Reliable Solution for Object Pose Determination using an Active Vision System

Yan Meng
Department of Electrical and Computer Engineering
Stevens Institute of Technology
Hoboken, NJ 07010, USA
ymeng1@stevens.edu

Hanqi Zhuang
Department of Electrical Engineering
Florida Atlantic University
Boca Raton, FL 33431
zhuang@fau.edu

Abstract – **The problem of pose determination using a calibrated structured light system is to determine the position and orientation of an object relative to a camera coordinate frame. By projecting a multiple-line light pattern to the three planes of the object, a number of line-to-plane correspondences can be obtained, from which an efficient and compact closed-form solution for the pose determination is derived. The performance of the proposed algorithm in the presence of noise is also investigated. Preliminary experiments conducted with a structured light system shows that this novel method is robust, efficient and is applicable to the real-time pose determination of mobile robot.**

Index terms: **pose determination, structured light system.**

I. INTRODUCTION

Determination of 3D rotation and translation is an important problem that has wide applications in robotics and computer vision. Pose determination involves the estimation of the pose of an object in a reference coordinate frame. Depending on the way the measurements are taken, the pose determination can be classified as, among others, 2D-to-3D and 3D-to-3D correspondence problems.

In 2D to 3D correspondence problem, pose determination are performed based on defining a match between sets of 2D image features and sets of corresponding 3D features. The Stereovision is one of the computer vision techniques that can be used to obtain 3D information of the environment. By solving the correspondence problem between cameras, triangulation can be used to reconstruct the 3D position of the matched pixels. However, the correspondence problem is not always easy to solve. On the other hand, structured light vision system projects a known pattern onto the environment and grabs the images with a camera, thereby avoids the standard correspondence problem. Furthermore, the structured light sensor is an active device so it will continue to work in dark environments as well as environments in which the objects are featureless. In contrast, stereovision would fail in such texture-free circumstances. The pose determination with a structure light system has been applied in object recognition and shape reconstruction [1, 2]. Recently, these measurements devices have also been used to collect 3D face data for biometrics applications [3].

In 3D-to-3D correspondence problem, there are two sub-catalogs: point-to-point and line-to-plane matching. A classical example of point-to-point correspondence applications is find the location of a robot in the world coordinate frame, provided that positions of the robot end-effector are measured in both the robot base and the world frame. In a line-to-plane correspondence problem, lines represented in one coordinate frame and planes on which the lines lie are represented in another frame. The problem is to determine the pose of the second frame relative to the first frame, given a set of matched line and plane measurements. A typical example of the line-to-plane pose determination problem is to find the position and orientation of an object in a reference coordinate frame, say camera coordinate frame, using a structured-light vision system [4]. This problem is illustrated in Fig. 1. A light plane is projected onto the object by a laser device. The image of the lines corresponding to the intersection of the light plane with the object planes is processed. The direction vector and distance of the lines in the camera coordinate frame are then computed by using the geometry of the camera and the laser scanner (Assuming that the geometry of the structured light system has been calibrated). As can be seen, the lines are represented in the camera frame, while the object planes are represented in the object frame. By projecting a multiple-line light pattern into the object planes, a number of line-to-plane correspondences can be obtained, from which the pose of the object in the camera coordinate frame can be determined.

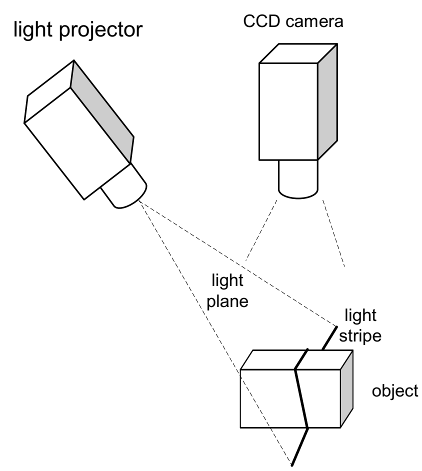


Fig. 1: A structured-light system

There have been active researches in solving line-to-plane correspondence problems. Various closed-form and numerical solutions were proposed in the literature [5-13]. As two typical examples, a closed-form solution method based on solving a high-order polynomial equation was developed in [4], together with the description of a set of necessary and sufficient conditions under which the line-to-plane pose determination problem can be solved. An iterative solution method based on the Levenberg-Marquardt nonlinear least squares optimization theory was given in [11]. Recently, results from comparative studies of four methods for registering 3D range data to 2D images were given in [13].

This paper deals with the 2D-line image and 3D-plane object correspondence problem initially, then with the calibrated structured light system, the 3D line correspondence in the camera coordinate frame can be obtained. Therefore, the problem becomes the 3D line-to-plane correspondence problem.

A closely related work to our research was reported in [12]. It proposed an algorithm to solve a special case of the line-to-plane pose determination problem. It was assumed that for each plane, there correspond at least three line segments. A light-stripe vision system depicted in Fig. 1 provides such measurement data. Under this assumption, a concise and computationally efficient closed-form solution, termed the triplet constraint algorithm, was derived based largely on the theory given in [6]. The idea of the algorithm is to use the invariant property of the angle between the two planes and the direction vector of the third line.

In this paper, it is assumed that there are at least three planes and each plane with two line segments on it. Under this assumption, a new method is proposed to solve the underline problem. With the method, the unknown rotation matrix and position vector can be obtained in a closed form. Furthermore, the necessary and sufficient conditions under which the pose determination problem can be solved become very intuitive.

The paper is organized as follows. In Section II, a formal statement of the problem is given. Section III presents our approach. Section IV provides experimental results. The paper concludes with a summary.

II. PROBLEM STATEMENT

Given 3-D data of multiple-lines in one coordinate frame (say the camera frame) and the corresponding planes in another frame (say the object frame), the problem at hand is to determine the transformation between the two coordinate frames (Fig. 2). Let the unit vector $\mathbf{m}_{i,j}$, denote the direction of a line $L_{i,j}$ in the camera frame, and the unit vector $\mathbf{n}_i$ denote the normal of a plane F_i in the object frame, where i denotes the number of planes and j denotes the number of lines on each plane.

$$\mathbf{m}_{i,j} = \begin{bmatrix} m_{xi,j} \\ m_{yi,j} \\ m_{zi,j} \end{bmatrix}, \qquad \mathbf{n}_i = \begin{bmatrix} n_{xi} \\ n_{yi} \\ n_{zi} \end{bmatrix}.$$

Let $\mathbf{p}_{i,j}$ be the vector of a point on line $L_{i,j}$ in the camera frame and d_i the shortest distance of plane F_i to the origin of the object coordinate frame. The problem can be formally stated as follows: given $\mathbf{m}_{i,j}$, $\mathbf{n}_i$, $\mathbf{p}_{i,j}$ and d_i ($i = 1,2,3,\ j = 1,2$), determine a rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$ such that the following equations are satisfied:

$$\mathbf{n}_i^T \mathbf{R} \mathbf{m}_{i,j} = 0 \tag{1}$$

and

$$\mathbf{n}_i^T (\mathbf{R}\mathbf{p}_{i,j} + \mathbf{t}) = d_i \tag{2}$$

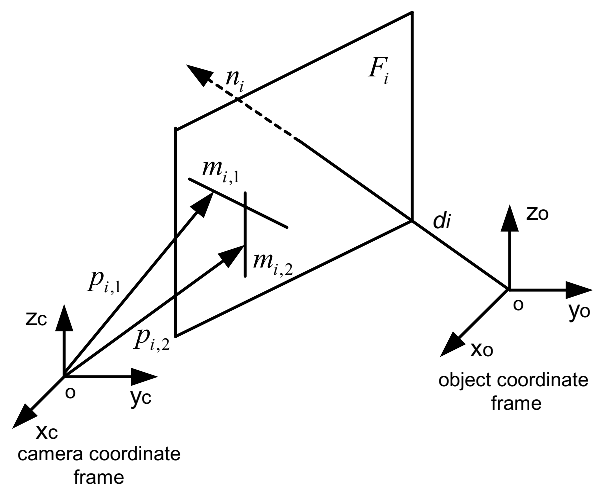


Fig. 2: The line data is described in the camera coordinate frame and the plane data is in the object coordinate frame

where T denotes matrix transposition, $\mathbf{R}$ and $\mathbf{t}$ are the rotation matrix and translation vector of the object coordinate frame related to the camera coordinate frame.

The geometric meaning of the above two equations is easy to understand. Equation (1) means that the rotated line vector will be perpendicular to the plane normal, and equation (2) says that the transformed point will line in the plane. Therefore, the whole line will be completely in the plane.

III. SOLUTION

Let

$$\mathbf{a}_i^T = \mathbf{n}_i^T \mathbf{R}, \qquad i = 1, 2, 3 \tag{3}$$

Then (1) can be rewritten as

$$\mathbf{a}_i^T \mathbf{m}_{i,j} = 0 \quad (4)$$

Each $\mathbf{n}_i$ corresponds to $\mathbf{m}_{i,j}$ on plane i, where $j = 1,2$. Since $\boldsymbol{a}_i$ is a unit vector according to equation (3), a cost function can be defined as

$$J = \sum_{j=1}^{2} \mathbf{a}_i^T \mathbf{m}_{i,j} \mathbf{m}_{i,j}^T \mathbf{a}_i + \lambda(1 - \mathbf{a}_i^T \mathbf{a}_i) . \quad (5)$$

If equation (4) is satisfied and $\boldsymbol{a}_i$ is a unit vector, the above cost function should equal to zero theoretically. Define

$$\mathbf{M}_i \equiv \sum_{j=1}^{2} \mathbf{m}_{i,j} \mathbf{m}_{i,j}^T . \quad (6)$$

The above cost function becomes

$$J = \mathbf{a}_i^T \mathbf{M}_i \mathbf{a}_i + \lambda(1 - \mathbf{a}_i^T \mathbf{a}_i) . \quad (7)$$

A solution to the problem can be obtained by minimizing J with respect to $\mathbf{a}_i$. Differentiating J with respect to $\mathbf{a}_i$ and setting it to zero, from

$$\frac{\partial J}{\partial \mathbf{a}_i} = 0 , \quad (8)$$

one obtains

$$\mathbf{M}_i \mathbf{a}_i = \lambda \mathbf{a}_i . \quad (9)$$

Equation (9) suggests that the solution $\mathbf{a}_i$ that minimizes the cost function J is the eigenvector corresponding to the minimum eigenvalue of $\mathbf{M}_i$. The following theorem states the condition under which $\mathbf{a}_i$ can be solved for:

Theorem 1. The necessary and sufficient conditions for $\mathbf{a}_i$ to be solved is that the two lines lie on the plane are not parallel.

Proof: In order to have a unique solution from (9), The rank of $\mathbf{M}_i$ must be two. It implies that the two lines on the plane should not be parallel. The converse is also true.

Q.E.D.

After $\mathbf{a}_i$ is known, the rotation matrix **R** can be determined from (3), assuming that the three planes are not mutually parallel. Since **R** is orthogonal matrix, $\mathbf{R}^{-1} = \mathbf{R}^T$. By multiplying $\mathbf{R}^T$ on both sides of Equation (3), we have

$$\mathbf{a}_i^T \mathbf{R}^T = \mathbf{n}_i^T \qquad i = 1, 2, 3. \quad (10)$$

The above equation can be rewritten as

$$\mathbf{R}\mathbf{a}_i = \mathbf{n}_i , \qquad i = 1, 2, 3. \quad (11)$$

There are a number of effective methods for solving rotation **R** in Equation (11): one is the Singular Value Decomposition (SVD) method, and another is the quaternion method [15]. Here SVD method was applied to calculate the rotation matrix **R**.

Once **R** is known, (2) is a linear equation in the translation vector. The solution is

$$\mathbf{n}_i^T \mathbf{t} = d_i - \mathbf{n}_i^T \mathbf{R} \mathbf{p}_{i,j} \quad (12)$$

It is easy to show the following fact:

Theorem 2: The necessary and sufficient condition for ***R*** and ***t*** to be uniquely solved is the three planes are not mutually parallel.

IV. EXPERIMENTAL STUDY

A. Simulation Study

In this section, the simulation procedure to illustrate the closed-form approach presented in Section III is described first. Then, the performance of the proposed algorithm in the presence of noise is given.

A.1. Simulation Procedure

The simulation data we need here are three planes represented in the object coordinate frame and two lines on each object plane represented in the camera coordinate frame. In order to get the above data, the following procedure is performed in the simulation.

- First, the normals of two planes, say light planes, were randomly generated in the camera coordinate frame, with the constraint that the two light planes should not be parallel to each other.
- The normals of three planes, say object planes, were then randomly generated in the object coordinate frame, with the constraint that they were not mutually parallel. The distances of the planes to the origin of the object frame were also randomly generated. After this step, $\mathbf{n}_i$ and d_i were obtained.
- The true value of the transformation from the object frame to the camera frame was assumed. It was a rotation around the axis (0.4514, 0.0402, -0.8914) by an angle $\pi/6$ followed by a translation (2, 5, 6).
- The object planes were then transformed with the transformation provided on the previous step from the object frame to the camera frame.
- For each object plane, the two intersection lines were calculated with the two light planes intersecting with the corresponding object plane. After this step, $\mathbf{m}_{i,j}$ were obtained.

- One point from each intersection line was then picked up. There were six of them totally. After this step, $\mathbf{p}_{i,j}$ were obtained.

With the above generated data, the transformation was estimated by applying the algorithm proposed in this paper without any noise presented. It turned out that the estimated pose was exactly the same with the ground truth data. It means that our approach is theoretically correct. Comparing with Chen's algorithm [4], which had to solve the relatively complex polynomial and need remove all the false roots, the proposed algorithm is very easy to implement.

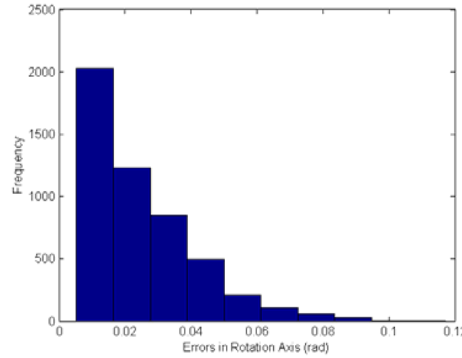
A.2. Sensitivity Study

In order to analyze the performance of the proposed algorithm, a simulation study on the sensitivity of the algorithm to errors in the orientation of line segments are reported here. It was assumed that the line vectors were perturbed randomly. The perturbed line vector located within a cone with the original line vector as the axis. The angle between the axis and the boundary of the cone was defined as the perturbation range (PR). The unit of the PR is degree.

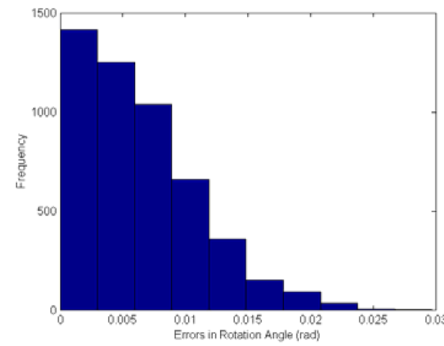
As is well known, the angle between two vectors can be defined as $\cos(\theta) = \mathbf{a} \bullet \mathbf{b}$, where θ is the angle, $\mathbf{a} \bullet \mathbf{b}$ is the inner product of the two vectors. If given the angle and one of the vectors, it is impossible to calculate the perturbed line vector directly, since the line vector has at least two unknowns (the third one can be calculated by using the constraint that the norm of the vector is equal to 1), and there is only one constraint available. Therefore, the following procedure was applied to generate the perturbed line vectors.

- Transform the original line vectors $\mathbf{m}$ to the unit vector along the z-axis of the camera frame. One way of implementing this transformation is to rotate the reference frame by an angle of α about an axis $\mathbf{k}$, where $\alpha \equiv \arccos(\mathbf{z} \bullet \mathbf{m}) = \arccos(\mathbf{m}_z)$ and $\mathbf{k} = \frac{\mathbf{z} \times \mathbf{m}}{\|\mathbf{z} \times \mathbf{m}\|}$ where $\mathbf{z}$ = $[0\ 0\ 1]^T$. The rotation matrix $\mathbf{R}_b$ can then be represented only by the vector $\mathbf{m}$.
- The PR was generated as the Gaussian white noise. Since the perturbed vector along the z-axis should be the unit vector, we can pick one special vector here as $[0\ \sin(\theta)\ \cos(\theta)]^T$.
- Then the perturbed line vector can be obtained by multiply $\mathbf{R}_b$ with the perturbed line vector along z-axis.

Fig. 3 and 4 show the histograms of rotation errors for PR $=1^o$ and 5^o. The results were obtained by running 5000 trials. The error in rotation axis was defined as the angular offset of the rotation axis; the error in rotation angle was simply the absolute difference between the computed rotation angle and the ground truth.

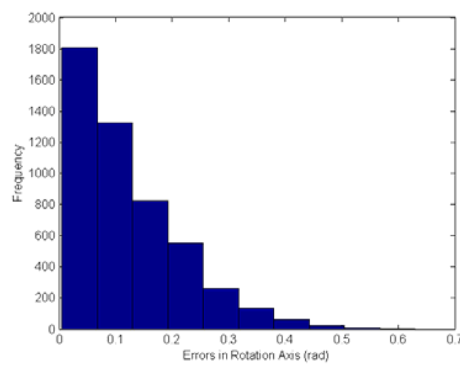


(a) Errors in rotation axis

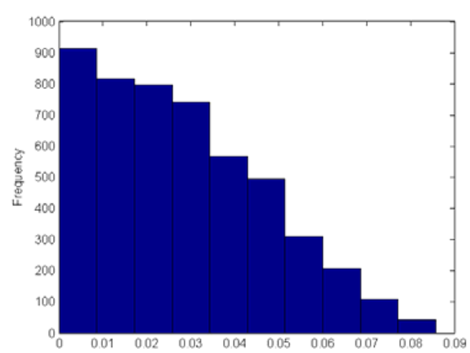


(b) Errors in rotation angle

Fig. 3. Histograms of rotation errors for PR = 1^o



(a) Errors in rotation axis



(b) Errors in rotation angle

Fig. 4. Histograms of rotation errors for PR = 5^o

It is clear that the rotation axis is more sensitive to noise than that of the rotation angle. Comparing Fig. 3 to Fig. 4, it is shown that, as a result of greater noise, the frequency distribution flattens out and more distribution sections shift to the right, and the maximum values of the errors increase. In other words, the probability of errors tends to fall into higher value areas when the noise increases. The number of the trials (in percentage) in which the error of the solution was found to be less than 0.05π (for rotation axis) and 0.01π (for rotation angle) are listed in Table I.

In order to make the comparison of sensitivity performance between our approach and Chen's[4], the errors of rotation axis and rotation angles obtained from both algorithms are listed in Table II. It is obviously that the sensitivity performance of our approach is much better than the other one. On top of it, the computation complex is also much efficient with our approach.

TABLE I
SENSITIVITY RESULTS USING OUR APPROACH

PR (DEG)	$E_{axis} < 0.05\pi$ (%)	$E_{ang} < 0.01\pi$ (%)
1	100	100
2	99.20	96.80
3	92.64	83.92
5	70.86	61.10

TABLE II
COMPARISON OF SENSITIVITY RESULTS

PR (DEG)	$E_{axis} < 0.1\pi$ (%)		$E_{ang} < 0.1\pi$ (%)	
	Chen's	Ours	Chen's	Ours
1	94.5	100	91.1	100
2	91.8	100	85.1	100
3	89.1	99.8	79.1	100
5	83.7	94.9	67.0	100

B. Experimental Results

The experimental setup consists of a LCD projector, a SONY CCD camera of 510 x 492 pixels and a 25-mm lens, and a cube object. Both the projector and the camera are rigidly mounted on one heavy-duty static system.

Before applying our method to the real experiment, the structured light system needs to be calibrated to obtain the geometry relationship between the projector and the camera. The method proposed in [14] has been adopted here for this purpose. After the calibration, the relative position of the projector and the camera should not be changed any more.

Since the proposed approach requires at least two nonparallel strip lines on each object plane, and three such object planes has to be acquired by the camera at one shot, the relative position of the object, the projector and the camera has has to be designed carefully so that the filed of view of the camera is able to cover at least three adjacent planes of the object at the same time while the project can also project the light planes on those three planes. After good relative positions for all three components are configured through trials, their positions should be fixed during the whole experiment. Since the project can only generate one light plane at a time, it has to project non-parallel light plane twice to generate the necessary data. The experimental procedure is described as follows.

- First, the projector shoots one light plan to the object. Camera takes the picture of the object with the light strips on. This is called as image 1.
- Second, the projector projects another light plane, which is not parallel to the first light plane, to the object. The camera takes the picture of the object again with the light strips on. This is called as image 2.
- Third, the camera takes the picture of the object with the light strips off. This is called image 3.

Since the positions of the camera and the object are fixed, the first light strip pattern can be easily acquired by the difference of image 1 and image 3. Similarly the second light strip pattern can be obtained. By using the geometry between the projector and the camera, which has been obtained through the calibration procedure, and the image of light patterns, the 3D line correspondences with respect to the camera can be obtained. Without losing the generality, the origin of the object coordinate frame was defined at the intersecting point of the three mutually orthogonal surfaces of the cube. Therefore, the plane correspondences in the object frame can be easily obtained. With the above acquired data, the proposed algorithm was applied to estimate the pose of the object relative to the camera.

In order to evaluate the experimental results obtained from our proposed approach, the ground truth data of the relative position of the object with respect to the camera is needed. However, in the real world, it is not easy to obtain this ground truth data. Usually there are two wyas, one is using an expensive well-setup measurement system to measure the position data. Another option is using a well-known camera calibration method to estimate the camera's position and orientation with respect to the object. The second approach is adopted here. The camera was calibrated with Zhang' method [15], where the calibration board can be generated by a chess-board paper posting on one object surface. Since the camera calibration results are the pose of the camera relative to the object coordinate frame, the inverse of the transformation matrix of the calibration results were calculated.

Then the estimated transformation matrix from the proposed approach is compared with the one from the camera calibration. The error of the rotation axis is 0.06π and the error of the rotation angle is 0.03π. The translation error is 1.045mm, which is defined as $\left\|\mathbf{t}_{est} - \mathbf{t}_{cal}\right\|$ where $\mathbf{t}_{est}$ is the estimated translation vector of the object relative to the camera using our method and $\mathbf{t}_{cal}$ is the translation vector estimated by the camera calibration. The experimental results show that the performance of our proposed algorithm is very close to the camera calibration one. Another advantage of the proposed algorithm is that the image noise can be reduced to its minimum since only the light strip pattern is needed for the data processing.

V. SUMMARY

A line-to-plane method for object pose determination using a structured light has been presented in this paper. Comparing to the other pose determination method using a structured light system, it can be seen that the closed-form procedure given in the paper is very efficient, compact, and computation time can be ignored in most cases. Furthermore, the necessary and sufficient conditions under which the pose determination problem can be solved are indeed intuitive. This characteristic would make the proposed method suitable for all the real-time applications, such as mobile robot localization and navigation.

Comparing to the stereo-vision-based pose determination method, the structured light system can avoid the challenging correspondence problems between two images. Furthermore, since the structured light sensor is an active device so it will continue to work in dark environments as well as environments in which the objects are featureless, which are the scenario that stereo vision system would fail to work.

Furthermore, extension of the procedure to a linear least squares solution for the case involving measurements from more lines in a plane and more planes is straightforward.

REFERNCES

[1] M. Berger, etc., “3D model based pose determination in real-time: strategies, convergence, accuracy,” *15TH International Conference on Pattern Recognition*, pp. 567-570, Sept. 2000.

[2] D. Skocaj and A. Leonardis, “Robust recognition and pose determination of 3-D objects using range images in eigenspace approach,” Third International Conference on 3-D Digital Imaging and Modeling, pp. 171-178, 2001.

[3] T. Sohmura, M. Nagao, M. Sakai, M. , K. Wakabayashi, T. Kojima, S. Kinuta, T. Nakamura, J. Takahashi, “High-resolution 3-D shape integration of dentition and face measured by new laser scanner,” *IEEE Trans. Medical Imaging*, vol. 23 , no. 5, pp 633 – 638, 2004.

[4] H.H. Chen, “Pose determination from line-to-plane correspondences: existence condition and closed-form solutions,”. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol.13, no. 6, pp. 530-541, 1991.

[5] M. Dhome, M. Richetin, J. T. Lapreste, and G. Rivers, “Determination of the attitude of 3D objects from a single perspective view,” *IEEE Trans. on Pattern Analysis and Machine Interlligence,* vol. 11, no. 12, pp. 1265-1278, 1989.

[6] W. E. L Grimson, and T. Lozano-Perez, “Model-based recognition and localization from sparce range or tactile data,”. *Int. J. Robotics Res.,* vol. 3, no. 3, pp. 3-35, 1984.

[7] S. J. Gordon, and W. P. Seering , “Real-time part position sensing,” *IEEE Trans. On Pattern Analysis and Machine Intelligence,* vol. 10, no. 3, pp. 273-310, 1988.

[8] T. S. Huang, A. N. Netravali, and H. H. Chen, “Motion and pose determination using algebraic methods,”. *Time-Varying Image Processing and Moving Object Recognition, vol.2, V.* Cappellini, Ed. Amsterdam, The Netherlands: Elsevier, pp. 243-249, 1990.

[9] Y. Liu, T. S. Huang, and O. D. Faugeras, “Determination of camera location from 2-D to 3-D line and point correpondences,” *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12, no. 1, pp. 28-37, 1990.

[10] G. Silverman, R. Y. Tsai, and M. Lavin,, “Locating polyhedral objects from edge point,” in *Proc. 1987 Workshop Spatial Reasoning and Multi Sensor Fusion,* St. Charles, IL, pp. 282-292, 1987.

[11] T. Q. Phong, R. Horaud, A. Yassine, and P. D. Tao, “Object pose from 2-D to 3-D point and line correspondences,” *Int. J. of Computer Vision*, vol. 15, pp. 225-243, 1995.

[12] K. Kemmotsu and T. Kanade, "Uncertainty in object pose determination with three light-stripe range measurements," *IEEE Transactions on Robotics and Automation*, vol. 11, no. 5, pp. 741-747, 1995.

[13] D. dCobzas, H. Zhang, and M. Jagersand, "A comparative analysis of geometric and image-based volumetric and intensity data registration algorithms," *Proc. IEEE Int Conf. Robotics and Automation*, pp. 2506 – 2511, 2002.

[14] D.Q. Huynh, “Calibration of a structured light system: a projective approach,” in *Proc. IEEE Conf. Computer Vision, Pattern Recognition*, 1997, pp. 225-230.

[15] Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, November 2000, pp. 1330-1334.