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

An Active Trinocular Vision System for Sensing Mobile Robot Navigation Environments

Min Young Kim, Hyungsuck Cho and Hyunki Lee
Department of Mechanical Engineering,
Korea Advanced Institute of Science and Technology,
373-1, Guseong-dong, Yuseong-gu, Daejeon, 305-701, Korea
(Tel : 82-42-869-3253; Fax : 82-42-869-3210 ; E-mail: hscho@lca.kaist.ac.kr)

Abstract—**Intelligent autonomous mobile robots must be able to sense and recognize 3D indoor space where they live or work. In this paper, we propose a new 3D sensing system using the laser structured lighting method, because of its robustness on the nature of the navigation environment and the easy extraction of feature information of interest. The proposed active trinocular vision system is composed of a flexible multi-stripe laser projector and two cameras arranged with a triangular shape. Based on the virtual camera concept and the trinocular epipolar constraints, the matching pairs of line features in two real camera images are established, and 3D information from one-shot image can be extracted on the patterned scene. Especially, for robust line feature matching, we propose a new correspondence matching algorithms based on line grouping and probabilistic voting. Finally, a series of experimental tests is performed to show its efficiency and accuracy.**

Keywords- trinocular visio; active 3D sensor; laser pattern

I. INTRODUCTION

A major research issue for service robots is the creation of environment sensing and recognition system for mobile robot navigation and task execution, which is robust and efficient. A variety of machine vision techniques have been developed for the determination of 3D scene geometric information from 2D images. Especially, although the binocular vision sensors among them have been widely used as representative ones of passive visual sensors for mobile robots, they still suffer from image intensity variation due to the illumination noise, insufficient feature information on environment composed of plain surfaces, and correspondence problem between multiple images. These reasons have made most mobile robot researches on 3D environment reconstruction using visual sensors to deal with just straight line edge and corner as interesting features [1]-[3], but these features are saliently observed in well arranged and structured environment with polygonal objects or polygon-textured surfaces. In addition, this information is not sufficient to describe the whole structure of 3D space. Therefore, robots frequently use active sensors for more reliable range sensing, and have become recognized as a hopeful alternative proposal, which include the infrared sensor, the ultrasonic sensor, and the laser sensor [4]. In many approaches to indoor robot applications, laser sensor has been used for detail sensing and modeling objects, which is commonly categorized to the laser visual sensor and the laser range finder measuring the time-of-flight. Though the laser range finder has more advantages in views of measuring range and relative accuracy, it need still high cost with high power consumption and heavy weight. In addition, the latter needs more scanning procedure than the former. This scanning procedure is a time consuming task to limit the sensing time, and needs a precisely controlled scanning mechanism. In order to keep up the advantages of the sensor system using the laser-structured light and to decrease the sensing time without degradation of the sensor resolution, it is necessary to develop a new visual sensor system different from the sensors mentioned above. In this work, we propose an active trinocular range sensor composed of a laser pattern projector and two cameras. As a general object-shape measuring device, not for mobile robots, Blake, et al. [5] developed a trinocular active range sensor with a similar configuration, which consists of two line pattern projectors and a camera for image acquisition. But, it employs careful alignment of two projector patterns with a camera to solve the ambiguity problem of node labels [6]. Without this alignment, our sensor system is just based on the trinocular epipolar constraints of the trinocular vision. The pattern projector is a flexible multi-stripe laser projector capable of varying the line-to-line spacing between stripes. Because the projector can be modeled as another virtual camera with previously known input image, this sensor system can be treated as a trinocular vision, and the acquired image can be analyzed using trinocular vision theory. For reliable correspondence matching between line features, we propose a special correspondence matching technique based on line grouping and probabilistic voting.

II. NAVIGATION ENVIRONMENT

A. Indoor navigation environment for mobile robots

Navigation environments can be split into one of the following categories: 1)known, 2)unknown and 3)partially known environments. Especially, a partially known environment means that the environment information modelled to a certain extent is given to the robots, but insufficient to fully support task completion. Since most applications for indoor robot navigation are closely related to this situation, we will deal with this case, and assume that most objects composing navigation environments are unknown and cluttered on the space. There might be many

This research was supported by University IT Research Center Project.

different types of objects that mobile robots encounter on navigation or task space. They are typically categorized according to their shapes and surface conditions, for example, texture and reflectance, as shown in Fig. 1.

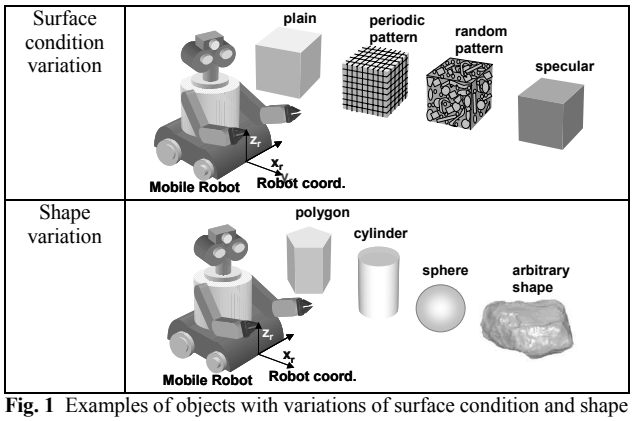


Fig. 1 Examples of objects with variations of surface condition and shape

B. Sensing system for mobile robots

Most robots for robustly sensing the environments are equipped with various sensors. Among them, laser range sensor and stereovision sensor are representatives commonly used for 3D perception. In case of stereovision, its performance is seriously influenced by surface characteristics and shape of the target object and the illumination condition on the scene. If the target object does not have sufficient texture information or the periodic pattern is printed on its surface, this sensor is impossible to acquire the whole 3D information of the object, but can deal with edges and corners of polygonal objects. For the curved object with plain surface, this sensor cannot reconstruct even boundary information accurately [7]. Similar to stereovision, laser range sensors except time-of-flight laser range finder are based on the optical triangulation principle. Since the laser used as the light source has more attractive features than other light sources, this sensor gives more reliable and accurate results than other optical sensors. If an object is not mirrorlike, it works successfully, and gives best results among four sensors for various objects with lambertian surface. Although the laser sensors are relatively attractive, they always need the scanning procedure for 3D scene perception. Development of faster and more flexible laser sensors has been for mobile robots [8][9] and robot manipulator [10][11].

III. MEASUREMENT OF 3D SURFACE

A. Sensing principle of the proposed sensor system

In general stereovision, correspondence problem has been the major research issue in stereovision, and studied until now since its birth. For easy handling of this difficulty, some researchers frequently have used the structured light method replacing one camera with a pattern projector [12]. The main stream of the related researches is the technique known as coded structured light with two categories: 1) temporal codes, and 2) spatial codes. Morano, et al. [13] proposed a vision system using structured light with pseudorandom codes. But, it has some disadvantages that the used color pattern projector is not compact, and the reflected pattern color may be different from the projected pattern color due to different reflection properties on the object surface, which often result in incorrect correspondence matching. Chen, et al. [14] adopted a vision system using a color structured lighting and two cameras for easy correspondence matching of the projected multi-line color stripe pattern. With one more camera, they replaced the lighting-to-image correspondence problem by an easier problem of image-to-image correspondence. This system is more robust than former to color distortion problem, however, approaches using color pattern projector have size, heat, and cost problem. Therefore, its application to mobile robots is still beyond, and we propose a novel 3D measurement system that utilizes a flexible multi-stripe laser pattern generator. Fig. 2 shows the basic concept of the proposed sensor system, which is composed of a multi-stripe laser generator and two cameras.

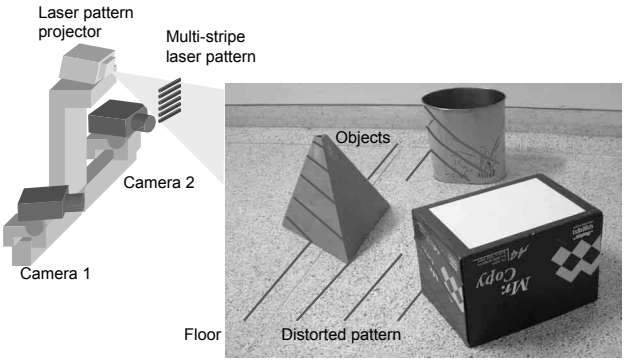


Fig. 2 Concept of a three-dimensional surface profile sensing system

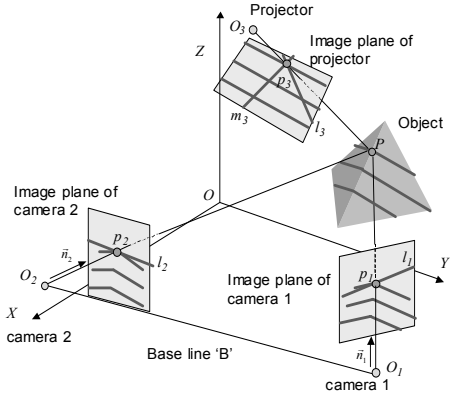


Fig. 3 View analysis in each image for finding the correspondence point

Since the pattern projector can be modeled as another camera from which we know the input image [5], the whole system can be treated as a trinocular vision system. To see how this sensor system works, we analyze the view of each image device as shown in Fig. 3, and imagine a point P on the object. $OXYZ$ denotes the world coordinates system, and the coordinates of a point on this frame are represented as (x_p, y_p, z_p). O_1, O_2 and O_3 represent the optical centers of two cameras and the projector. Assume that the object point P on the projected pattern is observable as the image point p_1 on the image of camera 1 and the matching point p_2 on the image of camera 2. Based on epipolar constraint, the latter must lie on the epipolar line l_2 that is the projection of the line O_1p_1 onto the image plane of camera 2 [1]. Let T_1, T_2, and T_3 be the perspective transform matrix of three cameras. (u_1, v_1) and (u_2, v_2) denote the image coordinates of p_1 and p_2. Relations

between each perspective transform matrix and image coordinates (u_i, v_i) are given by

$$\begin{bmatrix} U_i \\ V_i \\ s_i \end{bmatrix} = T_i \begin{bmatrix} x_p \\ y_p \\ z_p \\ 1 \end{bmatrix}, \begin{bmatrix} u_i \\ v_i \end{bmatrix} = \begin{bmatrix} U_i / s_i \\ V_i / s_i \end{bmatrix}, \text{ and } T_i = \begin{bmatrix} \vec{t}_1^{\,i} & t_{14}^i \\ \vec{t}_2^{\,i} & t_{24}^i \\ \vec{t}_3^{\,i} & t_{34}^i \end{bmatrix}$$

$$for\ i=1,2,3 \qquad (1)$$

where (U_i, V_i, s_i) denotes the projective coordinates of an image point on i^{th} camera, and i indicates i^{th} camera. The parametric equation of the line O_1p_1 in 3D space is defined by $\begin{bmatrix} x & y & z \end{bmatrix}^T = \vec{O}_1 + \lambda \cdot \vec{n}_1$ in which λ, $\vec{n}_1$, and $\vec{O}_1$ are arbitrary scaling factor, the directional vector of the line, and the camera origin, respectively. The epipolar line l_2 for the point p_1 can be obtained by

$$\begin{bmatrix} U_2 \\ V_2 \\ s_2 \end{bmatrix}_{l_2} = T_2 \begin{bmatrix} \vec{O}_1 + \lambda \cdot \vec{n}_1 \\ 1 \end{bmatrix} \qquad (2)$$

where $\vec{n}_1$ denotes the direction vector of the line O_1p_1. This epipolar line constraint reduces the search for a corresponding image point from two-dimension to one-dimension. Additionally, in our system, the solution of the correspondence problem can be supported by another third image. In Fig. 3, the epipolar line m_3 is another constraint that the relation of the projector and camera 2 makes on virtual image plane for image point p_2, and the epipolar line l_3 is the other constraint that the relation of the projector and camera1 make for image point p_1. Similar to Eq.(2), the epipolar lines, m_3 and l_3, are defined as follows:

$$\begin{bmatrix} U_3 \\ V_3 \\ s_3 \end{bmatrix}_{m_3} = T_3 \begin{bmatrix} \vec{O}_2 + \lambda \cdot \vec{n}_2 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} U_3 \\ V_3 \\ s_3 \end{bmatrix}_{l_3} = T_3 \begin{bmatrix} \vec{O}_1 + \lambda \cdot \vec{n}_1 \\ 1 \end{bmatrix} \qquad (3)$$

where $\vec{n}_2$ denotes the direction vector of the line O_2p_2. If the image points p_1 and p_2 are a corresponding pair, these epipolar lines must intersect at a point p_3 on the projected line feature of the virtual multi-stripe laser image. Using this characteristic, correspondence candidate points can be checked one by one. When the object point P on the pattern is visible at two cameras, these constraints are always effective. After correspondence matching, 3D coordinates of an object point P are easily computed [2]. Using a pair of correspondence points, $p_1(u_1, v_1)$, $p_2(u_2, v_2)$, and $p_3(u_3, v_3)$, and Eq.(1), a linear equation for 3D coordinates $a=(x_p, y_p, z_p)$ of the object point P is constructed as follows:

$$Aa = b \qquad (4)$$

where $A = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}$, $b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$, $A_i = \begin{pmatrix} \left(\vec{t}_1^{\,i} - u_i \vec{t}_3^{\,i}\right)^t \\ \left(\vec{t}_2^{\,i} - v_i \vec{t}_3^{\,i}\right)^t \end{pmatrix}$, and $b_i = \begin{pmatrix} u_i t_{34}^i - t_{14}^i \\ v_i t_{34}^i - t_{24}^i \end{pmatrix}$ $for\ i = 1,2,3.$

Using a least square error method, the solution of this equation is derived by

$$a = \left(A^T A\right)^{-1} A^T b \cdot \qquad (5)$$

For the proposed sensor system, the procedure of extracting the 3D information is represented in Fig. 4. For the robust correspondence matching between line features, we propose a novel correspondence matching algorithms.

B. Correspondence matching between line features

The correspondence matching between line features is based on the point-to-point correspondence check algorithms. Each line feature is composed of a set of connected points. In order that a line feature of the left image becomes matched with one of the right image, all correspondences between points consisting of two line features must be theoretically established. However, a line feature in one image is often broken into a few ones or disappeared in the other image due to own viewing pose and different optoelectronic characteristics of each camera. Since it makes correspondence problem more difficult, we develop a probabilistic voting method with two phases: 1) voting phase 2) ballot counting phase. In voting phase, a set of points on each line feature attends the voting. Fig. 5 shows the three-dimensional polls for correspondence voting between line features, which is composed of so-called accumulator array with l x m x n size. Here, m and n denote a number of line feature labels observed in left and right image, respectively, and l denotes a number of the laser lines projected on a target scene. In Fig. 5, each 2D slice of the whole 3D accumulator array represents the polls for a certain laser line. For multiple laser lines, a set of 2D accumulator array constructs the 3D polls. These cells clearly play a role to depict the correspondence relationship between line features and the corresponding laser stripe among the projected stripes. For example, a cell, $A(i, j, k)$, denotes the probabilistic confidence level for correspondence-matching between a line feature with i^{th} label in left image, a line feature with j^{th} label in right image and a laser stripe with k^{th} label in the pattern image. In the voting phase, after initialization of the accumulator array, for all pixels consisting of line features in one image, the trinocular epipolar constraint is checked. Whenever a pair of pixel points that pass successfully the correspondence check algorithm is found, the confidence value of the related accumulator cell is iteratively increased. The corresponding accumulator cell, $A(i, j, k)$, is updated as follows:

$$A(i,j,k) = A(i,j,k) + 1 \cdot \qquad (6)$$

However, if there are N candidate points passing through the check algorithms simultaneously, the amount of change of probabilistic confidence value is decreased as $1/N$, and all of the matching candidate pairs share the correspondence probability. A set of corresponding accumulator cell is updated as follows:

$$A(i,j,k) = A(i,j,k) + 1/N \cdot \qquad (7)$$

After the correspondence voting for all pixels consisting of line features is executed, the acquired accumulator array represents a list that can describe the matching situation between line features at a glance. The value of the cell, $A(i, j, k)$, depends on a number of the matched image points between i^{th} line in left image, j^{th} line in right image, and k^{th} laser stripe. Therefore, real

correspondences existing in three images will produce large values of the related cells. This results in local maxima along the vertical columns in the 3D accumulator space. In the ballot counting phase, Non-Maxima Suppression is performed for finding the global maximum along each vertical column of the 3D accumulator array. The results are stored at the essential accumulator array, C, in a following form:

$$C(i,j)=\{(A(i,j,k),k)\,|\,Max(A(i,j,k)) \quad for \;\; k=1,\cdots,l\} \quad (8)$$

where $Max(\cdot)$ is a maximum selection function. To get rid of the mismatched correspondence candidates due to noise, the essential accumulator array cells to have less confidence level than a predefined threshold value, T_{sub}, are cleared out. Finally, the uniqueness constraint is applied on the filtered essential accumulator array. This constraint denotes that each line feature in a pair of stereo images should be matched to only one laser stripe in the pattern image. This constraint can be implemented on the essential accumulator array as follows: 1) Search essential accumulator cells with the identical line label in the left image and the different laser stripe label in the pattern image, 2) Group them into subgroups with an identical laser stripe label, and Sum up the confidence of each matching candidate in a subgroup for evaluating total confidence that this line feature corresponds to the laser stripe, 3) Compare the integrated confidence of each subgroup with each other, and determine a candidate with the highest value as the correct matching one, and 4) the above procedure (1)~(3) is repeated for line features in right image. The main advantage of using 3D correspondence voting method lies in the fact that it is insensitive to missing parts of line features, image noise and other line features that sometimes disturb unique correspondence matching. For example, a noisy or rough line feature will not yield a cell with large value in the accumulator array but result in a cell with low value.

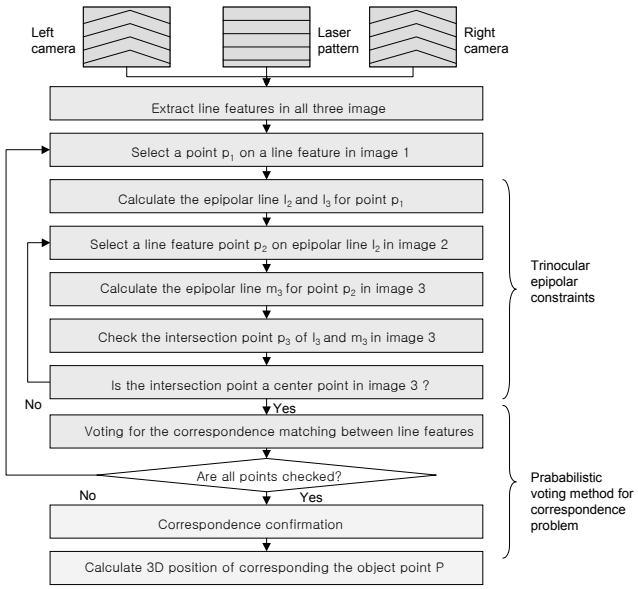


Fig. 4 Algorithmic flowchart for the correspondence matching

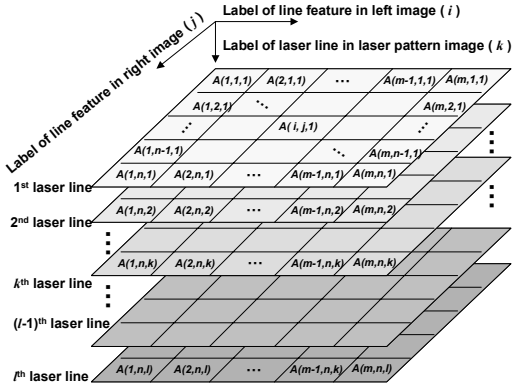


Fig. 5 Accumulator array used for correspondence matching between line features, with l x m x n size

IV. Sensor System and Its Calibration

Fig. 6 shows the experimental setup for evaluating the proposed sensing system and calibrating this sensor system. Through the lower circular window of the pattern generator, a laser pattern is projected on the scene, and the upper window is reserved for other purpose. The pattern generator consists of a rotating polygon mirror which makes a laser slit beam sweep on the scene, an assembly of a laser diode and a cylindrical lens that make a laser stripe, a photo diode for synchronous signal generation, and micro-processor for the speed control of polygon mirror and the on/off time of the laser diode. The whole dimension of the pattern generator is 150x170x60*mm*. Each camera has 640x480 pixels, and the focal length of the camera lens is 16*mm*. To increase the signal-to-noise ratio, a band-pass filter is attached in front of each camera. The length of the base line between two cameras is about 100*mm*. For calibration of each camera, a method proposed by Tsai is adopted. The world coordinate frame is fixed at a position with about 1,235*mm* distance from sensor.

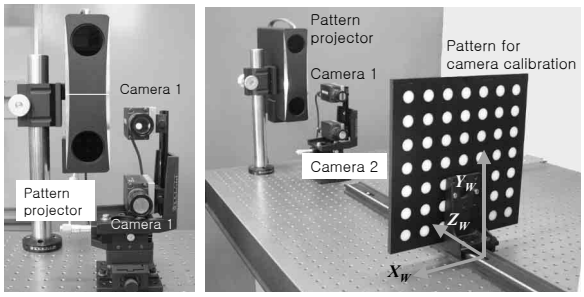


Fig. 6 Sensor system and camera calibration setup

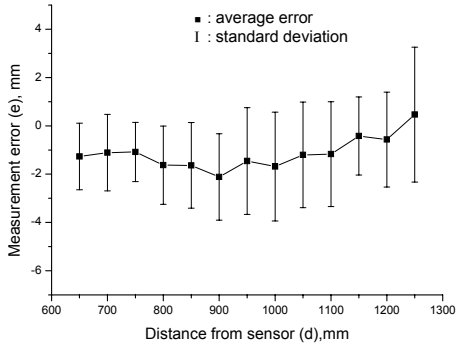


Fig. 7 Depth measurement using the proposed sensor (measurement error in direction of z axis)

V. EXPERIMENTS FOR SENSING ENVIRONMENTS

A. Image processing for line feature extraction

To implement the proposed sensing principle, first it is necessary to robustly find the line feature information from the acquired images. The line feature extraction process in each input image is a combination of several image processing algorithms, which are summarized as follows: 1) For image noise rejection, noise filtering is applied to each image. 2) For each column, a thickness of line features is predicted using derivative of gaussian filter. 3) For each column, a line mask with a form of laplacian of gaussian filter is generated according to the estimated laser line thickness, and the line pattern is enhanced using this filter. 4) For local line center detection, non-maxima suppression along normal direction of the curved line is applied to the local area corresponding to the line mask size. 5) After acquiring the essential line image with thinned laser line, line blobbing is performed for grouping the connected line features and segmenting the disconnected line features with the rejection process of short line features due to image noise.

B. Experiments on objects with various shapes

To evaluate the performance of the sensor system, it is necessary to examine a simple object first. For this purpose, a flat plane is placed in front of the sensor system instead of the calibration target shown in Fig.6, and a laser line pattern is projected on it. With variations of the distance from the sensor to this plane, d, the measurement tests were performed, and the experimental results are summarized in Fig. 7. The drift error is mainly dependent on the sensor calibration, and the error standard deviation, σ_e, results from the image processing and quantization error. The error standard deviation tends to increase with the object distance. At d=1,250mm, the average error and the error standard deviation in direction of z axis are 0.467mm and 2.79mm. Secondly, we performed measurement tests on four objects composed of polygons and curved ones. Polygonal objects are a tetrahedron and a pentagonal cylinder, and curved object are a cylinder and a cylindrical object with curved surface as shown in Fig. 8. In Fig. 9, the acquired 3D point data on each object is represented with the same order. In each case, a pattern with 32 lines is projected on the scenes. The experimental results show that the proposed sensor system can measure the objects with plain surface, without depending on variations of their shapes. The broken line on the results is caused by the incomplete line extraction on each input image, and the ambiguity of the correspondence between these incomplete line features. Shape measurement results on the objects were acquired through a surface fitting method based on the least square error minimization. First, a contained angle of two planes of tetrahedron after the plane fitting process was measured as 71.4° comparing with the ground truth, 70.5°. Second, in case of a cylindrical pentagon, a contained angle of its two planes was measured as 107.2° comparing with the ground truth, 108°. The angle measurement error of two polygonal objects shows less than 0.9°. Next, after cylinder fitting, the radius of cylinder was estimated as 105.3mm with 5.3mm error to the real value. In this case, the cylinder fitting method using just measurements of a front side of the object is a major reason to lead this error. If we use more information acquired at different views, it will be decreased. Fig. 10 represents the results of the wireframe modeling on four target scenes: 1) first scene with a cup, a box and a plane, 2) second scene with a flowerpot, two small boxes, a large box and a plane, and 3) third scene with a cup, a small box, a large box and a plane. The 3D point cloud data was acquired through a scanning of a projected laser pattern. For the scanning, the laser pattern composed of 16 lines is shifted 5 times with stepwise increment of a projection angle, about 0.4°. The acquired 3D point cloud data was input to a commercial software for the wireframe modeling. The result shows the proposed sensor is robust to variations of object shapes and surface conditions. Especially, these scenes include combinations of objects painted with a figure, objects with plain surface, objects with a periodic texture, and objects printed with commercial logos. The modeling results shown in Fig. 10 represent well the scene composed of cluttered objects with various shapes and surface conditions.

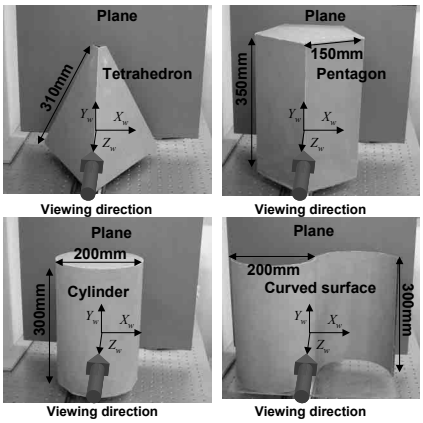


Fig. 8 Target objects for experiments

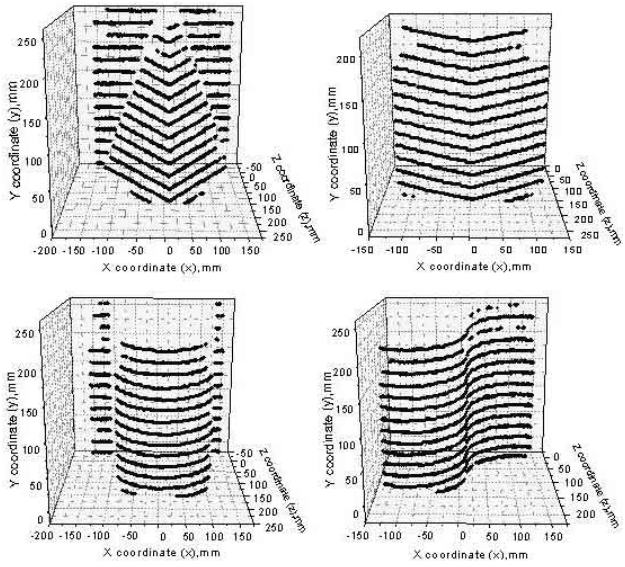


Fig. 9 Three-dimensional point data obtained for each target object (32 laser line projection)

VI. CONCLUSIONS

A major research issue associated with mobile service robots is the creation of environment sensing and recognition system for navigation and task execution that is robust and efficient. The widely used laser sensing technique needs the scanning procedure for acquiring the 3D information on the entire scene. In this paper, we proposed a novel three dimensional sensor system for environment perception needed for mobile robots with less scanning time by using multi-laser stripe pattern projection, composed of the laser pattern projector of generating multiple line stripes and two cameras with laser filters. The system can extract 3D information on the scene without sequential image acquisition and complex spatial pattern coding. Though the performed experiments are not quite general, the experimental results show the feasibility of applying this sensor system to sensing mobile robot navigation environment. In addition, because this sensor system use a laser as the light source, and the CCD camera tends to become miniaturized, the sensor size and power consumption do not make troubles for robot operation when it is implemented on the mobile robots.

REFERENCES

[1] O. Faugeras, *Three dimensional computer vision: A geometric viewpoint*, MIT Press, 1993.

[2] N. Ayache, *Artificial Vision for Mobile Robots: Stereo vision and Multisensory perception*, MIT Press, 1991.

[3] P. Weckesser and R. Dillmann, "Modeling unknown environments with a mobile robot," Robotics and Autonomous Systems, Vol. 23, pp. 293-300, 1998.

[4] H. R. Everett, *Sensors for Mobile Robots: Theory and Application*, AK Peters, 1995.

[5] A. Blake, D. McCowen, H. R. Lo, and P. J. Lindsey, "Trinocular Active Range-Sensing," IEEE Trans. on Pattern Analysis and Machine Vision, Vol. 15, No. 5, pp. 477-483, May 1993.

[6] G. Hu and G. Stockman, "3-D surface solution using structured light and constraint propagation," IEEE Trans. on Pattern Analysis and Machine Vision, Vol. 11, No. 4, pp. 390-402, 1989.

[7] C. Chen and Y. F. Zheng, "Passive and Active Stereo Vision for Smooth Surface Detection of Deformed Plates," IEEE Trans. on Industrial Electronics, Vol.42, No.3, pp. 300-306, 1995.

[8] H. Surmann, et al., "Fast acquiring and analysis of three dimensional laser range data," Proc. 6th Int. Workshop Vision, Modeling, and Visualization, pp. 59-66, 2001.

[9] S. Elgazzar, R. Liscano, F. Blais, and A. Miles, "Active range sensing for indoor environment modeling," IEEE Trans. on Instrumentation and Measurement, Vol. 47, No.1, pp. 260-264, 1998.

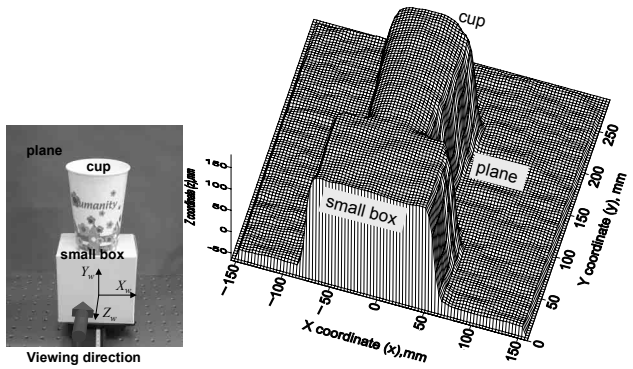
[10] K. Hattori and Y. Sato, "Handy Rangefinder for Active Robot Vision," Proc. IEEE Int. Conf. Robotics and Automation, pp. 1423-1428, 1995.

[11] G. Taylor, L. Kleeman and Å. Wernersson, "Robust colour and range sensing for robotic applications using a stereoscopic light stripe scanner", Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems, pp. 86-91, 2002.

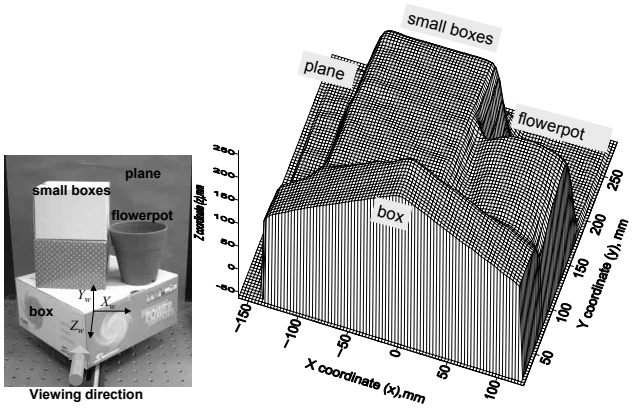
[12] E. Mouaddib, J. Batlle, and J. Salvi, "Recent Progress in Structured light in order to solve the correspondence problem in Stereo Vision," Proc. IEEE Int. Conf. Robotics and Automation, pp. 130-136, 1997.

[13] R. A. Morano, C. Ozturk, R. Conn, S. Dubin, S. Zietz, and J. Nissanov, "Structured light using Pseudorandom codes," IEEE Trans. on Pattern Anayisis and Machine Intelligence, Vol. 20, No. 3, pp. 322-327, March 1998.

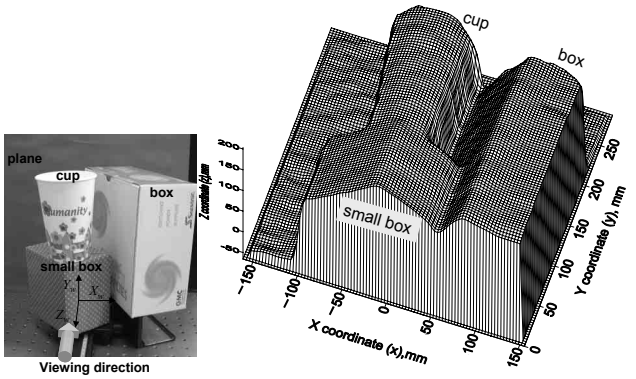
[14] C. S. Chen, Y. P. Hung, C. C. Chiang, and J. L. Wu, "Range data acquisition using color structured lighting and stereo vision," Image and Vision Computing, Vol. 15, pp. 445-456, 1997.



a) a wireframe model of first scene



b) a wireframe model of second scene



c) a wireframe model of third scene

Fig. 10 wireframe modeling results using the measured data
(for a scanning, the laser pattern is shifted 5 times with data acquisition)