CAD Model-based 3D Object Pose Estimation using an Edge-Based Nonlinear Model Fitting Algorithm

Chi-Yi Tsai*, Wei-Yi Wang, Chi-Hung Huang, and Bo-Ren Shih

Department of Electrical Engineering, Tamkang University
No. 151, Ying-zhuan Road, Tamsui District
New Taipei City 25137, Taiwan R.O.C.
*Corresponding Author: chiyi tsai@mail.tku.edu.tw

Abstract

This paper addresses the design of a model-based 3D object pose estimation algorithm, which is one of the major techniques to develop a robust robotic vision system using a monocular camera. The proposed system first extracts line features of a captured image by using edge detection and Hough transform techniques. Given a CAD model of the object-of-interest, the 6-DOF pose of the object can then be estimated via a novel edge-based nonlinear model fitting algorithm, which is a nonlinear optimization process for estimating the optimal object pose based on an edge-based distance metric. Experimental results validate the performance of the proposed system.

Keywords: CAD model-based tracking, object pose estimation, model fitting, nonlinear optimization.

1. Introduction

In recent years, the application of computer vision has become more and more popular in robotic studies. For example, in industrial applications, a robot manipulator needs to know pose information of an object-of-interest (OOI) in order to perform pick-and-place tasks. In this case, computer vision techniques provide an efficient solution to detect and track the OOI for the robot to complete the task automatically.

Estimating pose information of an OOI is one of the most important grasping tasks in robot manipulation, but most visual sensing systems only provide two-dimensional (2D) image information, which is difficult to complete the task of object pose estimation in 3D space. Therefore, how to estimate the 3D pose information of an OOI from the 2D image of the object is still an important issue in the

development of a visual sensing system. To achieve this purpose, this paper presents a novel model-based 3D object pose estimation algorithm, which employs a priori-known 3D CAD model of an OOI to estimate its 6-DOF pose with respect to (w.r.t.) the camera frame of a vision system from an observed object image. This feature allows increasing the applicability of a robotic vision system.

2. Related Work

In a robotic vision system, the existing object pose estimation techniques can be roughly divided into two categories: feature-based and model-based tracing methods. The feature-based tracking methods use basic geometric features (such as points, lines, and circles, etc.), boundary contours ^(1, 2), and regions of interest ⁽³⁾ of the object to detect and track the OOI appeared in the image plane. By contrast, the model-based tracking methods use a priori-constructed 3D CAD model of the object to assist the tasks of object detection and tracking ⁽⁴⁾. Compared with the feature-based tracking approaches, the model-based ones usually provide more robust tracking results, even the OOI is suffered from the partial occlusion problem ⁽⁵⁾ in a complex work environment.

To achieve 6-DOF pose tracking, an efficient method is to use an artificial mark attached to the OOI to simplify the design of object detection and tracking processes ⁽⁶⁾. This method provides a simple and robust way to resolve the 6-DOF pose estimation problem, but it cannot work without an additional marker on the OOI. Therefore, many researchers have focused on the development of natural feature-based 3D pose tracking methods, which can be divided into keypoint-based and edge-based pose tracking methods ⁽⁷⁾. Keypoints are image features that can be

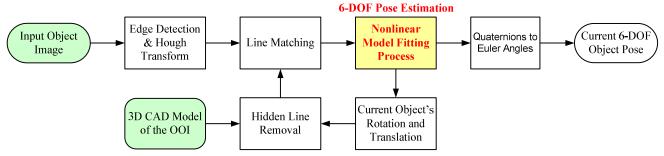


Fig. 1. Flowchart of the proposed 6-DOF pose tracking system.

repeatedly detected in images robust to translation, scaling, and rotation transformations. In keypoint-based approachs, the keypoints are first extracted from the incoming images to detect the OOI. The 6-DOF pose of the OOI can then be detected by a keypoint classifier, which was trained from a large training set with different view ports of the OOI. By doing so, the pose tracking problem is simplified as a keypoint recognition problem and can be efficiently resolved using random trees ⁽⁸⁾ for fast nearest neighbor keypoint matching.

On the other hand, edge-based pose tracking methods usually use a 3D CAD model to estimate the 6-DOF pose of the OOI detected in the image. For example, in reference ⁽⁹⁾, Choi and Christensen proposed a RAPiD (10) style tracking system, which consists of a global and a local pose estimation unit. In the global pose estimation, a keypoint-based object recognition method was employed to detect the OOI in the image and decide its initial pose information w.r.t. current camera frame. In the local pose estimation, an edge-based pose tracking algorithm was developed to estimate the 6-DOF object pose w.r.t. the camera via back-projection optimization given an initial pose and a 3D CAD model of the OOI. The proposed system provides an automatic object detection, recognition, and pose tracking solution; however, only polygon mesh models can be used as the object CAD model in this system.

In this study, a novel edge-based nonlinear model fitting algorithm is proposed to implement an image-based 6-DOF pose tracking system using a monocular camera. The proposed method is a nonlinear optimization process based on an edge-based distance metric, which can work with 3D polygon edge and polygon mesh models. Experimental results demonstrate the performance of the proposed system.

In the remainder of the paper, Section 3 introduces the proposed CAD model-based 3D object pose estimation algorithm. Section 4 presents the experimental results to

evaluate the effectiveness and efficiency of the proposed 6-DOF pose tracking method. Section 5 concludes the contributions of this paper.

3. Proposed Method

Figure 1 shows the flowchart of the proposed pose tracking system. Suppose that the OOI in the image has been detected by an object detection and recognition module. Then, the line features of the OOI are extracted from the input object image by using edge detection and Hough transform techniques. In this work, we use Canny edge detector combined with straight line Hough transform to extract line features of the OOI detected in the image. Given a CAD model of an OOI, a line matching process is employed to find the nearest edges between the detected line features and the projection of the CAD model without hidden lines. Next, the 6-DOF pose of the OOI can be estimated through a nonlinear model fitting process, which is the main subject of this paper.

After matching the detected line features of the object and the projected lines of the CAD model, the proposed nonlinear model fitting process aims to estimate the 6-DOF pose of the object by fitting the projected boundaries of the CAD model to the detected Hough lines of the object. Let **R** and **T** denote a quaternion rotation matrix (11) and a translation vector such that

$$\mathbf{R}(w_q, x_q, y_q, z_q) = \begin{bmatrix} 1 - 2y_q^2 - 2z_q^2 & 2x_q y_q + 2w_q z_q & 2x_q z_q - 2w_q y_q \\ 2x_q y_q - 2w_q z_q & 1 - 2x_q^2 - 2z_q^2 & 2y_q z_q + 2w_q x_q \\ 2x_q z_q + 2w_q y_q & 2y_q z_q - 2w_q x_q & 1 - 2x_q^2 - 2y_q^2 \end{bmatrix},$$

$$\mathbf{T}(t_x, t_y, t_z) = \begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T,$$

where (w_q, x_q, y_q, z_q) are the four components of a quaternion, (t_x, t_y, t_z) are the three components of a translation vector. Suppose that the camera intrinsic matrix **K** is computed a priori via a camera calibration process ⁽¹²⁾.

Then, the projection of a CAD model point $\mathbf{X}_c = [x_c, y_c, z_c]^T$ onto an image coordinate $\mathbf{p}_i = [x_i, y_i]^T$ with respect to an object pose $\mathbf{P} = (w_q, x_q, y_q, z_q, t_x, t_y, t_z)$ in camera coordinates can be computed by

$$\widetilde{\mathbf{p}}_{i}(\mathbf{P}) = s\mathbf{K}[\mathbf{R}(\mathbf{P}) \quad \mathbf{T}(\mathbf{P})]\widetilde{\mathbf{X}}_{c},$$
 (1)

where s is a scale factor, $\widetilde{\mathbf{p}}_i$ and $\widetilde{\mathbf{X}}_c$ are the homogeneous coordinates of \mathbf{p}_i and \mathbf{X}_c , respectively.

Now we define an edge-based distance metric for the following optimization process. Suppose that a 2D Hough line located on one of the boundaries of the object is parameterized by $\mathbf{h}=(a, b, c)$. Then, the distance between the projection of a CAD model control point (termed as a projected control point) and the 2D Hough line is given by

$$d(\mathbf{h}, \mathbf{p}_i) = \frac{|ax_i + by_i + c|}{\sqrt{a^2 + b^2}},$$
 (2)

where d is the distance between a projected control point \mathbf{p}_i and the corresponding Hough line \mathbf{h} . Suppose that there are m Hough lines, each of them having n_j corresponding projected control points. Then, a nonlinear cost function to evaluate the error distance between the projection of the CAD model and the boundaries of the object w.r.t. an object pose in the camera frame is defined as

$$f(\mathbf{P}) = \sum_{j=1}^{m} \sum_{k=1}^{n_j} d(\mathbf{h}_j, \mathbf{p}_i^{jk}(\mathbf{P})), \qquad (3)$$

where **P** is the object pose in camera coordinates defined previously, \mathbf{p}_{i}^{jk} denotes the *k*-th projected control point corresponding to *j*-th Hough line. Finally, the optimal object pose can be estimated by minimizing the cost function $f(\mathbf{P})$ such that

$$\hat{\mathbf{P}} = \arg\min_{\mathbf{P}} f(\mathbf{P}),\tag{4}$$

where $\hat{\mathbf{P}}$ is the optimal object pose. Note that the initial value of \mathbf{P} is set as (1, 0, 0, 0, 0, 0, 0) in our implementation. The nonlinear model fitting algorithm described above is validated in the next section.

4. Experimental Results

Figure 2 shows the 3D polygon CAD model used in the experiments, which is a rectangular cube model containing twelve line features. A rectangular box was used as an OOI in the experiments. Figure 3 presents the

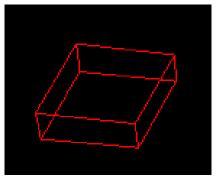


Fig. 2. The 3D polygon CAD model used in the experiments.

experimental results of the proposed model-based 6-DOF pose tracking algorithm. Figure 3(a) shows the OOI used in the pose tracking experiments, and Figs. 3(b)-3(d) shows the pose tracking results obtained from the proposed algorithm. It is clear that the proposed algorithm successfully fits the boundaries of the projected CAD model to the boundaries of the detected OOI in the incoming images. Table 1 records the 6-DOF pose tracking results of the detected OOI shown in Fig. 3. From Table 1, one can see that the yaw angle of the detected OOI is changed from -0.1515° to -14.8174° as the OOI rotates left along y-axis. This implies that the proposed tracking algorithm successfully estimates the motion trajectory of the detected OOI in the camera frame. Therefore, the above experimental results validate the tracking performance of the proposed method.

5. Conclusions

A novel CAD model-based 6-DOF pose tracking algorithm is presented in this paper. The proposed algorithm is a nonlinear model fitting process based on a new edge-based distance metric. Experimental results validate the tracking performance of the proposed method.

Acknowledgment

This work was supported by the National Science Council of Taiwan, ROC under grant NSC 103-2221-E-032-068 and 103-2632-E-032-001-MY3.

References

 B. Bascle, P. Bouthemy, R. Deriche and F. Meyer: "Tracking Complex Primitives in an Image Sequence", IEEE IRISA/INRIA, Campus Universitaire de

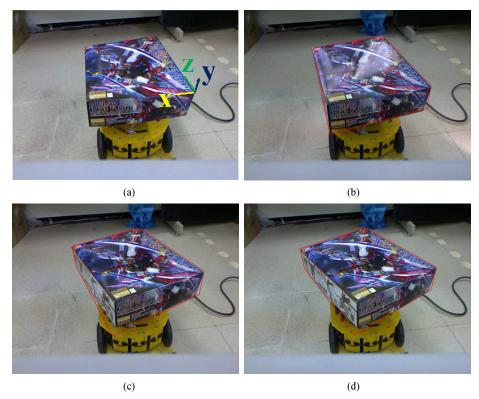


Fig. 3. Experimental results of the proposed model-based 6-DOF pose tracking algorithm.

Table 1. 6-DOF Pose Tracking Results.

Fig. 3	(b)	(c)	(d)
Pitch (degree)	-0.9762	-0.4792	-1.4983
Yaw (degree)	-0.1515	-9.0573	-14.8174
Roll (degree)	4.4816	10.0169	11.0560
Tx (m)	0.0016	0.1100	0.1803
Ty (m)	0.0083	0.0023	-0.0032
Tz (m)	0.0063	-0.0117	0.0009

Beaulieu, pp. 426-431, 1994.

- (2) M. Isard and A. Blake: "Contour Tracking by Stochastic Propagation of Conditional Density", In Proc. European Conf. Computer Vision, pp. 345-356, 1996.
- (3) G. Hager and K. Toyama: "The XVision system: A General-purpose Substrate for Portable Real-time Vision Applications", Yale University Department of Computer Science, 1995.
- (4) H. Kollning and H. H. Nagel: "3D Pose Estimation by Fitting Image Gradients Directly to Polyhedral Models", IEEE International Conference on Computer Vision, pp. 569-574, 1995.
- (5) T. Zhang, K. Jia, C. Xu, Y. Ma, and N. Ahuja: "Partial Occlusion Handling for Visual Tracking via Robust Part Matching", IEEE International Conference on Computer Vision and Pattern Recognition, pp.

1258-1265, 2014.

- (6) R. Katsuki, J. Ota, T. Mizuta, T. Kito, T. Arai, T. Ueyama, and T. Nishiyama: "Design of an Artificial Mark to Determine 3D Pose by Monocular Vision", IEEE International Conference on Robotics and Automation, pp. 995-1000, 2003.
- (7) V. Lepetit, P. Lagger and P. Fua: "Randomized Trees for Real-time Keypoint Recognition", IEEE International Conference on Computer Vision and Pattern Recognition, pp. 775-781, 2005.
- (8) Y. Amit and D. Geman: "Shape Quantization and Recognition with Ramdomized Trees", Neural Computation, Vol. 9, No. 7, pp. 1545-1588, 1997.
- (9) C. Choi and H. I. Christensen: "Real-time 3D Model-based Tracking Using Edge and Keypoint Features for Robotic Manipulation", IEEE International Conference on Robotics and Automation, pp. 4048-4055, 2010.
- (10) C. Harris: "Tracking with Rigid Objects", MIT Press, 1992.
- (11) S. Särkkä: "Notes on Quaternions", Internal Technical Document, Helsinki University of Technology, 2007.
- (12) W. I. Grosky and L. A. Tamburino: "A Unified Approach to the Linear Camera Calibration Problem", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No. 7, pp. 663-671, 1990.