directions:

1. The measurement method requires simple manual operation, high degree of automation, low measurement cost, real-time results.
2. The time required for tracking measurements is measured only by takeoff and landing phases, and measurements of the whole course of motion are made.
3. Parameters that need to be measured are developed from attitude parameters to pose, position, velocity, acceleration and other parameters.

References

- [1] W. Christian, W. Stefan, R. Stefan, S. Konrad, S. Bernt, Monocular visual scene understanding: understanding multi-object traffic scenes [J], IEEE Trans. Pattern Anal. Mach. Intell. 35 (4) (2013) 882–897.
- [2] S. Reza, S. Davide, Multi-body motion estimation from monocular vehicle-mounted cameras [J], IEEE Trans. Rob. 32 (3) (2016) 638–651.
- [3] L. Changjiang, Yi Zh, 3D trajectory reconstruction from monocular vision based on prior spatial knowledge [J], IEEE Sens. J. 16 (3) (2016) 817–822.
- [4] S. Zengyu, L. Yajun, S. Jincheng, G. Lei, Study on measuring the pose of the moving object based on binocular vision [J], Eighth International Symposium on Precision Engineering Measurement and Instrumentation 8759 (2013) 1–6.
- [5] G. Hua, C. Hao, Zh. Zhaoxiang, W. Kai, Method of binocular rotation angle measurement based on machine vision, Fourth International Conference on Instrumentation and Measurement,

- Computer, Communication and Control (2014) 462–467.
- [6] Yubo Guo, Dan Zhao, Optimization design of non-coplanar target for pose measurement with monocular vision system [J], *Optik* 131 (2017) 72–78.
 - [7] Fishier M, Boilers R, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [J]. *Communication of the ACM*, 1981, 24(6): 381-393.
 - [8] Hu Z Y, Lei C, Wu F C. A short note on p4p problem [J]. *Acta Automatica Sinica*, 2001, 27(6): 770—776 (in Chinese).
 - [9] Liu Y, Huang T, Faugeras O. Determination of camera location from 2-D to 3-D line and point correspondences [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, 12(1): 28-37.
 - [10] Chen H. Pose determination from line to plane correspondences: existence solution and closed form solutions [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1993, 13(6): 530-541.
 - [11] Yu Y, Guo L. Target passive location method based on line feature [J]. *Opto-Electronic Engineering*. 2009, 36(1): 41—46 (in Chinese).
 - [12] Han L W, Xu D. A Visual Dead Reckoning Positioning Method Based on Lines and Single Feature Point for Mobile Robot [J]. *Robot*, 2008(1): 79—84 (in Chinese).
 - [13] Qin L J, Hu Y L, Wei Y Z, et al. A algorithm for vision location based on plane quadrangle [J]. *Transactions of Shenyang Ligong University*, 2009, 28(2): 66-69 (in Chinese).
 - [14] Zhang L, Xu K J, Zhao R, et al. Improvement of position and orientation measurement algorithm of monocular vision based on circle features [J]. *Journal of Hefei University of Technology (Natural Science)*, 2009, 32(11): 1669-1673 (in Chinese).