Optik 125 (2014) 4051–4056

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

A monocular pose measurement method of a translation-only one-dimensional object without scene information

Yang Shang [a,b,*], Jinbo Liu [a,b], Tangtao Xie [a,b], Linshen Yao [a]

[a] College of Aerospace Science and Engineering, National University of Defense Technology, China
[b] Hunan Key Laboratory of Videometrics and Vision Navigation, China

ARTICLE INFO

Article history:
Received 22 July 2013
Accepted 15 January 2014

Keywords:
Pose measurement
Monocular
One-dimensional object
Translation-only

ABSTRACT

A monocular method to measure the pose of a translation-only one-dimensional (1D) object containing at least two known characteristic points is presented. The position of any other point on the 1D object can be located after the pose measurement. Scene information is not needed to apply the method. If the distance between the characteristic points is unknown, there is a scale factor between the object's real positions and the measurement results. The method can also be used to measure the pose and shape parameters of a parallelogram, prism or cylinder object.

1. Introduction

One important task of computer vision is to measure an object's three-dimensional (3D) pose, including position and attitude parameters. When the measurement system comprises two or more cameras, the 3D position and attitude parameters can be obtained by intersecting the cameras at three or more characteristic points on the object [1]. If only one camera is used (monocular), the object's pose can also be measured when there are at least three characteristic points on the object with known coordinates in the object system [2,3]. Such research is known as a PNP (perspective of N points) problem [4,5]. If there are more characteristic points on the object, the 3D structure and movement can be rebuilt by using the object's sequence images captured by a single camera [6]. For a point object, that is, the object is a point or only one characteristic point can be used on the object, its 3D position can be calculated by more than two cameras or by a single moving camera [7,8]. Apart from using characteristic points, the task can also be completed according to other characteristics such as lines or the object's contour [9–11].

For a one-dimensional (1D) object consisting of characteristic points on a line, most research on 3D pose measurement focuses on camera calibration with 1D objects. In camera calibration, all or some of the camera's intrinsic parameters and extrinsic parameters, i.e., the translation and attitude parameters between the camera and the objects, are calculated. The extrinsic parameters of the camera describe the pose of the object related to the camera. Traditional methods of camera calibration are based on 2D or 3D objects with reference points distributed on different planes, or on the same plane but along different lines [12,13]. In 2004, Dr. Zhengyou Zhang first proposed a method of camera calibration with 1D objects [14]. In Zhang's method, the reference object has at least three characteristic points on a line. The distances between the characteristic points are known. When one of the characteristic points is fixed and the others are moving, the camera's intrinsic and extrinsic parameters can be calculated from at least six images of the 1D object. After that, Wu, Peng, Qi and Lv further proposed some methods of camera calibration using 1D objects with the condition that the objects move under various restrictions, i.e., the object moves on a plane [15,16], the object moves under gravity (its centroid's trajectory is a parabola) [17], or the object translates in a direction perpendicular to a horizon plane and there are horizon lines to be used in the view field [18]. If there are two or more cameras used together, the multiple cameras can be calibrated with a 1D object under general motions [19].

This study presents a monocular method of pose measurement of a translation-only 1D object. The method can be used to obtain the 3D position and attitude of the translation-only object using at least two of its known characteristic points. The position of any other point aligned on the 1D object can also be located. If the distance between any two characteristic points is unknown, there is a scale factor between the object's real positions and

* Corresponding author at: Lab. 1302, College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China.
Tel.: +86 731 84576462; fax: +86 731 84576461.
E-mail address: shangyang1977@sina.com (Y. Shang).

http://dx.doi.org/10.1016/j.ijleo.2014.01.115
0030-4026/© 2014 Published by Elsevier GmbH.

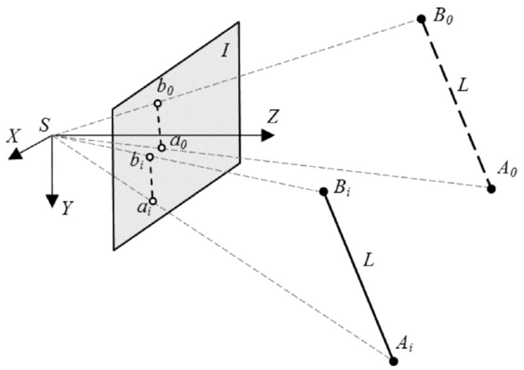

Fig. 1. Two characteristic points on a translation-only 1D object and their corresponding projections.

the measurement results. By the new method, the movement parameters of a translation-only object in a stationary camera's view field can be measured when there are two characteristic points on the object and the distance between them is known. Scene information, such as horizon plane, horizon lines, or any other geometrical restriction, is not needed for the object's pose measurement. This method can also be used to locate any other point on the 1D object and to obtain the movement and shape parameters of parallelograms, prisms and cylinders.

2. Principles of the new method

2.1. Projective relationship of the characteristic points on a translation-only 1D object

A pin-hole camera model is used in this method. The stationary camera system *S-XYZ* is the reference system. The coordinates in 3D for a characteristic point in the reference system are $\mathbf{M} = [X\ Y\ Z]^T$. The augmented matrix of $\mathbf{M}$ is $\tilde{\mathbf{M}} = [\,X \quad Y \quad Z \quad 1\,]^T$. The coordinates for the corresponding projection's image are $\mathbf{m} = [u\ v]^T$, with augmented matrix $\tilde{\mathbf{m}} = [\,u \quad v \quad 1\,]^T$. Then, the relationship between the characteristic point and its projection is as follows:

$$s\tilde{\mathbf{m}} = \mathbf{A}[\,\mathbf{R} \quad \mathbf{T}\,]\tilde{\mathbf{M}}$$
$$\text{with} \quad \mathbf{A} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{1}$$

where s is an arbitrary scale factor, the extrinsic parameters $[\mathbf{R}\ \mathbf{T}]$ are the rotation and translation, which relate the reference coordinate system to the camera coordinate system. Because the camera system in this case is the reference system, the translation vector $\mathbf{T}$ is a zero vector and the rotation matrix $\mathbf{R}$ is an identity matrix. $\mathbf{A}$ is the camera intrinsic matrix, and (u_0, v_0) are the coordinates of the principal point, α and β are the scale factors in the image u and v axes, and γ is the parameter describing the skew of the two image axes.

For example, consider the two characteristic points on a translation-only 1D object represented by the two endpoints on the 1D object shown in Fig. 1. The camera coordinate *S-XYZ* is the reference coordinate, and the image plane is *I*. The object's two characteristic points are *A* and *B*. The distance from *A* to *B* is *L*. The object is translated $m-1$ times from the initial location. At time t_i $(i = 0, 1, \ldots, m-1)$, the two characteristic points' locations are A_i and B_i, and the corresponding projections are a_i and b_i.

The coordinates of the characteristic points *A* and *B* are $\mathbf{M}_{Ai} = [X_{Ai}\ Y_{Ai}\ Z_{Ai}]^T$ and $\mathbf{M}_{Bi} = [X_{Bi}\ Y_{Bi}\ Z_{Bi}]^T$ at time t_i. The translation of the object from t_i to the initial time t_0 is $\mathbf{T}_i = [T_{Xi}\ T_{Yi}\ T_{Zi}]^T$, where $\mathbf{T}_0 = [0\ 0\ 0]^T$. Thus the following can be written:

$$\begin{cases} \mathbf{M}_{Ai} = \mathbf{M}_{A0} + \mathbf{T}_i \\ \mathbf{M}_{Bi} = \mathbf{M}_{B0} + \mathbf{T}_i \end{cases} \quad (i = 0, 1, 2, \ldots, m-1) \tag{2}$$

The image coordinates of the characteristic points' projections are $\mathbf{m}_{Ai} = [u_{Ai}\ v_{Ai}]^T$ and $\mathbf{m}_{Bi} = [u_{Bi}\ v_{Bi}]^T$ at time t_i. According to Eqs. (1) and (2), the projective relationships of the characteristic points *A* and *B* on the translation-only 1D object are obtained as follows:

$$\begin{cases} s_{Ai}\tilde{\mathbf{m}}_{Ai} = \mathbf{A}[\,\mathbf{R} \quad \mathbf{T}\,](\tilde{\mathbf{M}}_{A0} + \tilde{\mathbf{T}}_i) \\ s_{Bi}\tilde{\mathbf{m}}_{Bi} = \mathbf{A}[\,\mathbf{R} \quad \mathbf{T}\,](\tilde{\mathbf{M}}_{B0} + \tilde{\mathbf{T}}_i) \end{cases} \quad (i = 0, 1, 2, \ldots, m-1) \tag{3}$$

where $\tilde{\mathbf{m}}_{A0}$, $\tilde{\mathbf{m}}_{B0}$, $\tilde{\mathbf{m}}_{Ai}$, $\tilde{\mathbf{m}}_{Bi}$, $\tilde{\mathbf{M}}_{A0}$, $\tilde{\mathbf{M}}_{B0}$ and $\tilde{\mathbf{T}}_i$ are augment matrices of $\mathbf{m}_{A0}$, $\mathbf{m}_{B0}$, $\mathbf{m}_{Ai}$, $\mathbf{m}_{Bi}$, $\mathbf{M}_{A0}$, $\mathbf{M}_{B0}$ and $\mathbf{T}_i$. s_{A0}, s_{B0}, s_{Ai} and s_{Bi} are scale factors. $\mathbf{A}$ is the camera intrinsic matrix and $[\mathbf{R}\ \mathbf{T}]$ are the extrinsic parameters, where $\mathbf{T}$ is a zero vector and $\mathbf{R}$ is an identity matrix.

2.2. Calculation of the translation-only 1D object's pose parameters

In pose measurement problems, the camera's intrinsic parameters are known. The parameters to be calculated are $\mathbf{M}_{A0}$, $\mathbf{M}_{B0}$ and $\mathbf{T}_i$ $(i = 1, 2, \ldots, m-1)$, i.e., the locations of *A* and *B* at the initial time, t_0, and the translation of the object from t_0 to the later time, t_i. $\mathbf{M}_{A0}$ and $\mathbf{M}_{B0}$ describe the initial position and attitude of the 1D object, and $\mathbf{T}_i$ describes the purely translational movement of the object from the initial time to time t_i. The attitude of the object remains unchanged during translation-only movement.

The scale factor is eliminated from Eq. (3) and the equations are expressed in variable form as follows:

$$\begin{cases} u_{Ai} - u_0 = \alpha\dfrac{X_{A0} + T_{Xi}}{Z_{A0} + T_{Zi}} + \gamma\dfrac{Y_{A0} + T_{Yi}}{Z_{A0} + T_{Zi}} \\ v_{Ai} - v_0 = \beta\dfrac{Y_{A0} + T_{Yi}}{Z_{A0} + T_{Zi}} \\ u_{Bi} - u_0 = \alpha\dfrac{X_{B0} + T_{Xi}}{Z_{B0} + T_{Zi}} + \gamma\dfrac{Y_{B0} + T_{Yi}}{Z_{B0} + T_{Zi}} \\ v_{Bi} - v_0 = \beta\dfrac{Y_{B0} + T_{Yi}}{Z_{B0} + T_{Zi}} \end{cases} \quad (i = 0, 1, 2, \ldots, m-1) \tag{4}$$

Because the object is in front of the camera, the *Z* coordinates of the characteristic points are positive. Some intermediate variables can be defined as follows:

$$\begin{cases} g_0 = \dfrac{X_{A0}}{Z_{A0}}, g_1 = \dfrac{Y_{A0}}{Z_{A0}}, g_2 = \dfrac{X_{B0}}{Z_{A0}}, g_3 = \dfrac{Y_{B0}}{Z_{A0}}, g_4 = \dfrac{Z_{B0}}{Z_{A0}} \\ g_{5,i} = \dfrac{T_{Xi}}{Z_{A0}}, g_{6,i} = \dfrac{T_{Yi}}{Z_{A0}}, g_{7,i} = \dfrac{T_{Zi}}{Z_{A0}} \quad (i = 1, 2, \ldots, m-1) \end{cases} \tag{5}$$

Eq. (4) is transformed and the resulting equations are divided by Z_{A0}, to obtain the following:

$$\begin{cases} (u_{A0} - u_0) = \alpha g_0 + \gamma g_1 \\ (v_{A0} - v_0) = \beta g_1 \\ (u_{B0} - u_0)g_4 = \alpha g_2 + \gamma g_3 \\ (v_{B0} - v_0)g_4 = \beta g_3 \\ (u_{Ai} - u_0)(1 + g_{7,i}) = \alpha(g_0 + g_{5,i}) + \gamma(g_1 + g_{6,i}) \\ (v_{Ai} - v_0)(1 + g_{7,i}) = \beta(g_1 + g_{6,i}) \\ (u_{Bi} - u_0)(g_4 + g_{7,i}) = \alpha(g_2 + g_{5,i}) + \gamma(g_3 + g_{6,i}) \\ (v_{Bi} - v_0)(g_4 + g_{7,i}) = \beta(g_3 + g_{6,i}) \end{cases} \tag{6}$$