Optik 126 (2015) 1520–1526

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

A two-step pose estimation method based on four non-coplanar points

Zimiao Zhang [a,*], Bin Liu [b], Yongxiang Jiang [a]

[a] Tianjin University of Technology and Education, Tianjin Key Laboratory of High Speed Cutting & Precision Machining, 300222 Tianjin, China
[b] Tianjin University of Technology, School of Electrical Engineering, 300384 Tianjin, China

ARTICLE INFO

Article history:
Received 9 March 2014
Accepted 23 April 2015

Keywords:
Pose estimation
Scaled orthographic projection
Geometrical constraints
Perspective projection

ABSTRACT

The existing approaches to a solution of object pose fall into two distinct categories: analytical solutions and iterative solutions. In this paper, a two-step pose estimation method based on four non-coplanar points is studied by combining the advantages of the above two methods. In the first step, the scaled orthographic projection model is used to approximate the perspective projection model to construct the analytical solving model of points coordinates in the camera coordinate system. In the second step, an iterative method based on the geometrical constraints formed by the points is introduced, and the points coordinates obtained in the first step are passed to the current step being used as initial values of the iterative process. The proposed method is then compared with two other existing methods. Experimental results demonstrate that the proposed method has the advantage of lower computational cost. The accuracy, which is the main concerns during the real-time applications, is also improved.

1. Introduction

With the development of modern industrial technology, quickly and accurately determine the position and orientation between objects is becoming more and more important. For two objects in space, whether they are both moving, or one is still and the other is moving, one of them is selected as the observer, and the other as the target. The measurement of the relationship between them is the pose estimation problems. It can be widely applied in the fields of robot navigation [1], coordinates vision measurement system using light pen [2], helmet mounted sight [3,4], wheel alignment system [5], aerospace [6–8], so the research in this area is more active in recent years.

Pose estimation based on points is also known as PnP (Perspective n-Points Problem), in 1981 first proposed by Fischler and Bolles [9]. Given the relative positions of n control points, the angles between connection lines of control point and optical center, solve the distance from each control point to the optical center. The problem has been further studied by other researchers [10,11].

The existing approaches to a solution of object pose based on points fall into two distinct categories: analytical solutions and iterative solutions.

Analytical solutions: Hu and Wu [12] proved that 4 feature points can achieve a linear solution, but may have more than one solution. Tang et al. [13] presented a linear algorithm in the condition of five feature points. In [14] a non-linear pose estimation model is established on the basis of distance factor. With algebraic transformation method, the nonlinear mathematic mode is transformed into a linear non-homogeneous equation system. Ansar and Daniilidis [15] presented a general framework to directly recover the rotation and translation. Duan [16] introduced a new affine invariant of trapezium. Given the lengths of two parallel sides of a trapezium, pose estimation and plane measurement can be realized in a very simple way from the projection of trapezium.

Iterative solutions: DeMenthon et al. [17,18] proposed POSIT algorithm, which gets the initial value of solution using the scaled or orthographic model to approximate the perspective projection model. Chen et al. [19] solved the rotation and translation matrix using least squares method according to the orthonormal constraints which uses lots of points. Peng et al. [20] achieved the object pose non-linearly on basis of five feature points using least squares approach. Wildey et al. [21] proposed a positioning method based on three points, but this approach had a smaller measurement range (0°–5°). Liu and Wong [22] proposed a positioning method according to the corresponding geometrical contraints of four feature points. Zhang et al. [23] proposed an efficient solution for vision-based pose determination of a parallel manipulator.

Both analytical solutions and iterative solutions have their advantages and disadvantages. Compared with iterative solutions,

* Corresponding author. Tel.: +86 2228189031.
E-mail address: zzm19850126@aliyun.com (Z. Zhang).

http://dx.doi.org/10.1016/j.ijleo.2015.04.039

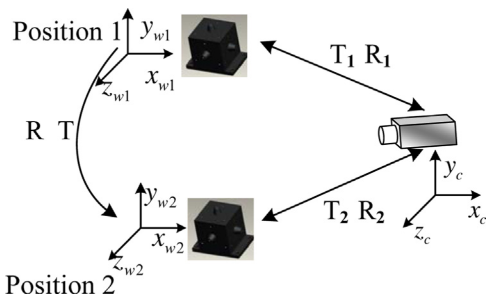

Fig. 1. Pose measurement principle.

analytical solutions generally take less computation time for the same accuracy. But on the other hand, the analytical solutions are extremely susceptible to noise because of the orthonormal constraints associated with the rotation matrix that are not taken into account. Whenever the number of points is larger than 4, the analytical solutions are not efficient. Iterative solutions are normally solved with some variation on gradient descent or Gauss–Newton methods and they have drawbacks. (1) Non-linear optimization needs a good initial estimate of the true solution. (2) They are more time consuming than analytical solutions. Therefore, such approaches cannot be used in tasks that require high speed performance.

Based on the discussion above, in this paper an efficient two-step method for pose estimation based on four non-coplanar is proposed. Firstly, the coordinates of four non-coplanar points in the camera coordinate system are calculated on the basis of geometrical constraints formed by the points and the corresponding image coordinates of points. During the process of solving, the scaled orthographic projection model is used to approximate the perspective projection model. In this way the coordinates of points in the camera coordinate system is solved analytically and the coordinates of points obtained are used as initial values of the iterative process to ensure the accuracy and convergence rate of non-linear algorithm.

The advantages of the two-step method are as follows. (1) The situation that the linear method is extremely susceptible to noise is improved. (2) The iterative process is given a preferable initial value.

The rest of the paper is organized as follows: Section 2 gives the principle of pose estimation. Section 3 proposes a method to get the initial value of object pose through scaled orthographic projection model. Section 4 proposes a method to obtain the exact pose of object through the geometrical constraints formed by the points. Section 5 provides some experiments to examine the two-step method from calculation precision and rapidity. Section 6 gives the conclusion.

2. The measurement principle of object pose

Pose measurement principle based on visual image is shown in Fig. 1. The intrinsic parameters of the camera are obtained by camera calibration method presented in [24]. The points on the object are captured by CCD camera and the corresponding image coordinates are obtained.

The target with four non-coplanar feature points is shown in Fig. 2, and their coordinates in the world coordinate system are known. The target is a cube with four non-coplanar feature points on each side of it except the bottom side, which is used to fix the target on the rotation and translation stage.

Pose of the object is represented with a rotation and translation matrix (**R**, **T**) in the corresponding coordinate system. The pose of

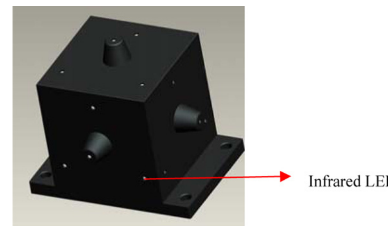

Fig. 2. Target with four non-coplanar points.

object in position 1 relative to the camera is $M_1 = R_1 \cdot T_1$, which can be expressed as the formula (1).

$$\begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = R_1 \begin{bmatrix} x_{c1} \\ y_{c1} \\ z_{c1} \end{bmatrix} + T_1 \tag{1}$$

The pose of object in position 2 relative to the camera is $M_2 = R_2 \cdot T_2$, which can be expressed as the formula (2).

$$\begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = R_2 \begin{bmatrix} x_{c2} \\ y_{c2} \\ z_{c2} \end{bmatrix} + T_2 \tag{2}$$

According to formulas (1) and (2), the pose of object can be expressed as the formula (3).

$$\begin{cases} R = R_2^{-1} \cdot R_1 \\ T = R_2^{-1}(T_1 - T_2) \end{cases} \tag{3}$$

It can be seen from the discussion above that coordinates of points in the camera coordinate system need to be obtained firstly to get the rotation and translation matrix that represent the pose of object. The process of solving point coordinates is introduced in Sections 3 and 4.

3. The analytical solving model of object pose

The scaled orthographic projection model (also known as the weak perspective projection model), which completely ignores the depth information, can be used to approximate the perspective projection model. This approximation can be regarded as synthesis of two successive projections. Firstly, the whole object is projected to a plane that is parallel to the image plane. Then the object figure on the plane mentioned above will be projected to the image plane of camera according to perspective projection model.

In this section, as shown in Fig. 3, according to the scaled orthographic projection, assuming that the four points have the same depth, a plane through P_0 parallel to the image plane is drawn, the points (from 1 to 3) are projected to this plane (represented with R_i) in direction parallel to the optical axis. The distance from this plane to O_c is at a Z_0. Then R_i is projected to the image plane by perspective projection.

The coordinates of points in the camera coordinate system are $P_i(X_i, Y_i, Z_i)$ and the coordinates of P_i on the image plane in the scaled orthographic projection model could be expressed as formula (4), in which $w = f/Z_0$ is scale factor.

$$\begin{cases} x_i' = fX_i/Z_0 \\ y_i' = fY_i/Z_0 \end{cases} \tag{4}$$

Three points are enough to solve the coordinates of points with the scaled orthographic projection and P_0, P_1, P_2 are chosen. Fig. 3 is simplified with preserving P_0, P_1, P_2 only. I_1' is the intersection point of the vertical line from I_1 to the plane through No. 0 point