

MOVING-OBJECT RECOGNITION USING PREMARKING AND ACTIVE VISION

D. He, D. Hujic, J.K. Mills[†] and B. Benhabib^{*}

Computer Integrated Manufacturing Laboratory (CIMLab), Department of Mechanical Engineering
University of Toronto, 5 King's College Road, Toronto, Ontario, M5S 3G8, *email:beno@me.utoronto.ca

ABSTRACT

This paper presents an active-vision system for the recognition of 3D objects moving along predictable trajectories. The novelty of this system lies in its unique approach to deal with the problem of moving-object recognition, by integrating object pre-marking, object-trajectory-prediction and time-optimal robot-motion techniques developed in our laboratory. The recognition technique is an extension of our earlier work on static-object recognition. Therein, objects were pre-marked optimally using circular markers, which are utilized during run time for guiding a robot-mounted camera to acquire 2D standard-views for efficient matching purposes. The Kalman-filter based prediction of the object trajectory and the time-optimal movement of the mobile camera for image acquisition are also based on our earlier research results on moving-object interception. The discussion of the various implementation issues and experimental results presented in this paper should provide researchers with useful tools and ideas.

1. INTRODUCTION

The development and implementation of an active-vision system for moving-object recognition involves numerous issues that currently constitute individual research endeavors in different laboratories. The two primary issues, however are (i) tracking and (ii) recognition of moving objects. The latter area attempts to obtain information, for the identification of the objects, from consecutive motion images. Two common approaches to the solution of motion-image processing and recognition problems have been: the optical-flow approach and feature-based approach [1]. Feature-based approaches require that correspondence be established between a set of features extracted from one image and those from its next image [2,3]. Optical-flow approaches, on the other hand, rely on local spatial and temporal derivatives of image-brightness values. No correspondence between successive images is required. Although many methods have been developed based on both approaches, they are only partial solutions, suitable

for simplified environments, sensitive to noise, and computationally expensive [1].

Integration of tracking and recognition techniques to form robust active-vision systems appear to be a logical answer to greatly simplify the above-mentioned problems. Feddema and Lee [4] proposed such an adaptive system for visually tracking a moving object with a single mobile camera. This system uses the concept of active sensing to reduce the time for feature searching and extraction. The vision system accurately predicts the object location at the next sampling time. When a tracking task begins, the vision system uses all *a priori* knowledge to recognize and locate the object. After the initial recognition stage, the image processing is reduced to a simple verification process. Hwang, Ooi and Ozawa [5] developed an alternate adaptive-sensing system with the capability of tracking and zooming onto moving objects using a fuzzy-logic controlled camera. Another active-vision system, which uses an active camera mounted on a pan/tilt platform, was proposed by Murray and Basu, for real-time motion detection [6]. This system successfully extracted moving edges from dynamic images. The camera motion is compensated by the tracking algorithm, which allows static techniques to be applied to active image sequences.

In contrast to the above proposed methods, the primary advantage of the active-vision system, designed and successfully implemented in the CIMLab, is the recognition of 3D moving-objects using a 2D modeling and matching process [7]. Based on an integration of active sensing and object pre-marking principles, the system is capable of successfully tracking and recognizing moving-objects defined *a priori* in a given object library. In the next sections, we will first briefly describe the individual components of the system, which would subsequently lead to the discussion of the implementation issues and experimental results.

2. THE ACTIVE-VISION SYSTEM – AN OVERVIEW

The task of moving-object recognition is broken herein into two sub-tasks: (i) tracking and prediction of the

[†] Laboratory for Non-Linear Systems Control, Department of Mechanical Engineering, University of Toronto

object's trajectory, and (ii) object recognition from consecutive images. The motivation behind this approach is twofold: to gain speed via parallel computing and to simplify the implementation of the system. Based on this rationale, the active-vision system was designed as a two-module architecture: namely, comprising the '*object recognizer*' module and the '*object-trajectory predictor and robot-motion planner*' module, Figure 1. The former module functions as an image processor and pattern recognizer. The latter module deals with the prediction of the object motion and planning of the robot trajectory, so as to guide the robot-mounted camera to desired locations for the acquisition of consecutive images.

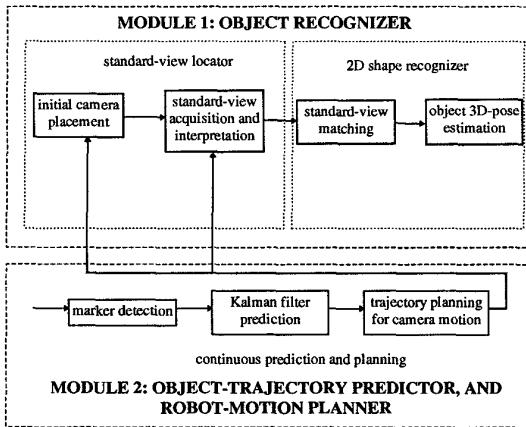


Figure 1. Structure of the Active-Vision System.

The object-recognizer module was based on the presumption that a 3D-object can be modeled by a pre-defined set of its 2D views referred to as *standard-views* [7]. Run-time recognition is then initiated by acquiring one of these standard-view images and followed by the matching process. Since markers placed on an object define a limited set of unique standard-views, by pre-marking objects, the 3D-recognition process is simplified into a *fast* 2D-matching process.

Based on the above principle, the recognition process is carried out in two stages. During Stage 1, a marker's motion is observed via a fixed camera and its trajectory is predicted by Module 2. In parallel, a robot-mounted camera is utilized by Module 1 to acquire two sequential images of the same viewed marker. The 3D pose of the circular feature is determined through the use of these two images. The correspondence between the poses of the same marker in the two consecutive images identifies the true orientation of the marker. During Stage 2, a time-optimal robot trajectory is planned to position the camera for the acquisition of a standard-view of the moving object at the right instant. Once a standard-view is acquired, recognition is conducted by matching the

object's shape signature with those in the standard-view library.

3. THE OBJECT-RECOGNIZER MODULE

An optimal number of circular markers of known size are used for pre-marking the objects, whose normals define the necessary standard-viewing axes. A standard-view is acquired by aligning the optical axis of the camera with one of the standard-viewing axes of the object. Visible markers, however, undergo perspective projection and would be perceived as elliptical shapes in arbitrarily acquired images. Thus, the parameters of these elliptical shapes must be used to determine the 3D position and orientation (pose) of the marker.

3.1 Calculation of the Elliptical Shape Parameters [8]

The parameters of the elliptical shape of a circular marker's acquired image can be calculated as follows.

Let

$$Q(X, Y) = aX^2 + bXY + cY^2 + dX + eY + f = 0 \quad (1)$$

be the general equation of an ellipse, and

$$(X_i, Y_i) \quad i = 1, N \quad (2)$$

be a set of boundary points on the marker's image to be fitted. The (five) elliptical parameters can be then computed by minimizing the following error function:

$$J_2 = \sum_{i=1}^N [w_i Q(X_i, Y_i)]^2, \quad (3)$$

where w_i are the weighing factors that take into account the non-uniformity of the data points along the ellipse's boundary.

3.2 Estimation of a Circular-Marker's 3D-Pose [9]

In order to move the robot-mounted camera to a standard-viewing position, the 3D pose of a circular marker has to be determined first. Circular-marker pose estimation is equivalent to the solution of the following problem: Given a 3D conic surface, defined by an elliptical base (the perspective projection of a circular feature in the image plane) and a vertex (the center of the camera's lens) with respect to a reference frame, determine the pose of a plane (with respect to the same reference frame) which intersects the cone and generates a circular curve.

The general form of the equation of a cone with respect to the image frame is as follows:

$$\begin{aligned} ax^2 + by^2 + cz^2 + 2fyx + 2gzx \\ + 2hzy + 2ux + 2vy + 2wz + d = 0. \end{aligned} \quad (4)$$

An intersection plane can be defined by $lx + my + nz = 0$. Therefore, the problem of finding the coefficients of the equation of a plane, for which the intersection is circular, can be expressed mathematically as: determine l, m and n such that the intersection of the conical surface with the following surface is a circle: $lx + my + nz = 0$, where $l^2 + m^2 + n^2 = 1$.

To solve the problem, first the equation of the conical surface can be reduced to a more compact form:

$$\lambda_1 X^2 + \lambda_2 Y^2 + \lambda_3 Z^2 = \mu, \quad (5)$$

where the XYZ -frame is called the canonical frame of conicoids. It can be shown that the reduction of the general equation of a cone to the above form results in a closed-form analytical solution. There exist two possible solutions to the problem. To obtain a unique acceptable solution, as part of the moving-object recognition process, an extra geometrical constraint, such as the change of eccentricity in a second image, has to be obtained.

To obtain a unique solution for a marker's position, the radius of the circular feature has to be known. There exist two solutions: one on the positive Z -axis, and one on the negative Z -axis. Since only the positive one is acceptable in our case (being located in front of the camera), the coordinates of the center of the circle (X'_0, Y'_0, Z'_0) with respect to the $X'Y'Z'$ -frame are found to be:

$$\begin{cases} X'_0 = -\frac{B}{A} Z'_0 \\ Y'_0 = -\frac{C}{A} Z'_0 \\ Z'_0 = \frac{Ar}{\sqrt{B^2 + C^2 - AD}} \end{cases} \quad (6)$$

where A, B, C and D are defined in terms of the elements of the transformation between the XYZ -frame and $X'Y'Z'$ -frame, λ_i ($i=1,2,3$) and the known radius r .

3.3 Solving the 3D-Orientation-Duality in a Single-Marker-Scene

As mentioned in Section 3.2, in order to obtain a unique solution for the orientation of a circular marker, acquisition of multiple images of the same marker would be necessary. In general, obtaining a unique solution for the 3D pose of a circular marker in an unknown motion

requires two consecutive images with at least three visible circles [3]. In our proposed system, however, a unique solution can be calculated with one circular marker given that the object is subject to certain motion constraint, and the size of the marker is known. For example, when the object is moving in translation or planar motion the orientation can be uniquely determined. Figure 2 depicts a situation where the marker undergoes 3D translation. Point O is the camera's focal center and Plane I is the image plane. From the first image, using the technique presented in Section 3.2, two possible surface normals of a circular marker, denoted as unit vectors \mathbf{n}_1 and \mathbf{n}_2 , can be computed. As the marker moves to the second position, another two possible solutions \mathbf{n}'_1 and \mathbf{n}'_2 are obtained. By the definition of translation, the surface normal vector of a translating plane remains constant. Therefore, the true solution of the surface orientation can be distinguished as the one that remains unchanged. In Figure 2, since $\mathbf{n}_1 = \mathbf{n}'_1$, while $\mathbf{n}_2 \neq \mathbf{n}'_2$, \mathbf{n}_1 and \mathbf{n}'_1 are found to be the true solution.

If the object motion is limited to planar motion, then a similar concept can be applied. We know that when a vector in space undergoes planar motion, its z component remains constant. Therefore, defining m_i and m'_i as the z component of \mathbf{n}_i and \mathbf{n}'_i respectively, if $m_1 = m'_1$, while $m_2 \neq m'_2$, then \mathbf{n}_1 and \mathbf{n}'_1 are the true orientations of the circular marker.

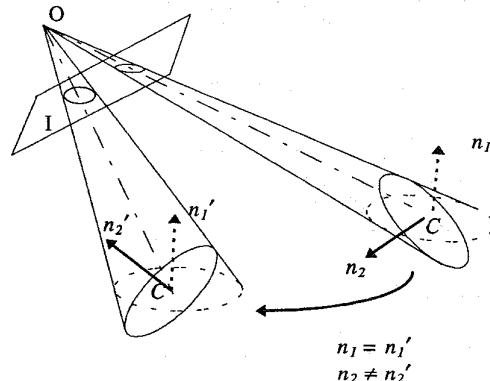


Figure 2. Elimination of the false solution in the case of translation.

3.4 Standard-View Matching [11]

When a standard-view image is acquired and the silhouette of the object is extracted, the resulting chain-coded contour of the object is used to compute the global eccentricity measure and the shape signature. The object is then identified by matching its feature vector, consisting of a global eccentricity measure and the shape signature, with the feature information of the standard views in the database. The identity of the shape is determined by the minimum-distance rule. However, if the measures of

dissimilarity between the acquired shape and multiple standard views are below a certain threshold, the object cannot be uniquely recognized. In this case, the system will identify the several candidates in the database and pass the control back to the active vision system for the acquisition of an additional standard view to resolve the ambiguity at hand.

4. THE OBJECT-TRAJECTORY PREDICTOR AND ROBOT-MOTION PLANNER MODULE

4.1 Determination of Optimal Initial Camera Orientation [12]

Optimal camera placement is necessary to maximize marker detectability, since this would consequently allow us to minimize the number of markers placed on objects. Several optimization problems were formulated and solved in [12]. The specific problem pertinent to this work is finding: the minimum number of markers to be placed on a given set of objects, such that the visibility of at least one marker (on a randomly appearing object) is guaranteed in a single-camera environment. The outcome of the optimization is the minimum number and locations of the markers on the object, as well as the optimal initial-viewing angle (γ) of the camera.

4.2 Object-Trajectory Prediction [13]

A recursive Kalman filter (KF) was proposed in [13] to obtain optimal estimates of the object's present two-dimensional position, as well as predictions of its future trajectory. The recursive KF is a computationally efficient algorithm, which yields an optimal least-squares estimate of a state from a series of noisy observations. It produces a new optimal estimate from an additional observation without having to reprocess past data. The KF can also be used to obtain multiple-step-ahead predictions by propagating the KF extrapolation equations (i.e., one-step ahead predictor). As will be discussed in the next sub-section, these multiple step-ahead predictions are used in our system to guide the mobile camera to optimal viewing locations.

4.3 Motion Planning for Mobile-Camera

The problem addressed here is two-fold: (i) finding an optimal viewing point within the robot's workspace, and (ii) generate time-optimal robot trajectory to this viewing-point. As previously shown in [14] for a moving-object interception problem, these two issues are strongly coupled and should be addressed simultaneously for achieving time-optimal results.

(i) First Marker-Viewing Location

Given the predicted trajectory of an object's travel through the robot's workspace, a corresponding camera placement trajectory, $\{G(t)\}$, can be determined, via a constant transformation, to represent potential viewing-points. Using this camera trajectory and the robot's latest position, our objective is to find a time-optimal initial-viewing point on $\{G(t)\}$. A solution of this problem was provided in [14] and will not be repeated here.

It should be noted however, that while the current robot trajectory is executing, the planning module continuously re-plans the unexecuted portion of the robot-mounted camera's motion in response to new information generated by the object-motion prediction algorithm.

(ii) Standard-Viewing Location

The robot-mounted mobile camera has to be re-located, from its first viewing location, in order to acquire a standard-view of the object. As for the solution utilized above, this standard-view acquisition location can be optimally determined for the mobile camera, based on the information provided by the object-trajectory prediction module and the calculated standard-viewing-axis orientation, Section 3.3. In order to allow real-time applicability, time-optimal task-space quintic polynomials are used in our work for robot trajectories.

5. EXPERIMENTAL SYSTEM

5.1 Experimental Setup

The experimental system is an integration of the following, Figure 3:

(a) The "object-trajectory predictor and robot-motion planner" module:

- *Host computer 1*: 80486 PC DX4 100 MHz.
- *Imaging subsystem*: A Hitachi 30Hz CCD camera fixed above the object motion plane. A PC-based PIP Matrox digitizer board with 640x480 resolution.
- *Software*: A KF-based object-motion-prediction algorithm. A camera-viewing-location-planner and robot-motion-planner algorithm.

(b) The "object recognizer" module:

- *Host computer 2*: 80486 PC DX 33MHz.
- *Imaging subsystem*: A JVC 30Hz CCD camera mounted on the a six degree-of-freedom GMF S-100 robot's wrist (fifth link). A PC-based PIP Matrox digitizer board with 640x480 resolution.
- *Software*: Object-recognition algorithms.

(c) Other auxiliary subsystems:

- *Object-motion simulator*: A NC X-Y table, controllable from a RS-232 port, used to produce the object motion.

- Communication interface: Implemented on a 9600-bps RS-232 serial communication line, to provide exchange of commands and data amongst Host 1, Host 2, robot controller and the NC-table controller.

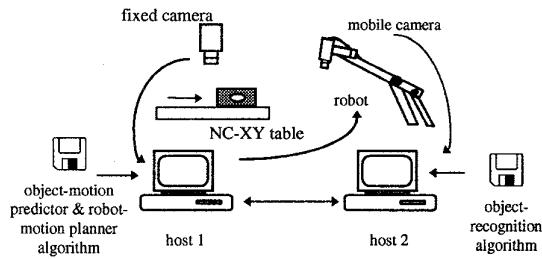


Figure 3. Experimental setup.

5.2 System Implementation

The Hitachi CCD camera with a 25 mm lens is placed 1.8 m above the surface of the X-Y table. This setup yields a 600×400 mm² field of view with ~ 0.93 mm/pixel resolution. The JVC CCD camera, mounted near the robot's end-effector, also had a 25 mm lens. Both cameras were calibrated using the mono-view non-co-planar point technique proposed in [15]. The error in X and Y directions was less than 0.5% for both. Different object-motion trajectories were induced via programmed movement of the NC table at speeds from 4~12 mm/s (limited by the travel length and field of view in our experiment, but can be increased to higher values in a different setting).

Since the objects were pre-marked using red markers, a red-color filter, was used to threshold the analog camera-signals such that only one feature, the circle's centroid, is tracked. After locating the marker in the camera's field of view, the marker's centroid is determined, and used to update the KF. One-step-ahead KF prediction is used to follow the marker across the field of view. At present, the entire process (i.e., grabbing an image, finding the object's centroid, and updating the KF) takes ~ 65 msec.

Once the object's trajectory has been predicted and the robot has reached its initial viewing-position, two consecutive images are taken. Markers on the object are detected by the mobile camera. Based on the algorithm presented earlier, a marker's pose, represented by a set of 3D coordinates of the marker center and the surface normal vector of the marker, is determined. The marker's pose is passed to the motion-planning system, which in turn sends the mobile camera to its standard-view-acquisition location. Subsequently, the recognition system performs the matching of 2D signature.

5.3 Results

The recognition system was implemented and tested successfully for a set of seven different objects, shown in Figure 4. Their sizes ranged from $40 \times 40 \times 35$ mm³ (Object #2) to $90 \times 90 \times 70$ mm³ (Object #5). All the objects were distinguished and classified successfully in our experiments. Figure 5(a)-(d) show images taken by the robot-mounted camera at different stages. Initially (Time 0), the camera is placed at a *home* position. When the overhead camera detects the object, the trajectory planner guides the robot to move to the initial viewing position, which is approximately 1 m above the X-Y table, aiming at the randomly posed object at an angle of 63 degrees. At Time 1, the first image is obtained and a unique solution for the position of the marker is calculated, Fig. 5(a), (the elliptical image of the marker is highlighted, and its major and minor axes are shown). Two possible surface normal vectors of the surface are calculated and stored. At Time 2, when the X-Y table moves to a second position at an arbitrary speed, a second image is taken, Fig. 5(b), and the true surface normal is identified. The surface normal value is sent to the trajectory planner for the planning of the camera path. Subsequently, the robot automatically moves to the standard-viewing position. Without any delay, the mobile camera is able to take a snapshot of the standard-view as soon as it arrives at the standard-viewing position, Fig. 5(c). From this view, the object is successfully recognized as Object #7, Fig. 5(d).

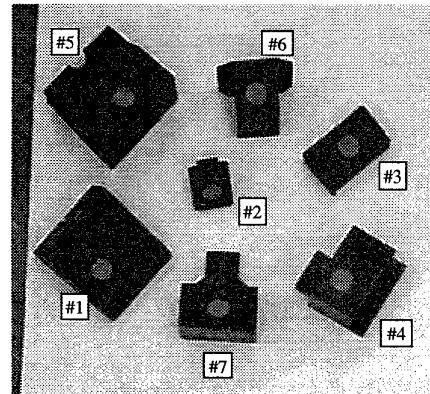


Figure 4. Tested objects.

6. CONCLUSIONS

As existing general approaches for moving-object-recognition have proven to be difficult to implement, active-vision systems have shown their potential in simplifying the problem and thus expediting the process. The active-vision system presented in this paper demonstrated a collection of novel techniques used in tracking, trajectory planning, recognition and camera

placement. This system, which features the principles of active sensing and object-pre-marking, is capable of recognizing pre-marked objects moving along predictable paths.

Our system is only a potential implementation example and should not be viewed as globally optimal. Variety of issues, especially in real-time imaging, still remain to be addressed.

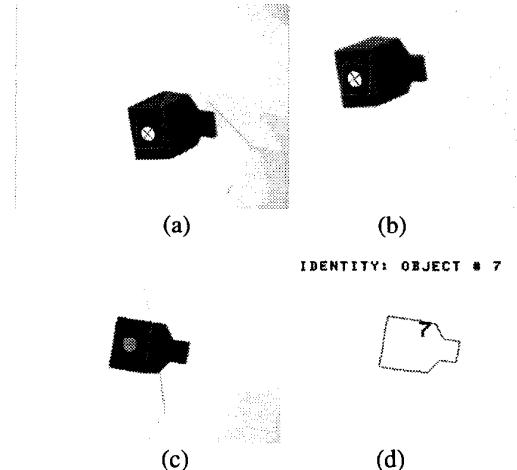


Figure 5. (a) Image of the object taken at Time 1;
(b) Time 2; (c) Time 3, a standard-view of the object;
(d) Object is recognized as Object #7.

REFERENCES:

- [1] J.K. Aggarwal and N. Nandhakumar, "On the Computation of Motion from Sequence of Images: A Review", *Proceedings of the IEEE*, Vol. 76, No. 8, August 1988, pp. 971-935.
- [2] T.S. Huang and A.N. Netravali, "Motion and Structure from Feature Correspondence: A Review", *Proceedings of the IEEE*, Vol. 82, No. 2, Feb. 1994, pp. 252-268.
- [3] S.D. Ma, "Conics-Based Stereo, Motion Estimation, and Pose Determination", *International Journal of Computer Vision*, Vol. 10, No. 1, 1993, pp. 7-25.
- [4] J.T. Feddema and C.S.G. Lee, "Adaptive Image Feature Prediction and Control for Visual Tracking with a Hand-Eye Coordinated Camera", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 20, No. 5, September/October 1990, pp. 1172-1183.
- [5] J. Hwang, Y. Ooi, and S. Ozawa, "An Adaptive Sensing System with Tracking and Zooming a Moving Object", *IEICE Transactions on Information & Systems*, Vol. E76-D, No. 8, Aug. 1993, pp. 926-934.
- [6] D. Murray and A. Basu, "Active Tracking", *International Conference on Intelligent Robots and Systems*, Yokohama, Japan, July 1993, pp. 1021-1028.
- [7] R. Safaee-Rad, I. Tchoukanov, X. He, K.C. Smith, and B. Benhabib, "An Active-Vision System for Recognition of Pre-Marked Objects in Robotic Assembly Workcells", *IEEE Conference on Computer Vision and Pattern Recognition*, N.Y. City, U.S.A., June 1993, pp. 722-723.
- [8] R. Safaee-Rad, K.C. Smith, B. Benhabib and I. Tchoukanov, "Application of Moment and Fourier Descriptors to the Accurate Estimation of Elliptical Shape Parameters", *Pattern Recognition Letters*, Vol. 13, July 1992, pp. 479-508.
- [9] R. Safaee-Rad, K.C. Smith, B. Benhabib, and I. Tchoukanov, "3D-Location Estimation of Circular Features for Machine Vision", *IEEE Transactions on Robotics and Automation*, Vol. 8, No. 5, 1992, pp. 624-640.
- [10] D. He, M. Tordon and B. Benhabib, "An Active-Vision System for the Recognition Moving Objects", *IEEE International Conference on Systems, Man and Cybernetics*, Vancouver, B.C. Oct. 1995, pp. 2274-2279.
- [11] I. Tchoukanov, R. Safaee-Rad, K.C. Smith, and B. Benhabib, "The Angle-of-Sight Signature for 2D-Shape Analysis of Manufactured Objects", *Journal of Pattern Recognition*, Vol. 25, No. 11, Dec. 1992, pp. 1289-1305.
- [12] X. He, B. Benhabib, and K.C. Smith, "CAD-Based Off-line Planning for Active Vision", *Journal of Robotic Systems*, Vol. 12, October, 1995.
- [13] D. Hujic, G. Zak, E.A. Croft, R.G. Fenton, J.K. Mills and B. Benhabib, "An Active Prediction, Planning and Execution System for Interception of Moving Objects", *IEEE International Symposium on Assembly and Task Planning*, Pittsburgh, Aug. 1995, pp. 347-352.
- [14] E.A. Croft, R.G. Fenton, and B. Benhabib, "Time-Optimal Interception of Objects Moving Along Predictable Paths", *IEEE International Symposium on Assembly and Task Planning*, Pittsburgh, PA, Aug. 1995, pp. 419-425.
- [15] R.Y. Tsai, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", *IEEE Journal of Robotics and Automation*, Vol. 3, No. 4, Aug. 1987, pp. 323-344.