Proceedings of 2004 IEEE/RSJ International Conference on
Intelligent Robots and Systems
September 28 - October 2, 2004, Sendai, Japan

SLAM in Indoor Environments with Stereo Vision

S. Takezawa
Mechatronics Lab in
Mechanical Engineering Systems
Faculty of Engineering
Hokkaido Institute of Technology
Sapporo, Japan
Email: takezawa@hit.ac.jp

D. C. Herath
ARC Centre of Excellence in
Autonomous Systems(CAS)
Faculty of Engineering
University of Technology, Sydney
Broadway, NSW, Australia
Email: damith.herath@eng.uts.edu.au

G. Dissanayake
ARC Centre of Excellence in
Autonomous Systems(CAS)
Faculty of Engineering
University of Technology, Sydney
Broadway, NSW, Australia
Email: gdissa@eng.uts.edu.au

Abstract—**This paper proposes a method for simultaneous localisation and mapping (SLAM) in an indoor environment using stereo vision. Specially designed artificial landmarks distributed in the environment are observed and extracted from a camera image. The disparity map obtained from the stereo vision system is used to obtain the ranges to these landmarks. The main contribution of the paper is the formulation of the mathematical framework for SLAM for a robot moving on a planar surface among landmarks distributed in three dimensional space. The paper also presents the results of experiments conducted using a Pioneer robot and a Triclops stereo vision system. It is demonstrated that accurate robot and feature locations can be obtained using the proposed technique.**

I. Introduction

The general SLAM problem has been the subject of substantial research in the past few years [1]. SLAM using vision is becoming more and more important due to the recent developments in image processing. Therefore, vision based robot navigation has attracted significant attention [2], [3] and [4].

Current stereo vision systems can provide depth information from a scene at frame-rate. The disparity map provided by a stereo system can be used to determine range, bearing and elevation to point features in the environment. Although extracting feature points from a given scene image are a complicated process in its own right, utilizing this knowledge is able to infer the feature locations with respect to a three dimensional world coordinate system and in turn using such feature points to localize the robot itself in this coordinate system is nothing less of a challenge [5].

In case of vision based SLAM, the challenge is to provide a consistent map building method that allows the unknown coordinates of features in the environment together with the coordinates of the robot. In this paper we provide a techniques to achieve this and provide some experimental results to verify the proposed algorithms.

We improve on the estimation process based on the extended Kalman filter (EKF) by extending it to accommodate a three-dimensional world coordinate model. This paper extends the "the process model" module and "the observation model" from 2D to 3D. The other contribution made in this paper to the three-dimensional SLAM problem is the use of disparity maps as a means ofextracting spatial information of identified feature positions.

II. Formulation

As the robot with a known kinematic model starts at an unknown location and moves through an environment containing a number of features or landmarks, we must provide the procedure how we know and estimate of the robot position and landmark locations in Cartesian coordinates.

A. Discrete Robot and Landmark Models

We elaborate a discrete time index in this section replacing the continuous time index. The absolute locations of the landmarks are not available. Without prejudice, a linear (synchronous) discrete time model of the evolution of the robot state and the observations of landmarks is adopted. Although robot motion and the observation of landmarks is almost always nonlinear and asynchronous in any real navigation problem, the use of linear synchronous models does not affect the validity of the proofs.

We require the same linearization assumptions as those normally employed in the development of an extended Kalman filter. In the following, the robot state is defined by $X_r = [x_r, y_r, z_r, \varphi_r]^T$ where x_r, y_r and z_r denotes the location of robot on the coordinates of the centre at the rear axle of the robot with respect to some global coordinate frame and φ_r is the heading with reference to the x-axis. The landmarks are modelled as point landmarks $P_i = [x_i, y_i, z_i]^T$, $i = 1 \ldots N$ and represented by Cartesian co-ordinates in Fig.1.

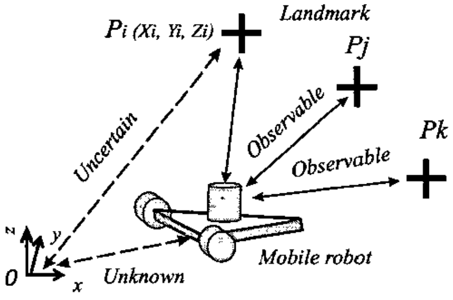


Fig. 1. Uncertain environment for robot

0-7803-8463-6/04/$20.00 ©2004 IEEE

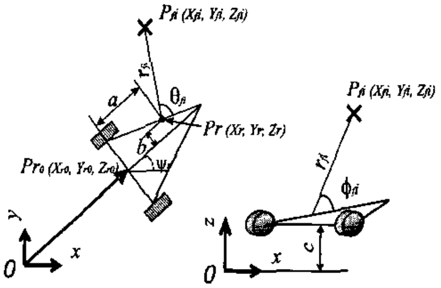


Fig. 2. Configuration for model

Fig. 3. Triclops and Pioneer

Figure 2 shows a schematic observation kinematics diagram of the robot in the process of observing one of landmarks. The following kinematic equations can be used to predict the robot state from rear wheel velocity inputs V

$$\begin{bmatrix} \dot{x}_r \\ \dot{y}_r \\ \dot{\varphi}_r \end{bmatrix} = \begin{bmatrix} V\cos(\varphi_r) \\ V\sin(\varphi_r) \\ \omega \end{bmatrix} \tag{1}$$

where V, ω denote the velocity and of the angular velocity of the robot respectively. These equations can be used to obtain a discrete time robot process model in the form

$$\begin{bmatrix} x_r(k+1) \\ y_r(k+1) \\ \varphi_r(k+1) \end{bmatrix} = \begin{bmatrix} x_r(k) + \Delta TV(k)\cos(\varphi_r(k)) \\ y_r(k) + \Delta TV(k)\sin(\varphi_r(k)) \\ \varphi_r(k) + \Delta T\omega(k) \end{bmatrix} \tag{2}$$

for use in the prediction stage of the robot state estimator. ΔT is the time step. The landmarks in the environment are assumed to be stationary point targets. The landmark process model is thus

$$\begin{bmatrix} x_i(k+1) \\ y_i(k+1) \\ z_i(k+1) \end{bmatrix} = \begin{bmatrix} x_i(k) \\ y_i(k) \\ z_i(k) \end{bmatrix} \tag{3}$$

for all landmarks $i = 1 \dots N$. Together, Eq.(2) and Eq.(3) define the state transition matrix f for the process model.

Regarding to an observation at time instant k from the sensor, the location of the landmark possibly responsible for this observation $P_{fi} = [x_{fi}, y_{fi}, z_{fi}]^T = g(x_r, y_r, \varphi_r, r_{fi}, \theta_{fi}, \phi_{fi})$ and its covariance P_{fi} is calculated using the following relationship.

$$\begin{bmatrix} x_{fi} \\ y_{fi} \\ z_{fi} \end{bmatrix} = \begin{bmatrix} x_r(k) \\ y_r(k) \\ z_r(k) \end{bmatrix} + \begin{bmatrix} x_{rf}(k) \\ y_{rf}(k) \\ z_{rf}(k) \end{bmatrix} \tag{4}$$

where

$$\begin{bmatrix} x_{rf}(k) \\ y_{rf}(k) \\ z_{rf}(k) \end{bmatrix} = \begin{bmatrix} r_{fi}\cos(\phi_{fi})\cos(\varphi_r(k) + \theta_{fi}) \\ r_{fi}\cos(\phi_{fi})\sin(\varphi_r(k) + \theta_{fi}) \\ r_{fi}\sin(\phi_{fi}) \end{bmatrix} \tag{5}$$

and

$$P_{fi} = \nabla g_{xy\varphi} P_r \nabla g_{xy\varphi}^T + \nabla g_{r_f\theta_f\phi_f} R \nabla g_{r_f\theta_f\phi_f}^T \tag{6}$$

where P_r is the covariance matrix of the robot location estimate extracted from the state covariance matrix $P(k|k)$ and R is the measurement noise covariance. ∇g is the Jacobian of g. The update of the state estimate covariance matrix is of paramount importance to the SLAM problem. We consider the SLAM problem is to estimate the aggregated robot and landmark locations, in the form of a state vector, provided with the measurement. The state vector is given as

$$X = [X_r, P_{fi}]^T = [x_r, y_r, \phi_r, x_{fi}, y_{fi}, z_{fi}]^T \tag{7}$$

Both robot and landmark states are registered in the same frame of reference. Understanding the structure and evolution of the state covariance matrix are the key components to the solution of the SLAM problem.

B. Nonlinear Observation Models

The sensor signals used in the simulations returns the range $r_{fi}(k)$, bearing $\theta_{fi}(k)$ and bearing $\phi_{fi}(k)$ to a landmark i. Referring to Fig.2, the observation model can be written as

$$\begin{cases} r_{fi}(k) = R_{fi}(k) + w_r(k) \\ \theta_{fi}(k) = \arctan\left(\dfrac{y_{fi} - y_r(k)}{x_{fi} - x_r(k)}\right) - \varphi_r(k) + w_\theta(k) \\ \phi_{fi}(k) = \arcsin\left(\dfrac{z_{fi} - z_r(k)}{R_{fi}(k)}\right) + w_\phi(k) \end{cases} \tag{8}$$

where $R_{fi}(k) = \sqrt{(x_{fi} - x_r(k))^2 + (y_{fi} - y_r(k))^2 + (z_{fi})^2}$ is the Euclidean distance between robot and a landmark,w_r, w_θ and w_ϕ are the noise sequences associated with the range, bearing and elevation measurements. Vector $X_r = [x_r(k), y_r(k), z_r(k)]^T$ is given by laser sensor or odometer accumulation in global coordinates generally.

III. Triclops Stereo System

Figure 3 shows the test mobile robot, Pioneer, as well as the trinocular stereo vision camera, Tricrops which is located on the top of the robot. The camera module simultaneously obtains three images of the environment. While the appearance of images are quite similar, closer inspection reveals a shift between closer objects and those that are further away. Based on the amount of shift, the system is able to determine the distance to the objects in the image. If the calibration of camera is corrected perfectly, closer objects are represented with brighter shades grey

TABLE I

STEREO CAMERA SPECIFICATION

Image Sensor	1/3' CCD Progressive scanning
Effective Pixels	640x480 VGA format
Focal Length	3.8mm or 6.0mm
Baseline	100mm
Frame Rate	24Hz at 640x480 each camera
Gain Control	-3dB to 33dB
Shutter Speed	1/25s - 1/15000s dB
Size	15.5 x 15.5 x 5.0cm
Mass	500g

TABLE II

ROBOT SPECIFICATION

parameter	values	unit
x-velocity	0.02	m/s
angular velocity	0.00	rad/s
radar analysys	0.01	msec
bearing angle	$-\pi/4 < \theta < \pi/4$	rad
elavation angle	$-\pi/4 < \theta < \pi/4$	rad

while objects further away are represented with darker shades of grey. Main camera module and mobile robot specifications are described in Table I and Table II. The robot is moved at a constant velocity along a straight line and a sequence of depth maps are collected. In the evaluation of the 3D SLAM algorithm, this information is employed to deduce estimation for both robot position and landmark locations. The camera coordinate frames are co-planar and aligned so that the epipolar lines of the camera pairs lie along the rows and columns of the images.

As shown in Fig.4, the X, Y and Z position of a point in the scene can be determined through triangulation between points in the images, P_l, P_c and P_t, obtained by the left, center and top camera respectively. B_h and B_v are horizontal and vertical baseline displacement between the camera respectively. f denotes a focal length and (u_0, v_0) denotes a principal point. Regarding to X-Y coordinate, a pare of rectified image points (X_l, Y_l) on left side camera and (X_c, Y_c) on center camera are two corresponding points in that image pair.

TABLE III

LANDMARK AND BEACON POSITION

LMK No. for Vision	x[m]	y[m]	z[m]
1	1.40	0.50	0.30
2	1.40	-0.30	-0.30
3	2.80	0.00	0.00
BCN No. for Laser	x[m]	y[m]	z[m]
1	3.00	0.30	0.00
2	3.00	-0.60	0.00

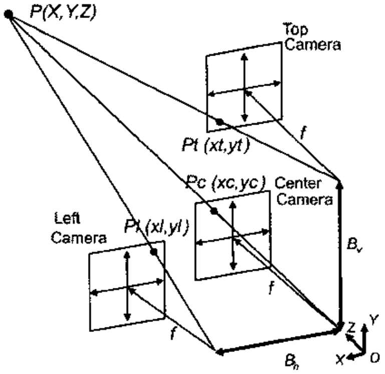


Fig. 4. Camera layout

The calculating dense disparity map from stereo images has been studied for long time and is popular in computer vision. Essence of the procedure adopted is described below. Since the corresponding points must line on epipolar lines, the relation between the two points is

$$\begin{cases} x_l = x_c - d_h \\ y_l = y_c \end{cases}$$

where d_h is defined as disparity of the point (X_c, Y_c) As the same way, in the vertical direction

$$\begin{cases} x_t = x_c \\ y_t = y_c - d_v \end{cases}$$

where d_v is defined as vertical disparity of the point (x_c, y_c). From above definition, d denotes the average of the d_h and d_v After replacing (x_c, y_c) to (x, y), we define $(\bar{x}, \bar{y}) = (x - u_0, y - v_0)$ the centered image point coordinate of (X, Y). Let (X, Y, Z) be the Euclidean coordinate of the 3D point M corresponding to a point in the reference image. The relation between (X, Y, Z) and $(\bar{x}, \bar{y}, d)$ is

$$\begin{cases} \bar{x} = x - u_0 = f\frac{X}{Z} \\ \bar{y} = y - v_0 = f\frac{Y}{Z} \\ d = \frac{fB}{Z} \end{cases} \tag{9}$$

where B equals B_h and B_v. Equation 9 is widely used in the vision literature for estimating (X, Y, Z) from $(\bar{X}, \bar{Y}, d)$. In our study, we obtain $(\bar{X}, \bar{Y}, d)$ from experiment through image segmentation and foreground background detection.

Relative positions of the Laser Beacons (BCN) for the laser based localizer and the artificial landmarks (LMK) for the 3D vision based SLAM are shown in Table III. From the BCN position data, the location of a mobile robot $X_r = [x_r(k), y_r(k), z_r(k)]^T$ is detected and utilized to measure and observe the range, bearing and elevation between a robot and landmarks.

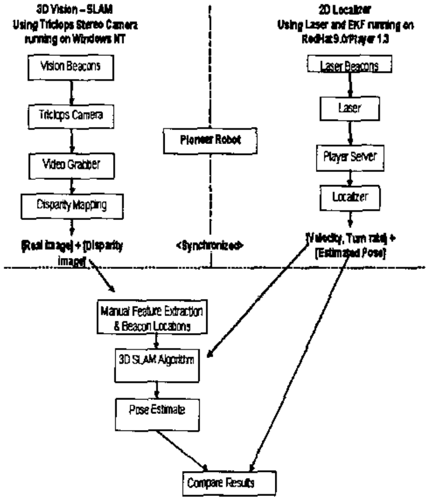
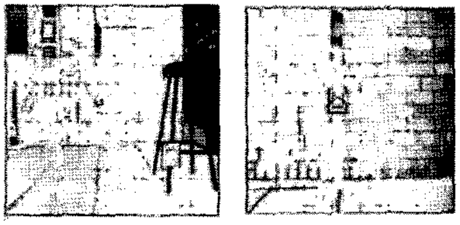
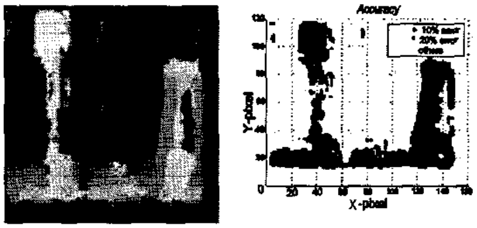


Fig. 5. SLAM experimental flow



(a) Initial position (b) Final position

Fig. 6. Rectified image



(a) Initial position (b) Disparity noise

Fig. 7. Disparity image

TABLE IV
CAMERA CALIBRATION

unit [m]	holizontal	vertical
FocalLength	0.886054	1.18036
BaseLine	0.0995951	0.0996837

IV. Experimental Results

The experiments were conducted using the Pioneer robot and the Triclops vision system. Textured landmarks similar to those proposed in [6] were used in the experiments to minimise the effect of image processing issues. However, the general framework proposed would be applicable to point natural landmarks extracted using an algorithms such as KLT. The robot starts at an unknown location with no knowledge of the location of landmarks in the environment. As the robot moves through the environment, it makes relative observations of the location of individual landmarks. A laser sensor installed on the robot together with reflective beacons placed in the environment at known locations were used to compute the precise location of the robot so that the outcome of the 3D SLAM can be appropriately evaluated. A standard localisation algorithms based on the EKF was used for this purpose [7].

A laser sensor installed on the robot together with reflective beacons placed in the environment at known locations were used to compute the precise location of the robot so that the outcome of the 3D SLAM can be appropriately evaluated. A standard localisation algorithms based on the EKF was used for this purpose.

We obtain discrete vision data at each half-second period. The experimental arrangement for the 3D SLAM is shown in Fig.5. From the 2D Localizer algorithm, we can compare and evaluate the location of a stationary $P_{fi} = [x_{fi}, y_{fi}]^T$ without z_{fi} in order to check and design our experimental methodology.

A. Feature Detection

Figure 6(a) shows the rectified initial position image from the reference camera while the test robot is moving in an environment that contains several laser reflectors to aid the laser based localization and three artificial landmarks for estimation by stereo vision. Figure 6(b) shows the rectified final position image from the reference camera mounted on the robot. It has 160x120 resolution.

Figure 7(a) shows disparity image data. This image was created with algorithms using Triclops 2.1.4 library. The disparity results are validated in two ways. First, we have done a "sufficient texture" test. This test checks that there is sufficient variation in the image patch that is to be correlated by examining the local sum of the Laplacian of Gaussian of the image. Secondly, we achieved a "quality of match" test. In this test, the value of the score is normalized by the sum of all scores for this pixel. If the result is not below a threshold, the match is considered to be insufficiently unique and therefore a likely mismatch. This kind of failure generally occurs in occluded regions where the pixel cannot be properly matched [8].

Figure 7(b) shows the disparity error of the image at the initial position. The noise associated with $\bar{x}$ and $\bar{y}$ is due to the image disceritization. Without any a priori information, the variances $\sigma_{\bar{x}}$ and $\sigma_{\bar{y}}$ of this noise are the same for all image points. Using camera calibration parameters in Table IV, the range of values for disparity variation from 31 to 63 relates to 2.81 m to 1.39 m.

In this figure, the error of disparity including noise and measurement position data are presented depending on differences of error percentages. A manually calculated depth value for each landmark at each frame was used to compute the error distribution in each disparity image. Our own experiments indicate that the Triclops stereo vision module produces results with standard deviations well under expected values. Simultaneously, we recognized the value of depth especially was unstable and unreliable if the range to the landmark beyond about 4 m.

Figure 8 shows the state estimates of two landmarks as the robot moves towards them. Ideally the landmarks would move towards the edges of the image diagonally since the robot moves in a straight line towards them. The upper left corner data presents landmark 1 and the remaining plots landmark 2. The observation line includes the motion of relative angle and transition error. In this case, rotation angle equals zero because of motion of robot is straight forward along Z axis. The estimation line is calculated by EKF. This figure shows that the proposed algorithms is accurate [9].

B. Robot Localization

Figure 9(a) shows the robot and landmark locations after one sampling step. The sampling rate of the estimator was 2 Hz, while the images were captured once every second. The ellipses show the error covariances and '+' and '*' indicate the true landmark locations while circles are used to show the estimated landmark locations.

Figure 9(b) shows the estimated robot and and landmark locations at the end of the experiment. The path of the robot is also shown. The error covariance of the estimated landmark locations are smaller then those at the beginning, as expected. This is generally true for SLAM as for any diagonal element of the map covariance matrix,

$$\sigma_{ii}^2(k+1|k+1) \leq \sigma_{ii}^2(k|k) \tag{10}$$

Thus the error in the estimate of the absolute location of every landmark also diminishes.

Figure 10 shows the true and estimated path of the robot. We note here that the true path was estimated with an EKF that uses information from the laser range finder and reflective beacons placed in the environment.

Figure 11 shows the error between true and SLAM estimated position of the robot in details. This figure shows the actual error in estimated robot location as a function of time (the central solid line). The figure also shows 95% (2σ) confidence limits in the estimate error. These confidence bounds are derived from the state estimate covariance matrix and it is clear that the estimator is consistent as the error is within the confidence limits.

The innovations in range, bearing and elevation observations are shown in Fig.12 and together with the estimated 95% (2σ) confidence limits. The innovations are the only available measure for analysing on-line filter behaviour when true robot states are unavailable. The innovations here also indicate that the filter and models employed are consistent.

Figure 13 shows the error between the actual and estimated landmark locations for two of the landmarks. One of these landmarks (landmark 1) was observed from the initial robot location at the start of the run. This figure also shows the associated 95 % confidence limits in the location estimates. As before, these are calculated using the estimated landmark location covariance. The landmark location estimates are thus consistent (and conservative) with actual landmark location errors being smaller than the estimated error. This figure also shows that there is some bias in the landmark location estimates. However, this bias is well within the accuracy of the true measurements.

V. Discussion and Future Work

An algorithm for estimating robot location using a stereo vision sensor is presented in this paper. It is seen that a SLAM algorithm can be easily augmented to incorporate three-dimensional data. The problem becomes even simpler as the robot is moving on a plane. The effect of small roll and pitch motions experienced as the robot moves appear to be not significant. In the experiments conducted, the environment is populated with a series of artificial landmarks specially selected to enable reliable landmark extraction. Although this requires installation of infrastructure, as the locations of the landmarks can be arbitrary, it can be argued that this is a relatively easy task.

There are number of issues that can be addressed to progress this work further. If installing infrastructure is acceptable, then the key question is how to determine the number and distribution of the landmarks to achieve a desired accuracy of location estimates. While the answer to this question may be specific to a given environment, an indicative relationship between the landmark density and the accuracy achievable will be valuable.

The other major area of work is on the investigation of the use of natural features. This is clearly advantageous in exploring genuinely unknown environments. The challenges in this situation are the efficient extraction and tracking of viewpoint invariant landmarks. The fact that a landmark can be extracted from the image does not mean that the depth to this landmark is available. Even when the depth is available, the accuracy of the depth measurement may not be adequate. Use of natural features in this situation is therefore a challenge. Adequate measures for feature quality are required. It is also necessary to investigate the effect of the parameters used to define the quality of the disparity map.

Acknowledgment

This work is supported by the ARC Centre of Excellence programme, funded by the Australian Research Council (ARC) and the New South Wales State Government.

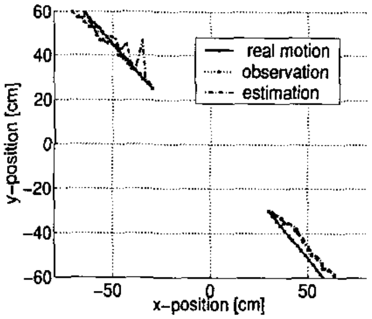
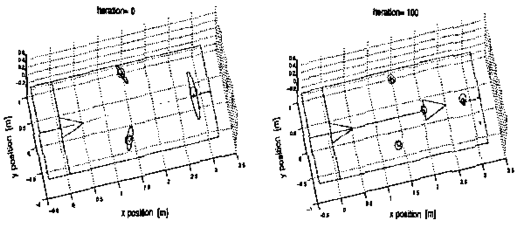


Fig. 8. Estimation on vision



(a) Initial uncertainty

(b) Final uncertainty

Fig. 9. Three-dimensional view of the robot and landmark locations

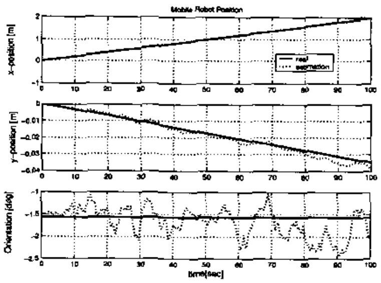


Fig. 10. Robot position and orientation

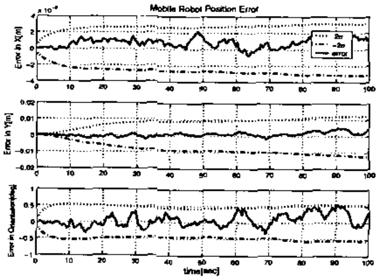


Fig. 11. Robot localization error

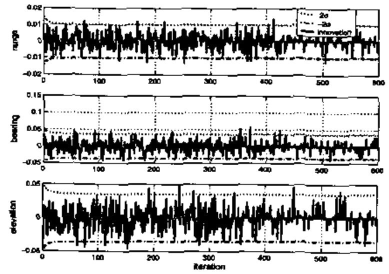


Fig. 12. Innovation time history

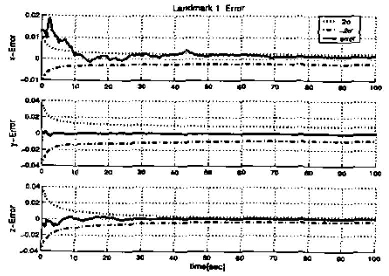


Fig. 13. Error covariance

REFERENCES

[1] Dissanayake G., Newman P., Clark S., Durrant-Whyte H.F. and Csorba M., *A Solution to the Simultaneous Localization and Map Building (SLAM) Problem.* IEEE Transaction on Robotics and Automation, Vol.17, No.3, 229-241, June 2001.

[2] Moon I., Miura J. and Shirai Y., *Dynamic Motion Planning for Efficient Visual Navigation under Uncertainty.* Proceedings of 5th Int. Conf. on Intelligent Autonomous Systems, 172-179, 1998.

[3] Moon I., Miura J. and Shirai Y., *On-line Modeling and Selection of Visual Landmarks under Uncertainty.* JSME Proceedings of 99' Conf. on Robotics and Mechatronics, June 1999.

[4] Ayache N. and Faugeras O.D., *Maintaining Representations of the Environment of a Mobile Robot.* IEEE Transaction on Robotics and Automation, Vol.5. No.6, 804-819, December 1989.

[5] Davison A.J. and Murray D.W., *Mobile Robot Localisation Using Active Vision.* In Proc. 5th European Conf. on Computer Vision, 809-825, Freiburg, 1998.

[6] Ota J., Yamamoto M.,Ikeda K., Aiyama Y. and Arai T., *Environmental Support Method for Mobile Manipulators Using Visual Marks with Memory Storage.* Journal of the Robotics Society of Japan, Vol.17, No.5, 66-72, July 1999.

[7] Smith R., Self M. and Cheeseman P., *Estimating uncertain spatial relationships in robotics.* In Autonomous Robot Vehicles, I.J. Cox and G.T. Wilfon, Eds, New York, Springer Verlag, 167-193, 1990.

[8] Murray D. and Little J., *Stereo vision based mapping for a mobile robot.* In Proc. IEEE Workshop on Perception for Mobile Agents, June 1998.

[9] Demirdjian D. and Darrell T., *Motion Estimation from Disparity Images.* In Proc. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.19, No.10, 213-218, October 2001.