ICIC Express Letters ICIC International ©2016 ISSN 1881-803X
Volume **10**, Number **3**, March **2016** pp. **563–568**

3D OBJECT POSE TRACKING USING A NOVEL MODEL-BASED CONTOUR FITTING ALGORITHM

Chi-Yi Tsai*, Wei-Yi Wang and Shu-Hsiang Tsai

Department of Electrical Engineering
Tamkang University
No. 151, Yingzhuan Road, Tamsui District, New Taipei City 25137, Taiwan
*Corresponding author: chiyi_tsai@mail.tku.edu.tw

Received October 2015; accepted January 2016

Abstract. *This paper presents a novel three-dimension (3D) object pose tracking algorithm, which is an important function to implement a robust vision system using a monocular camera. The proposed algorithm is based on a novel computer-aided design (CAD) model-based contour fitting algorithm, which is a nonlinear optimization process for estimation of the optimal object pose using a contour-based distance metric. Given a CAD model of the object-of-interest, the proposed system first computes a projected contour model of the object in the image plane. Then, the 6-DOF pose of the object is estimated by minimizing sum of fitting errors between the projected contour model and the detected contour of the object in a captured image. Experiment results show that the proposed pose tracking algorithm not only efficiently estimates 3D pose of the object, but also overcomes partially occlusion problem during visual tracking process.*

Keywords: 3D object pose tracking, Object pose estimation, Contour fitting, Nonlinear optimization

1. **Introduction.** In recent years, visual sensing systems have become more and more popular used in robotic studies. For example, industrial robot manipulators require the information of 3D pose of an object-of-interest (OOI) in order to perform complex pick-and-place tasks. In this case, vision-based pose tracking techniques provide an efficient solution to track 3D pose of the OOI for the robot to automatically complete more complex tasks.

Estimation of 3D pose information of targets is an important task in applications of robot manipulation, but it is also a difficult task because of most visual sensing systems only providing 2D images without depth information. Therefore, the development of a vision-based pose tracking system to estimate the 3D pose information of an OOI from 2D images of the OOI is an important issue in practice. To develop a vision-based 3D pose tracking system, the existing object pose estimation techniques are roughly divided into feature-based and model-based tracing methods. The former uses basic geometric features (such as points, lines, and circles), boundary contours [1,2], and regions of interest [3] of the object to estimate 3D pose of the OOI appearing in the image, and the latter uses a priori-constructed 3D CAD model of the target to detect the OOI in the image while tracking its 3D pose in camera frame [4]. Empirically, the model-based tracking approaches can provide better robustness than the feature-based ones; even the OOI is suffered from the partial occlusion problem [5] in a complex work environment.

To enable six degree-of-freedom (6-DOF) pose tracking, a fast and efficient method attaches an artificial mark to the OOI to simplify the computation of object detection and pose tracking processes [6]. This method is a simple and robust design to address the 6-DOF pose estimation problem; however, it only works with an additional marker on the OOI. In contrast, many researchers 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 repeatedly detected in images robust to translation, scaling, and rotation transformations. In keypoint-based approaches, the keypoints are first extracted from incoming images to detect the OOI using a keypoint matching operation. Then, a keypoint classifier, which was trained from a large training dataset with different view ports of the OOI, is used to directly determine the 6-DOF pose information of the OOI. By doing so, the 3D pose tracking problem is simplified as a keypoint recognition problem, which can be efficiently resolved using fast nearest neighbor keypoint matching methods, such as k-d tree method [8], and multiple random trees method [9].

On the other hand, edge-based pose tracking methods employ a 3D CAD model of the OOI to estimate its 6-DOF pose from the observed OOI image. In [10], Harris introduced a real-time attitude and position determination (RAPiD) algorithm, which is a real-time model-based tracking method for a known 3D object executing arbitrary motion observed from a single video camera. In [11], Choi and Christensen combined a global and a local pose estimation design to implement a RAPiD style tracking system. In global pose estimation, a keypoint-based object recognition method was employed to decide initial pose information of the OOI detected from the image. In local pose estimation, an edge-based back-projection fitting algorithm was developed to continuously track its 6-DOF pose using a given 3D CAD model of the OOI. The RAPiD style tracking system is an efficient 6-DOF pose tracking system; however, only polygon mesh models can be used as the object CAD model in this system.

In this study, a novel contour-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 contour fitting process based on an edge-based distance metric, which can work with 3D polygon edge and polygon mesh models to address model-based 3D pose tracking problem efficiently. Suppose that a 3D CAD model of the OOI is known a priori. Then, the proposed model-based 3D pose tracking algorithm employs the CAD model to estimate 3D pose of the OOI with respect to (w.r.t.) the camera frame of a vision system from an observed object image by minimizing sum of fitting errors between the projected contour model and the detected contour of the OOI through a nonlinear optimization process. This feature allows increasing the robustness and applicability of a robotic vision system. Experimental results demonstrate the pose tracking performance of the proposed method.

In the remainder of the paper, Section 2 introduces the proposed CAD model-based 3D object pose estimation algorithm. Section 3 presents the experimental results to evaluate the effectiveness and efficiency of the proposed model-based 6-DOF pose tracking method. Section 4 concludes the contributions of this paper.

2. The Proposed Method.

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

After matching the detected contour lines of the OOI and the contour lines of the model projection, the proposed nonlinear model fitting process aims to estimate the 6-DOF pose of the OOI by fitting the contour projection of the CAD model to the observed contour lines of the OOI. Let $\mathbf{R}$ and $\mathbf{T}$ denote a quaternion rotation matrix [12] and a translation

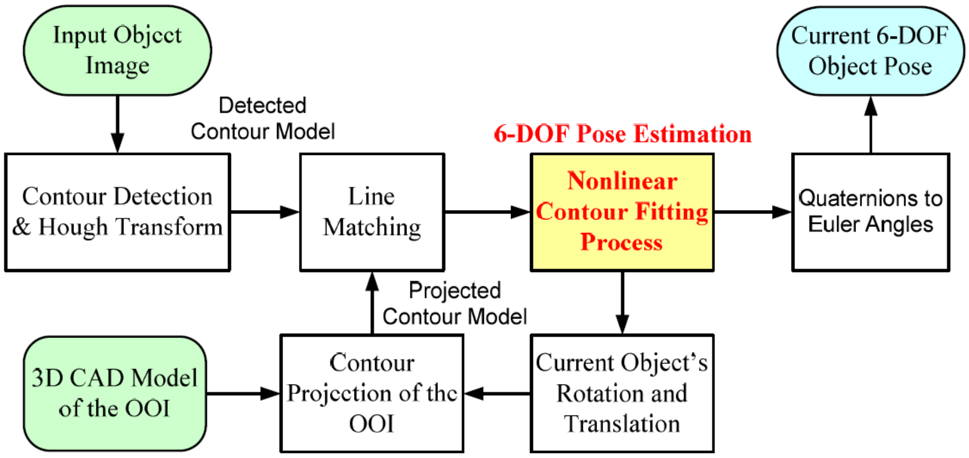


Figure 1. Flowchart of the proposed 6-DOF pose tracking system

vector such that

$$\mathbf{R}(w_q, x_q, y_q, z_q) = \begin{bmatrix} 1-2y_q^2-2z_q^2 & 2x_qy_q+2w_qz_q & 2x_qz_q-2w_qy_q \\ 2x_qy_q-2w_qz_q & 1-2x_q^2-2z_q^2 & 2y_qz_q+2w_qx_q \\ 2x_qz_q+2w_qy_q & 2y_qz_q-2w_qx_q & 1-2x_q^2-2y_q^2 \end{bmatrix} \tag{1}$$

$$\mathbf{T}(t_x, t_y, t_z) = \begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T \tag{2}$$

where (w_q, x_q, y_q, z_q) are the four components of a quaternion, and (t_x, t_y, t_z) are the three components of a translation vector. Suppose that a camera intrinsic matrix $\mathbf{K}$ is computed a priori via a camera calibration process [13]. Let $\mathbf{X}_c = [x_c, y_c, z_c]^T$ denote a 3D control point on the target CAD model, and $\mathbf{p}_i = [x_i, y_i]^T$ is the corresponding image coordinate of $\mathbf{X}_c$ projected onto image plane. Then, the image control point $\mathbf{p}_i$ can be obtained by perspective projection of $\mathbf{X}_c$ w.r.t. an object pose $\mathbf{P} = (w_q, x_q, y_q, z_q, t_x, t_y, t_z)$ in camera frame such that

$$\tilde{\mathbf{p}}_i(\mathbf{P}) = s\mathbf{K} \begin{bmatrix} \mathbf{R}(\mathbf{P}) & \mathbf{T}(\mathbf{P}) \end{bmatrix} \tilde{\mathbf{X}}_c \tag{3}$$

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

Now we define an edge-based distance metric for the nonlinear contour fitting process. Suppose that a 2D Hough line located on the contour of the detected OOI is parameterized by $\mathbf{h} = (a, b, c)$. Then, the distance between the image control point $\mathbf{p}_i$ and the 2D Hough line is computed by

$$d(\mathbf{h}, \mathbf{p}_i) = \frac{|ax_i + by_i + c|}{\sqrt{a^2 + b^2}} \tag{4}$$

where d can be defined as a fitting error between the image control point $\mathbf{p}_i$ and the detected contour model in image plane. Suppose that there are m Hough lines, each of them having n_j corresponding image control points. Then, a nonlinear cost function to evaluate the fitting error between the projection of the CAD model and the detected contour model of the OOI w.r.t. an object pose in camera frame is defined as

$$f(\mathbf{P}) = \sum_{j=1}^{m} \sum_{k=1}^{n_j} d\left(\mathbf{h}_j, \mathbf{p}_i^{jk}(\mathbf{P})\right) \tag{5}$$

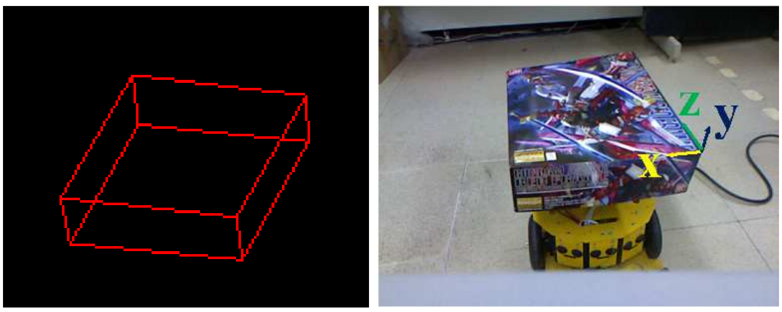
where $\mathbf{P}$ is the object pose in camera frame defined previously, and $\mathbf{p}_i^{jk}$ denotes the k-th image control point corresponding to the j-th Hough line. Finally, the optimal object pose can be estimated by minimizing the cost function $f(\mathbf{P})$ such that

$$\hat{\mathbf{P}} = \arg\min_{\mathbf{P}} f(\mathbf{P}) \tag{6}$$

where $\hat{\mathbf{P}}$ is the optimal object pose. Note that the initial value of $\mathbf{P}$ was set as (1, 0, 0, 0, 0, 0, 0) in our implementation, and the Levenbrg-Marquardt algorithm was used to resolve the nonlinear fitting problem defined in (6). The performance of the nonlinear contour fitting algorithm described above is validated in the next section.

3. **Experimental Results.** Figure 2(a) shows the 3D polygon CAD model used in the experiments, which is a rectangular cube model containing twelve line features. Figure 2(b) shows a rectangular box used as the target to be tracked in the pose tracking experiments. Figure 3 presents the experimental results of the proposed model-based 6-DOF pose tracking algorithm. Figures 3(a1)-3(a3) show the pose tracking results obtained from the proposed algorithm when the target rotated along y-axis. It is clear that the proposed algorithm successfully fits the contour lines between the projected CAD model and the detected OOI in the incoming images. Similar results also can be observed from Figures 3(b1)-3(b3), in which the target rotated along a different axis. Table 1 records the 6-DOF pose tracking results of the target shown in Figure 3. From Table 1, the yaw angle of the target firstly is changed from $-0.1515°$ to $-14.8174°$ as the OOI rotates left along y-axis. Next, the pitch angle of the target is changed from $-5.1972°$ to $-13.2776°$ when the OOI rotates left along x-axis. Thus, the proposed tracking algorithm successfully estimates the motion trajectory of the target in camera frame. Moreover, Figure 4 shows the pose tracking results under partial occlusion conditions. It is clear that the proposed algorithm can overcome partial occlusion problem during pose tracking process. Therefore, the above experimental results validate the tracking performance and robustness of the proposed method.

4. **Conclusions and Future Work.** A novel CAD model-based 6-DOF pose tracking algorithm is presented in this paper. The proposed algorithm is a nonlinear contour fitting process based on a new edge-based distance metric. Experimental results validate the tracking performance and robustness of the proposed method. In future work, some

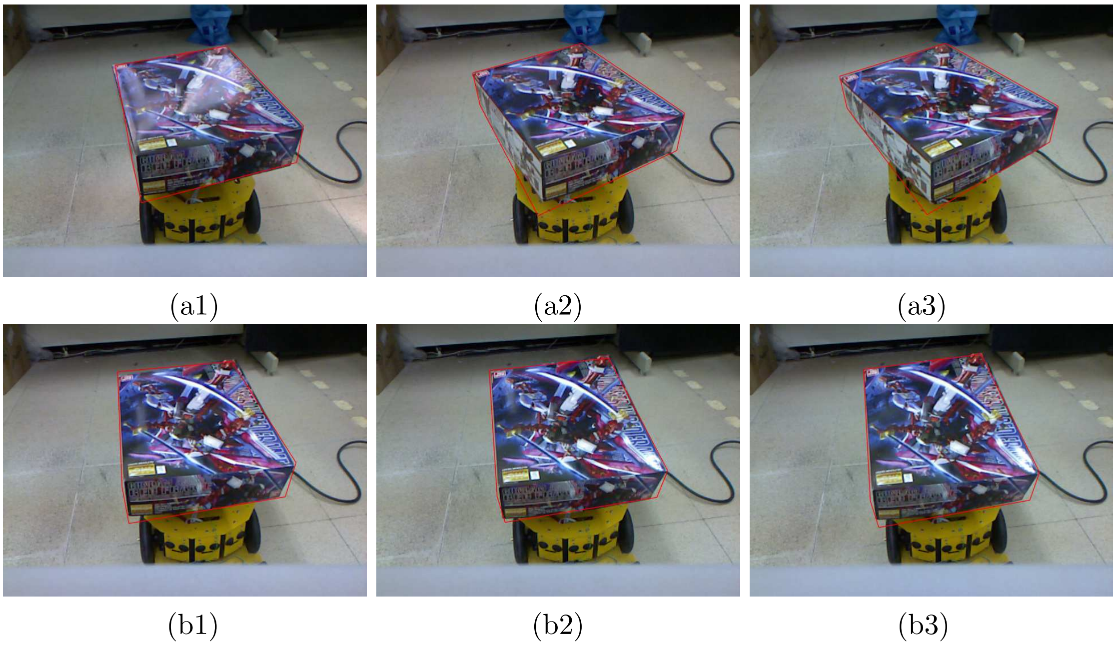


(a) (b)

Figure 2. Experimental setup: (a) the 3D polygon CAD model used in the experiments, and (b) the target object used in the experiments

Table 1. 6-DOF pose tracking results

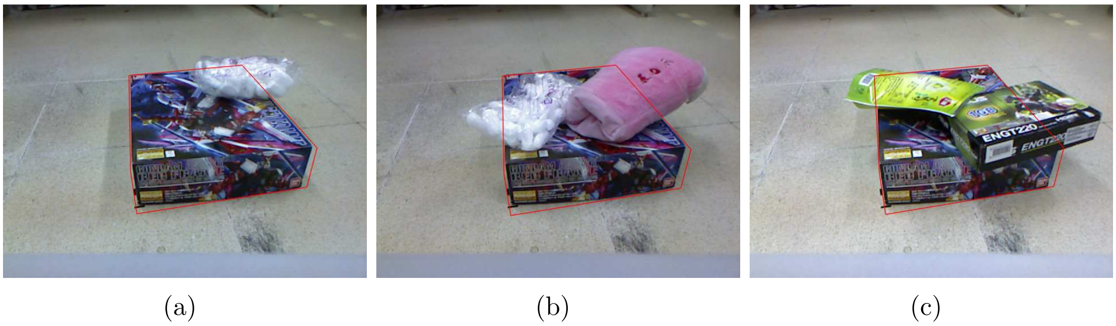
Figure 3	(a1)	(a2)	(a3)	(b1)	(b2)	(b3)
Pitch (degree)	−0.9762	−0.4792	−1.4983	**−5.1972**	**−13.0595**	**−13.2776**
Yaw (degree)	**−0.1515**	**−9.0573**	**−14.8174**	0.3372	−1.3862	2.4264
Roll (degree)	4.4816	10.0169	11.0560	−0.9124	−0.5219	0.3422
t_x (m)	0.0016	0.1100	0.1803	−0.0178	−0.0274	−0.0383
t_y (m)	0.0083	0.0023	−0.0032	0.05379	0.1309	0.1369
t_z (m)	0.0063	−0.0117	0.0009	−0.0028	0.0375	0.0343



(a1) (a2) (a3)

(b1) (b2) (b3)

FIGURE 3. Experimental results of the proposed model-based pose tracking algorithm



(a) (b) (c)

FIGURE 4. Pose tracking results under partial occlusion conditions

comparisons between the proposed method and previous methods will be carried out to further validate the advantage of the proposed algorithm.

Acknowledgment. This work was supported by the Ministry of Science and Technology (MOST) of Taiwan, under grant MOST 103-2221-E-032-068 and 103-2632-E-032-001-MY3. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] M. Isard and A. Blake, Contour tracking by stochastic propagation of conditional density, *Proc. of European Conference on Computer Vision*, Cambridge, UK, pp.345-356, 1996.
[2] Y. Fu, A. T. Erdem and A. M. Tekalp, Tracking visible boundary of objects using occlusion adaptive motion snake, *IEEE Trans. Image Processing*, vol.9, no.12, pp.2051-2060, 2000.
[3] G. D. Hager and K. Toyama, X vision: Combining image warping and geometric constraints for fast visual tracking, *Lecture Notes in Computer Science*, vol.1065, pp.507-517, 1996.
[4] A. I. Comport, É. Marchand and F. Chaumette, Robust model-based tracking for robot vision, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sendai, Japan, pp.692-697, 2004.
[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*, Columbus, USA, 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*, Taipei, Taiwan, 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*, San Diego, USA, pp.775-781, 2005.

[8] E. Kushilevitz, R. Ostrovsky and Y. Rabani, Efficient search for approximate nearest neighbor in high dimensional spaces, *SIAM Journal on Computing*, vol.30, no.2, pp.457-474, 2000.

[9] V. Lepetit and P. Fua, Keypoint recognition using randomized trees, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.28, no.9, pp.1465-1479, 2006.

[10] C. Harris, *Tracking with Rigid Objects*, MIT Press, 1992.

[11] 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*, Anchorage, Alaska, pp.4048-4055, 2010.

[12] S. Särkkä, Notes on quaternions, *Internal Technical Document*, Helsinki University of Technology, 2007.

[13] C. Chatterjee and V. A. Roychowdhury, Algorithms for coplanar camera calibration, *Machine Vision and Applications*, vol.12, no.2, pp.84-97, 2000.