Three Dimensional Structure Estimation and Planning with Vision and Range

Li Zhang[†] and Bijoy K. Ghosh[‡]

Department of Systems Science and Mathematics, Washington University
St. Louis, Missouri 63130, USA

[†]lzhang@zach.wustl.edu, [‡]ghosh@zach.wustl.edu

Abstract

In this paper, a novel 3D structure estimation approach is proposed. The uniqueness of the approach lies in the fusion of vision and 2D range, which is a common sensor combination for mobile robots. 2D range information can be used for hypothesizing 3D structures, feature association, and improving the estimate accuracy. These problems are difficult to solve if using only vision. The proposed approach will be beneficial to 3D map building application of the mobile robots. Experimental results validate the effectiveness of the approach.

1 Introduction

An important and somewhat basic task in mobile robotics is to explore an unknown environment and estimate the surrounding structure, with an eventual goal of making a map of the surroundings and to locate important objects of interest in the map. The process of "map making" and "object recognition and localization" aids subsequent tasks of "Navigation" and "Material Handling" in an unstructured and uncalibrated terrain. Maps that most mobile robots build are two dimensional which are unusable to represent a three dimensional structure in a terrain. In this paper, we propose a new three dimensional structure estimation approach with the ultimate goal of building 3D maps using structures as the basic elements.

Three dimensional structure estimation is a difficult task. Most approaches are vision based: stereo vision [1], [2], structure from shape [3], [4], and structure from motion [5], [6], etc. The effectiveness of these methods rely on an assumption that the structure of interest is known to be a certain type of object. Also, occlusions is hard to deal with. So, these methods are not very effective in mobile robot 3D mapping task.

We consider the application of a 2D laser range finder and a CCD camera mounted on a mobile platform. The range information is encoded in the term of geometric features such as line segments and circular arcs. The camera computes the "perspective projection" of the surrounding structures. We propose using 2D range information to hypothesize, and then together with vision to estimate 3D structures. Emphasis is first placed on estimating the most common 3D feature – planes (Section 2). Then the techniques are extended on more complicated structures (Section 3). All the techniques are implemented in to experiments (Section 4). Conclusion is drawn in Section 5.

2 Plane Estimation

Perhaps the simplest structure to estimate is a plane (planar patch) – a subject that has been studied widely by researchers interested in Machine Vision, [11], [8], [9], [10],. Earlier attempts have been focusing on structure from motion approach, where the motion cues (optical flow) provide information about the shape. The typical problems encountered are that the algorithms are non-recursive and are therefore sensitive to noise. Recursive algorithms based on local linearization methods suffer from the fact that they require good initialization and convergence is not guaranteed. The existence of the plane is assumed and feature association is assumed solved.

We try to eliminate these assumptions and improve the weaknesses by utilizing additional information – 2D range. The laser range finder gives sufficiently accurate measurements. For an arbitrary plane, a range scan measures a 3D line segment lying on it. This information can be used to: 1)hypothesize the presence of a plane, and 2)improve the estimation accuracy. Multiple viewpoints are needed to improve the estimation accuracy. The movement between the viewpoints can be well calibrated by localization techniques developed in [14]. Extended Kalman Filter is used for the estima-

tion by fusing vision and range. The experiment setup is shown in Figure 1.

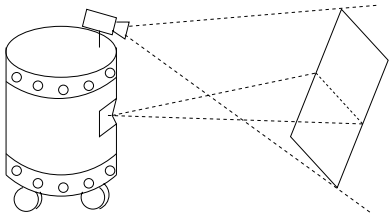

Figure 1: The mobile robot with camera and laser range finder

2.1 Image Dynamics

Due to the viewpoint changes, a feature point on the image creates a optical flow, which includes shape information. In [8], [9], the dynamics is derived for continuous time and constant but unknown motion. Estimating both shape and motion is a very hard problem, and the solution is up to certain degrees of ambiguity. In the map building case, the motion, although varying, can be computed very accurately. Then the plane estimation is simplified.

The plane is estimated in the robot coordinate and is parameterized by p, q, r as in plane equation:

$$pX^r + qY^r + rZ^r + 1 = 0 \tag{2.1}$$

The camera observes feature points on the plane using perspective model in the camera coordinate:

$$x = \frac{X^c}{Z^c}, \quad y = \frac{Y^c}{Z^c} \tag{2.2}$$

The transformation between robot and camera coordinate can be well calibrated in advance and is described by:

$$\mathbf{X}^r = R_{r|c}\mathbf{X}^c + T_{r|c} \tag{2.3}$$

where $\mathbf{X} = [X, Y, Z]^T$.

Suppose the movement from k_{th} view to $(k+1)_{th}$ view is computed to be $[d\theta, dx, dy]$ in k_{th} coordinate, the feature points in the two mobile robot coordinates ($\mathbf{X}^r_k$ and $\mathbf{X}^r_{k+1}$) has the following relationship:

$$\mathbf{X}^r_{k+1} = R^{k+1}_k \mathbf{X}^r_k + T^{k+1}_k \tag{2.4}$$

where R^{k+1}_k and T^{k+1}_k are functions of $[d\theta, dx, dy]$.

A feature point in the camera coordinate will have the following dynamics due to the viewpoint change:

$$\mathbf{X}^c_{k+1} = A\mathbf{X}^c_k + b \tag{2.5}$$

where

$$A = R^T_{r|c} R_{k+1|k} R_{r|c}$$
$$b = R^T_{r|c}(R_{k+1|k} T_{r|c} + T_{k+1|k} - T_{r|c})$$

Then the dynamics of an image point (x, y) can be derived to be:

$$\begin{aligned} x_{k+1} &= \frac{a_{11}x_k + a_{12}y^c_k + a_{13} + b_1/Z^c_k}{a_{31}x_k + a_{32}y_k + a_{33} + b_3/Z^c_k} \\ y_{k+1} &= \frac{a_{21}x_k + a_{22}y^c_k + a_{23} + b_1/Z^c_k}{a_{31}x_k + a_{32}y_k + a_{33} + b_3/Z^c_k} \end{aligned} \tag{2.6}$$

where a_{ij}, b_i, $(i, j = 1 \sim 3)$ are elements of A, b.

From the constraint that the feature point is on the plane to be estimated, we can describe $1/Z^c_k$ by plane parameters and other known data:

$$1/Z^c = \frac{[\ p \quad q \quad r\ ] R_{r|c} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}}{[\ p \quad q \quad r\ ] T_{r|c} + s} \tag{2.7}$$

By substituting $1/Z_c$ into equation (2.7), we can write the image feature dynamics as:

$$\begin{cases} x_{k+1} = f_x(x_k, y_k, p, q, r, s) \\ y_{k+1} = f_y(x_k, y_k, p, q, r, s) \end{cases} \tag{2.8}$$

where f_x, f_y contains the motion and calibration parameters (the details cannot fit in the space here).

In the mobile robot coordinate, the plane has a dynamics due to the mobile robot motion, and is described as the following:

$$\begin{bmatrix} p \\ q \\ r \end{bmatrix}_{k+1} = \frac{R_{k+1|k} \begin{bmatrix} p \\ q \\ r \end{bmatrix}_k}{1 - [\ p \quad q \quad r\ ]_k R^T_{k+1|k} T_{k+1|k}} \tag{2.9}$$

The actual observed image points $(\bar{x}, \bar{y})$ are the observations:

$$\bar{x}_k = x_k + v_x, \quad \bar{y}_k = y_k + v_y \tag{2.10}$$

where v_x, v_y are noises for the image measurements.

The states for the plane estimation EKF are plane parameters and all image points:

$$[p, q, r, x^1, y^1, \cdots, x^n, y^n] \tag{2.11}$$

The state dynamics are from (2.9) and (2.8). The observations are from (2.10) with prefixed variances.

The initial states can be calculated by observations from two viewpoints. We have developed a method that utilizes the 2D range line segment to better make feature association and compute the initial plane parameters [15]. Also, the range information will be integrated into the EKF later.

2.2 Convergence Anaylysis

There is no guarantee that the EKF method will converge. In general, the convergence depends on the observability of the system and the accuracy of the initial value. It is better to have some guidelines that show under what condition(s) the EKF for the plane estimation is more likely to converge than under the others.

The plane estimation is actually a structure from motion problem, in which the choice of motion affects the estimation quality (thus the convergence of the EKF). Intuitively, what determines the quality of the estimation is the noise to signal ratio on image measurements. The signal can be referred to the effectiveness of the viewpoint change for stereo triangulation. More effective viewpoint change always results in stronger stereo information. But this has never been quantitatively verified.

Since the plane is static, it can also be solved algebraically. We can compute the error bound of the solution under different ways of viewpoint changes (different robot motions). The plane parameters can be seperated from the rest of the information in image dynamics (2.8) by rearranging all the items:

$$\begin{aligned} 0 &= \begin{bmatrix} f_p & f_q & f_r & f_1 \\ g_p & g_q & g_r & g_1 \end{bmatrix} \begin{bmatrix} p \\ q \\ r \\ 1 \end{bmatrix}_k \\ &= \begin{bmatrix} f_p & f_q & f_r & f_1 \\ g_p & g_q & g_r & g_1 \end{bmatrix} \begin{bmatrix} R_{0|k}^T & 0 \\ T_{0|k}^T & 1 \end{bmatrix} \begin{bmatrix} p \\ q \\ r \\ 1 \end{bmatrix}_0 \\ &\triangleq \begin{bmatrix} U_k & -V_k \end{bmatrix} \begin{bmatrix} p \\ q \\ r \\ 1 \end{bmatrix}_0 \\ \Longrightarrow \quad & U_k X = V_k \end{aligned} \tag{2.12}$$

where f's and g's are the coefficients of (p, q, r, s) after the rearrangement. U_k is 2×3 matrix and V_k is 2×1 vector. Subscript 0 indicates the world coordinate.

Equation (equ:staticplaneest) is based on the observation of one image point from two viewpoints. If we have at least three linear independent feature points, by stacking up U_k and V_k, $[p, q, r]_0$ can be obtained by solving equation $U^T U X = U^T V$ (denote as $AX = B$). For more viewpoints, we can keep stack these equation up and obtain an over-constrained system.

Matrix U and vector V are functions of the image points and the mobile robot motion parameters. We can study the quality of the solution by using the results from numerical analysis. For equation $(A + dA)(X + dX) = B + dB$, the relative error on solution X, $||dX||/||X||$, is bounded by μ:

$$\mu = \frac{cond(A)}{1 - cond(A)\frac{||dA||}{||A||}} \left(\frac{||dB||}{||B||} + \frac{||dA||}{||A||}\right) \tag{2.13}$$

with the assumption that

$$1 - cond(A)\frac{||dA||}{||A||} > 0 \tag{2.14}$$

Simulations can be carried out to check μ for the plane estimation problem under various motions. dA, dB come from the image noise. To eliminate the dependence on a specific set of the random noise, the simulation runs 100 times for each type of motion and the error bounds μ are averaged.

We checked three types of motions: moving sideways (1), moving toward (2) and away from (3) the plane (Figure 2). For each motion, we compute the μ for single step (using two views) and overall steps.

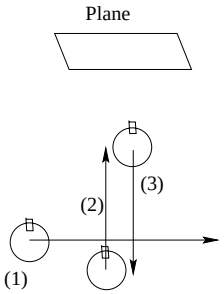

Figure 2: Three types of motions (top view)

Figure 3 and 4 show the error bound μ of both single step (solid) and overall steps (dotted) for different motions. Figure 3 shows results for motion 1. Left side shows the plot for 0.1m per step while right side for 0.05m. It is obvious that bigger movements (thus bigger image movements) have smaller error bounds (check the amplitudes). Also, image movements is bigger when right in front of the plane. That is reason there is a valley in the middle. Figure 4 is for motion 2 and 3. The error bounds are smaller when the mobile robot is closer so that the image movements are bigger. We also have run the EKF estimation algorithm. The convergence rates coincide with the error bounds. That is, the motions that renders smaller error bound always yield better EKF estimations.

The error bound is actually too conservative to be meaningful bound (see the magnitude of the plots). But it shows relatively the estimation quality verse different motions. So, the method provides a way to quantitatively check the estimation quality based on different motions. From the simulation, we know that better movements are usually moving sideways with big enough distance.

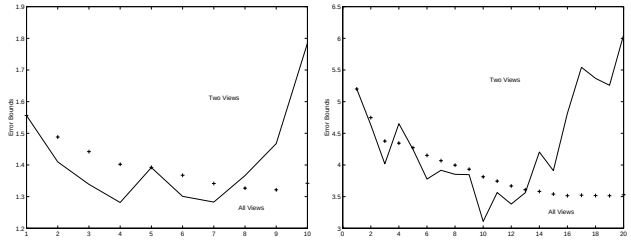

Figure 3: Motion 1. Left:0.1m step. Right:0.05m step.

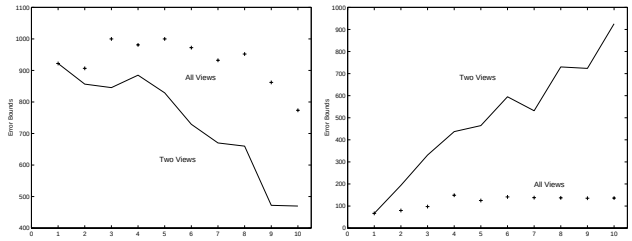

Figure 4: Moving 2(left) and 3(right), 0.1m step

2.3 Integrating range

The laser range finder can observe a line segment on the plane of interest (fitted from range points). This constraint can be a very good candidate to form an extra observation for the plane estimation EKF. The range line segment lies in the horizontal scanning plane and can be modeled using COG description [14]. It can be parameterized by the intersection of the scanning plane and the vertical plane passing though the range line segment, which can be written in a matrix form:

$$0 = \begin{bmatrix} 0 & 0 & 1 & -h_{lsr} \\ c1 & c2 & 0 & c3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \triangleq L \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (2.15)$$

where h_{lsr} is the height of the scanning plane, and c_i's are from the COG line description.

Since the plane to be estimated passes the range line, its parameter $[p,\ q,\ r,\ 1]^T$ must be the linear combination of row vectors of matrix L, and thus perpendicular to the kernel of L. Let K be the kernel of L. Then,

$$0 = K \begin{bmatrix} p & q & r & 1 \end{bmatrix}^T \quad (2.16)$$

This constraint is an implicit observation for the plane estimation EKF, which is dealt with by linear approximation [12]. We will show in the experimental results that range information greatly improves the plane estimation.

3 Box Estimation

Most 3D structures in the indoor environments consist of planes. Although it is impossible to parameterize a complex 3D structure using simple equations like plane equation, a set of parameters can be chosen so that they can define all the planes on the structure. The 3D structure can then be estimated by estimating these planes using the EKF plane estimation techniques.

We use cubic shape boxes as the structure of interest. It is a relatively simple 3D structure yet contains all the basic ingredients for common 3D structures. Using it we can get good understanding of the problem.

3.1 Modeling Boxes

We are able to make the models more general by using range and vision. Here, models are built so that each model describes a class of 3D structures that have similar structure in terms of planes. The model of a box includes dimensional parameters (length, width, height), and the geometric relations between the planes. We are also interested in the pose of the box, which consists of 6 parameters, three for orientation angles (tilt: α, yaw:β, pan:γ) and three for a reference point.

The reference point is chosen from range measurements which are used to hypothesize the box. The intersection of the two connected line segments indicates a point on one edge of the box. Since the range measurement is very accurate, this point is a good reference point.

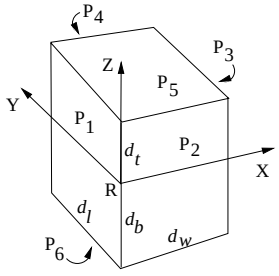

Figure 5: Box modeling

A local box coordinate can be chosen to be aligned with the normal vectors of the box planes (Figure 5). d_w and d_l are the width and length. The reference point splits the height into two parts: d_t and d_b. The six planes are numbered based on their position with respect to the reference point. The orientation of the box is between the box and the world coordinate. There are totally seven parameters to estimate after the reference point is found: $\alpha, \beta, \gamma, d_l, d_w, d_t, d_b$.

In the box coordinate, the parameters of all six planes $P_i^b, i = 1 \sim 6$ can be expressed by box parameters

d_l, d_w, d_t, d_b, using $P_i^b = [p, q, r, s]$ homogeneous parameters as in $px + qy + rz + s = 0$ plane description.

$$\begin{array}{ll} P_1^b = [1\ 0\ 0\ 0]; & P_2^b = [0\ 1\ 0\ 0]; \\ P_3^b = [1\ 0\ 0\ -d_l]; & P_4^b = [0\ 1\ 0\ -d_w]; \\ P_5^b = [0\ 0\ 1\ -d_t]; & P_6^b = [0\ 0\ 1\ d_b]; \end{array} \quad (3.17)$$

To utilize the plane estimation techniques, all six planes on the box are transformed into the robot coordinate. The reference point $T_{r|b}$ and the rotation matrix $R_{r|b}$ between box and robot coordinate can be computed based on box parameters. Six plane parameters in the mobile robot coordinate $P_i^r,\ i = 1 \sim 6$ are computed by:

$$P_i^r = P_i^b \begin{bmatrix} R'_{r|b} & -R'_{r|b} T_{r|b} \\ 0 & 1 \end{bmatrix} \quad (3.18)$$

3.2 EKF estimation of the box

It is then straight forward to use the plane estimation techniques to estimate these box parameters. The states consist of the seven box parameters plus all image points which are associated with some box planes:

$$X = [\alpha, \beta, \gamma, d_l, d_w, d_t, d_b, x_1, y_1, \cdots, x_n, y_n] \quad (3.19)$$

To get the image point dynamics, the plane parameters p, q, r, s (3.18) of each visible plane are substituted into the image point dynamics (2.8) in plane estimation:

$$\begin{cases} x_{k+1} = f_x(x_k, y_k, p(\psi), q(\psi), r(\psi), s(\psi)) \\ y_{k+1} = f_y(x_k, y_k, p(\psi), q(\psi), r(\psi), s(\psi)) \end{cases} \quad (3.20)$$

where ψ stands for all the box parameters.

The camera can see at most three planes on the box at one viewpoint. To estimate more parameters, the mobile robot needs to go around the box. During the process, some planes are appearing and others disappearing. Since image points are the states, we need to deal with appearing and disappearing states.

To better illustrate this issue, some of the states and their variance matrix are shown below:

$$X = \begin{bmatrix} x_b \\ \vdots \\ x_i \\ \vdots \\ x_j \end{bmatrix}; \quad P = \begin{bmatrix} V_b & \cdots & V_{bi} & \cdots & V_{bj} \\ \vdots & & \vdots & & \vdots \\ V_{ib} & \cdots & V_i & \cdots & V_{ij} \\ \vdots & & \vdots & & \vdots \\ V_{jb} & \cdots & V_{ji} & \cdots & V_j \end{bmatrix} \quad (3.21)$$

where b stands for the box parameters, i and j for the i_{th} and j_{th} image points. V's are the variances and covariances for the corresponding states.

If image point j disappears, x_j is discarded together with all related items in P (on the same row and column with V_j). This keeps the state space from growing to a huge one after observing more and more feature points. When a new feature point (i) appears, it is added directly to X together with its V_i added into P. Its covariance with other states are set to zero.

The observations of the box estimation EKF consist of all the observed image points and the range line segments. The range line segments are associated with certain planes of the box first. Then the plane parameters p, q, r, s (3.18) are substituted into (2.16):

$$0 = K \begin{bmatrix} p(\psi) & q(\psi) & r(\psi) & s(\psi) \end{bmatrix}^T + v_L \quad (3.22)$$

where ψ stands for the box parameters.

The initial box estimation can also be calculated based on the initial plane estimation technique [15].

4 Experiments

4.1 Plane Estimation

The EKF plane estimation techniques has been implemented on the mobile robot Normadic XR4000. The mobile robot moves on both x and y direction and rotates to estimate a tilted plane. Since image processing is not a focus of this research, clear feature points are put on the planes to make the image processing easier. Established image processing and stereo vision computation can be applied easily, especially with the help of 2D range.

Two plane estimation results are shown in figure 6. One shows the results of using both vision and range information (solid curves), and the other one uses vision only (dashed curves). The thick curves indicate the actual parameter values. Normalized $[p, q, r]$ and the center of feature points are used instead to describe the plane for better comparison. It is obvious that the integration of the range greatly improves the estimation.

4.2 Box Estimation

An experiment has been carried out to test the proposed box estimation method. The experimental setup picture is shown in figure 7. The mobile robot circles the box about a half circle to see all the visible faces. Initial estimate of the box parameters can be computed based on techniques developed in [15].

The estimation results are shown in figure 8. The estimated seven parameters are drawn (as 'x') against the actual value (in solid line). All box parameters are well estimated during the estimation process, except d_b (height to the bottom), which is unobservable.

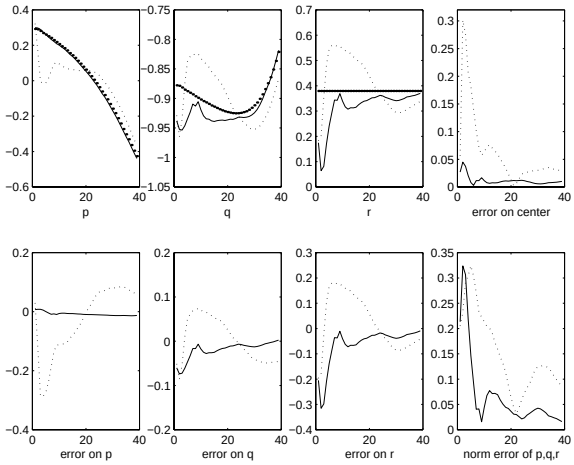

Figure 6: Plane estimation experimental results

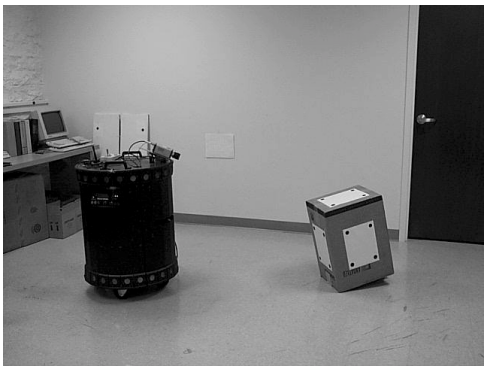

Figure 7: The setup in box estimation experiment

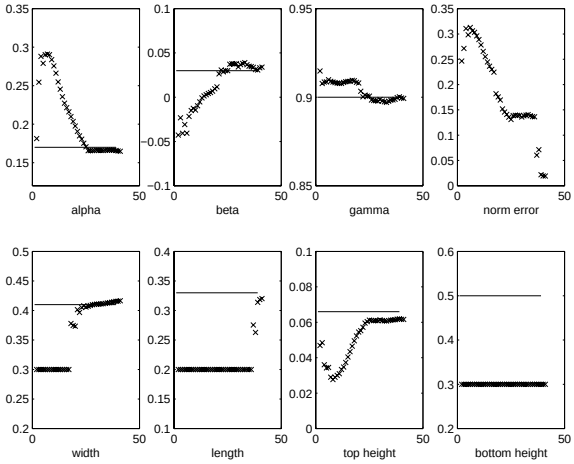

Figure 8: Box estimation experimental result

5 Conclusions

A novel 3D structure estimation approach has been established. It effectively integrates vision and 2D range, which are commonly available on mobile robots, to estimate 3D structures. The 2D range plays various roles in the scheme: localization, 3D hypothesizing, improving the estimation accuracy. Some difficult issues in vision-only approaches, including feature corresponding and association, can be solved with the help of range. Experimental results validate the effectiveness of the approach. The proposed approach is part of our 3D map building research.

References

[1] Y. Shirai, *Three dimensional computer vision*, Springer-Verlag, 1987.

[2] Z. Zhang, O. Faugeras, A 3D world model builder with a mobile robot, *Int. j. of robotics research*, vol. 11, no. 4, pp. 269 – 285, 1992.

[3] B. Horn, M. Brooks, *Shape from shading*, The MIT Press, 1989

[4] R. Zhang, P. S. Tsai, J. E. Cryer, M. Shah, Shape from shading: a survey, *IEEE Trans. On PAMI*, vol. 21, no. 8, pp. 690, 1999.

[5] S. Ullman, *The interpretation of visual motion*, The MIT Press, Cambridge, MA, 1979

[6] C. J. Taylor and D. J. Kriegman, Structure and motion from line segments in Multiple images, *IEEE Trans. On PAMI*, vol. 17, no. 11 pp. 1021, 1995.

[7] M. D. Wheeler and K. Ikeuchi, Sensor modeling, probabilistic hypothesis generation and robust localization for object recognition, *IEEE Trans. On PAMI*, vol. 21, no. 1, 1999.

[8] B. K. Ghosh, M. Jankovic, Y. T. Wu, Perspective problems in system theory and its application to machine vision, *J. of Math. Sys., Est. and Contr.*, vol. 4, no. 1, pp. 3-38, 1994.

[9] B. K. Ghosh, E. P. Loucks, A perspective theory for motion and shape estimation in machine vision, *SIAM J. on Contr. and Optim.*, vol. 33, No. 5, pp. 1530-1559, 1995.

[10] B. K. Ghosh, H. Ianaba, S. Takahashi, "Identification of Riccati dynamics under perspective and orthographic observations", *IEEE Trans. on Autom. Contr.*, Sept. 2000, to appear.

[11] K. Kanatani, *Group-Theoretical Methods in Image Understanding*, Spring-Verlag, 1990.

[12] S. Soatto, R. Frezza, P. Perona, "Motion estimation via dynamic vision", *IEEE Trans. on Autom. Contr.*, vol. 41, no. 3, pp. 393-413, 1996.

[13] S. Umeyama, "Least-Squares estimation of transformation parameters between two point patterns", *IEEE Trans. on PAMI*, vol. 13, no. 4, pp. 376-380, Apr. 1991.

[14] L. Zhang, B. K. Ghosh, "Line Segment Based Map Building and Localization Using 2D Laser Rangefinder", *Proc. of ICRA 2000*, pp. 2538-2543, 2000

[15] L. Zhang, *Map Building, Localization, and Structure Estimation Problems in Mobile Robotics*, Ph.D Thesis, Washington Univ., Sept. 2000.