Optik 125 (2014) 6106–6112

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Extended gravitational pose estimation

Peng Chen*, Guang-Da Hu, Jiarui Cui

School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

ARTICLE INFO

Article history:
Received 1 November 2013
Accepted 3 June 2014

Keywords:
Pose estimation
Correspondenceless
Gravitational field

ABSTRACT

The model-to-image registration problem is a problem of determining the position and orientation (the pose) of a three-dimensional object with respect to a camera coordinate system. When there is no additional information available to constrain the pose of the object and to constrain the correspondence of object features to image features, the problem is also known as simultaneous pose and correspondence problem, or correspondenceless pose estimation problem. In this paper, we present a new algorithm, called extended gravitational pose estimation (EGPE), for determining the pose and correspondence simultaneously. The algorithm is based on gravitational pose estimation (GPE) algorithm. In our algorithm, the original GPE has been revised to deal with the problem with false image points. For problems with both occluded object points and false image points, we firstly applied single-link agglomerative clustering algorithm to pick out occluded object points when a local minimum has been found, then the revised GPE is applied again on the clustering result to update rotation and translation of the object model. EGPE has been verified on both synthetic images and real images. Empirical results show that EGPE is faster, more stable and reliable than most current algorithms, and can be used in real applications.

1. Introduction

Vision-based pose estimation problem is to determine position and orientation of a camera and a target with a set of n feature points expressed in the object coordinates and its 2D projection expressed in the camera coordinates. It is one of the key problems in object recognition, visual navigation, robot localization, augmented reality and other areas.

When the correspondence between 3D object feature points and 2D image feature points are given, the problem is also known as perspective-n-point (PnP) problem. Existing methods to solve the problem can be divided into three categories: non-iterative, iterative and globally optimal algorithms. The non-iterative algorithms apply linear methods to obtain algebraic solutions. Most research is focused on P3P, P4P and P5P problems [1,2], since PnP problem is actually a classical Direct Linear Transformation (DLT) problem and can be solved linearly when $n > 5$ [3]. The most popular algorithms to handle arbitrary value of n are proposed in [4,5]. Especially, the EPnP algorithm proposed in [6] is believed to be faster and more accurate than other non-iterative algorithms. As to iterative algorithms, the classical methods are to formulate the pose estimation as a nonlinear least-square problem with the constraint that the rotation matrix is orthogonal. The problem can be solved by using non-linear optimization algorithms, most typically, Levenberg–Marquadt or Gaussion–Newton method [7–10]. A widely used iterative algorithm is orthogonal iterative (OI) algorithm proposed by Lu et al. [11]. The algorithm reformulated a new objective function to minimize the object-space collinearity error. Compared with other iterative algorithms, OI algorithm performs higher accuracy, speed and noise-resistance. Globally optimal algorithms are recently developed algorithms based on L_∞-norm minimization method. Since the object-space collinearity error can be expressed as a quasi-convex function, techniques of linear programming or second-order cone programming can be applied. Typical algorithms are proposed in [12–14].

However, the input data are often noisy in a real application, the extracted features can be missing or false, and their position may be inaccurate due to poor image quality, bad light conditions, partial occlusions and/or precision of acquisition and feature-extraction process. In those cases, the problem actually consists of two subproblems: pose estimation and correspondence determination. When there is no additional information available with which to constrain the pose of the object and to constrain the correspondence of object features to image features, the problem is known as the simultaneous pose and correspondence problem,

* Corresponding author. Permanent address: School of Automation and Electrical Engineering, University of Science and Technology Beijing, No. 30 Xueyuan Road, Haidian District, Beijing 100083, China. Tel.: +86 13522300113.
E-mail address: holmes83@163.com (P. Chen).

http://dx.doi.org/10.1016/j.ijleo.2014.06.109
0030-4026/

or correspondenceless pose estimation problem. The problem is difficult because it requires solution of two coupled problems, each easy to solve only if the other has been solved first. The classic approach to solving these coupled problems is the hypothesize-and-test approach, of which the best known example is the RANSAC algorithm [15]. The RANSAC algorithm can achieve a high probability of success, but at a very heavy computation cost. A more effective algorithms of this type is Blind PnP, which is developed in [16] using a Gaussian mixture model. However, it needs some prior information on the object pose. A genetic algorithm based pose estimation algorithm with correspondences determination called EvoPose is proposed in [17]. Inspired by EvoPose, an algorithm by using differential evolution algorithm, called DePose, is proposed in [18]. The problem of the both algorithms is that poor local minima may cause the search to converge to false solutions, especially when there are missing or false image points. By integrating an iterative pose estimation technique called POSIT [19], and an iterative correspondence assignment technique called SoftAssign [20] into a single iteration loop, David et al. proposed an algorithm called SoftPOSIT [21]. It is arguably one of the best algorithms for the simultaneous pose and correspondence problem. However, SoftPOSIT do not ensure to find a pose when the initial guess is poor.

Gravitational pose estimation (GPE) algorithm, which is inspired by classical mechanics, is proposed in [22]. The algorithm creates a simulated gravitational field from the image and lets the object model to move and rotate in that force field, starting from an initial pose. Unlike SoftPOSIT, GPE algorithm is more robust, consistent and fast even when starting from a bad initial pose. Because SoftPOSIT can find pose with great precision when it is able to converge, GPE and SoftPOSIT are integrated together in [23] to improve the performance of the correspondenceless pose estimation algorithm. However, the usage of GPE algorithm is limited by the assumption that there are no false image feature points.

In this paper, we proposed an algorithm, called extended gravitational pose estimation (EGPE). The algorithm can solve correspondenceless pose estimation problem even when there are occluded object points or false image points. We firstly revised the original GPE algorithm to make it capable of dealing with correspondenceless pose estimation problem with false image feature points. Then we further improved the revised GPE algorithm, in order to deal with the case when there are both occluded object points and false image points. Experimental results show that our algorithm is faster, more stable and reliable than most current correspondenceless pose estimation algorithm.

2. Problem formulation

Suppose the object model can be described by a set of feature points $\boldsymbol{P}_k$ (k = 1, . . ., L), the coordinate of $\boldsymbol{P}_k$ in the object coordinate system, $OX^wY^wZ^w$, is $\boldsymbol{P}_k{}^w = [X_k{}^w, Y_k{}^w, Z_k{}^w]^T$, while the coordinate of $\boldsymbol{P}_k$ in the image coordinate system, O_Ixy, is $\boldsymbol{p}_i = [x_i, y_i]^T$ (i = 1, . . ., N), according to perspective projection, the relationship between $\boldsymbol{P}_k{}^w$ and $\boldsymbol{p}_i$ is

$$\tilde{\boldsymbol{p}}_i = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} [R \quad t]\tilde{\boldsymbol{P}}_k^w, \tag{1}$$

where $\tilde{\boldsymbol{p}}_i$ and $\tilde{\boldsymbol{P}}_k^w$ are the corresponding homogeneous coordinate of $\boldsymbol{p}_i$ and $\boldsymbol{P}_k^w$ respectively, f is the focal length. $\boldsymbol{R} = [\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{R}_3]$ is the rotation matrix. If $\boldsymbol{u}_i$, $\boldsymbol{u}_j$, $\boldsymbol{u}_k$ represents the three unit direction vectors of axes X^w, Y^w, Z^w of the object coordinate system, then $\boldsymbol{R}_1$, $\boldsymbol{R}_2$, $\boldsymbol{R}_3$, are respectively the coordinates of $\boldsymbol{u}_i$, $\boldsymbol{u}_j$, $\boldsymbol{u}_k$ expressed in the camera coordinate system. $\boldsymbol{t}$ is the translation vector. The relationship between the object coordinate system, the camera coordinate system and the image coordinate system is shown in Fig. 1.

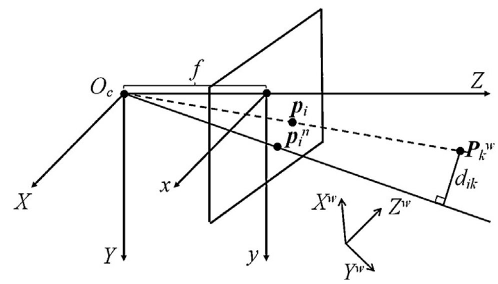

Fig. 1. Perspective projection model.

Ideally, $\boldsymbol{P}_k^w$ should align with the line of sight (LOS) $\bar{O_c\boldsymbol{p}_i}$; however, because of noise and inaccurate feature extraction, the position of $\boldsymbol{P}_k$ on the image plan is usually $\boldsymbol{p}_i^n = [\boldsymbol{x}_i^n, \boldsymbol{y}_i^n]^T$, but not $\boldsymbol{p}_i$. Therefore, the distance between $\boldsymbol{P}_k^w$ and LOS $\bar{O_c\boldsymbol{p}_i^n}$ can be represented as [11]

$$d_{ik}^2 = \left\| (\boldsymbol{I} - \boldsymbol{V}_i)\left(\boldsymbol{R}\boldsymbol{P}_k^w + \boldsymbol{t}\right) \right\|^2, \tag{2}$$

where $\boldsymbol{P}_k^c = \boldsymbol{R}\boldsymbol{P}_k^w + \boldsymbol{t}$, is the coordinate of $\boldsymbol{P}_k$ in the camera coordinate system, $\boldsymbol{V}_i$ is the LOS projection matrix, can be expressed as

$$\boldsymbol{V}_i = \frac{\boldsymbol{v}_i \boldsymbol{v}_i^T}{\boldsymbol{v}_i^T \boldsymbol{v}_i}, \quad \boldsymbol{v}_i = [x_i^n, y_i^n, f]^T. \tag{3}$$

According to [18], when all the object points and image points have their own corresponding points, the simultaneous pose and correspondence problem can be formulated as a minimization of the objective function

$$E = \frac{1}{L}\sum_{k=1}^{L} \min_i \left\| (\boldsymbol{I} - \boldsymbol{V}_i)\left(\boldsymbol{R}\boldsymbol{P}_k^w + \boldsymbol{t}\right) \right\|^2. \tag{4}$$

To apply Eq. (4) to more complicated cases, such as there are false image points or occlusions of object points, Eq. (4) is modified into the following form

$$E = \frac{1}{L}\sum_{i=1}^{N}\sum_{k=1}^{L} m_{ik} \left\| (\boldsymbol{I} - \boldsymbol{V}_i)\left(\boldsymbol{R}\boldsymbol{P}_k^w + \boldsymbol{t}\right) \right\|^2. \tag{5}$$

where m_{ik} is a weight, equals to 0 or 1 for its corresponding square distance $d_{ik}{}^2$. Given a set of object points $\boldsymbol{P}_k^w$, k = 1, . . ., L, and a set of image points $\boldsymbol{p}_i$, i = 1, . . ., N, the square distance $d_{ik}{}^2$ can be got from Eq. (2), so that we can get a distance matrix $\boldsymbol{D} = [d_{ik}{}^2]_{N\times L}$. To find an zero-one assignment matrix $\boldsymbol{M} = [m_{ik}]_{N\times L}$, the following steps can be applied:

Step 1: Find an ordered pair (i, k) so that $d_{ik}{}^2$ is the minimum entry of matrix $\boldsymbol{D}$.

Step 2: Set m_{ik} = 1, meanwhile assign the ith row and kth column of matrix $\boldsymbol{D}$ to an extremely large constant C.

Step 3: If all the entries of $\boldsymbol{D}$ equal to C, then return matrix $\boldsymbol{M}$, otherwise go to Step 1.

Therefore, if all the entries of the kth column of $\boldsymbol{M}$ equal 0, then the corresponding image feature of $\boldsymbol{P}_k$ is missing, so that $\boldsymbol{P}_k$ is probably occluded. If all the entries of the ith row of $\boldsymbol{M}$ equal 0, then image point $\boldsymbol{p}_i$ does not match any object feature, so that $\boldsymbol{p}_i$ is likely a false image feature point.

3. The algorithm

3.1. The revised gravitational pose estimation

To minimize Eq. (5), $d_{ik}{}^2$ should be minimized, which means that the object point $\boldsymbol{P}_k^w$ should be as close as possible to the LOS $\bar{O_c\boldsymbol{p}_i^n}$.