

Novel Visual Sensor System Fusing An Active Trinocular Vision and A Stereo Vision for 3D Range Data Acquisition

Min Y. Kim^a, Hyungsuck Cho^a, and Hyun Ki Lee^a

^aDepartment of Mechanical Engineering, Korea Advanced Institute of Science and Technology

ABSTRACT

One of major research issues associated with 3D range acquisition is the creation of sensor systems with various functionalities and small size. A variety of machine vision techniques have been developed for the determination of 3D scene geometric information from 2D images. As one of active sensors, structured lighting method has been widely used because of its robustness on the illumination noise and its extractability of feature information of interest. As one of passive sensors, stereo vision does also due to its simple configuration and easy construction. In this work, we propose a novel visual sensor system for 3D range acquisition, using active technique and passive one simultaneously. The proposed sensor system includes inherently two types of sensors, an active trinocular vision and a passive stereo vision. In the active vision part of this sensor, the structured lighting method using multi-lasers is basically utilized. In its stereo vision part, a general passive stereo is constructed. Since each of them has its own advantages and disadvantages on the measurements of various objects, we propose sensor fusion algorithms for acquiring more reliable range information from them. To see how the proposed sensing system can be applied to real applications, we mount it on a mobile robot, and a series of experimental tests is performed for a variety of configurations of robot and environment. The sensing results are discussed in detail.

Keywords: Active trinocular sensor, stereo vision, 3D visual sensor, sensor fusion

1. INTRODUCTION

A major research issue for service robots is to develop 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 a novel visual sensor system combining an active trinocular range sensor and a stereo vision, which is composed of a laser pattern projector and two cameras. It is a practical

*mykim@lca.kaist.ac.kr; phone +82-42-869-3253; fax +82-42-869-3210; Dept. Mech. Eng., Korea Advanced Institute of Science and Technology, 373-1, Guseung-dong, Yuseung-gu, Daejeon, Korea

implementation on a mobile robot of the sensing principle presented in our previous paper[5]. In case that the laser projector is utilized, because the projector can be modeled as another virtual camera with previously known input image, this sensor system can be treated as an active trinocular vision, and the acquired image can be analyzed using trinocular vision theory. For reliable correspondence matching between line features made by the projected laser lines, it uses a special correspondence matching technique based on line grouping and probabilistic voting. However, in case that the laser projector is not used, two cameras configure a normal stereo vision and can acquire stereo intensity images on scenes. This information can be used for extracting another 3D range information based on stereo vision theory or for supporting the robustness of the active trinocular vision sensor.

Recently, some researchers have performed researches combining laser range data and intensity image: 1) texture mapping on the range data and registration for realistic 3D modeling [6][7], 2) efficient edge finding and image segmentation using both range image and intensity image[8][9], and 3) range data updating [10][11]. Especially, Tate and Li[11] developed a multi-resolution method for acquisition of high resolution depth map from a low-resolution laser range image and a stereo pair of high-resolution intensity images. In this paper, we propose a method solving the correspondence ambiguity problem of the active trinocular range sensor using stereo intensity information.

The mobile robot used in this paper is shown in Fig. 1. For autonomous navigation in unknown space, it is equipped with a number of sensors, e.g., ultrasonic sensors, infrared sensors, and the proposed sensor head. During the navigation process, it perceives the navigation environment by using these sensors. For obtaining the detail 3D range information on environment, the proposed sensor head is utilized.

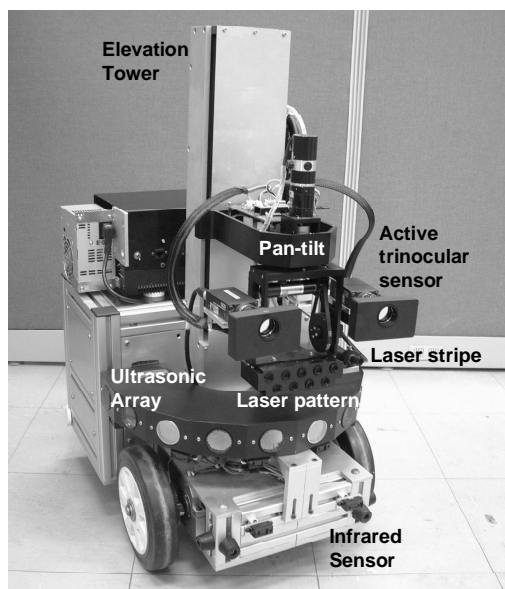


Fig. 1 An autonomous mobile robot, LCARIII

2. A VISUAL SENSOR SYSTEM WITH ACTIVE TRINOCULAR VISION AND STEREO VISION

Fig. 2 a) shows an enlarged photo of the proposed sensor system mounted on mobile robot, which consists of two cameras and a laser projector. For changing the viewing direction freely, this sensor head is mounted on pan-tilt unit. The vergence angle of two cameras is fixed, and the configuration of two cameras and a laser projector does too. For sensing environments robustly under variations of illumination conditions, the filter changer with the optical band pass filter is attached to front part of each camera. For taking laser images, the band pass filter is placed at each camera lens. Therefore, from the scenes on which the laser pattern is projected, only distorted laser information except unnecessary ones can be grabbed in cameras. Conversely, for taking intensity images, the band pass filter is removed, and the conventional intensity image can be obtained in each camera. The insertion and removal of each optical filter is controlled by microprocessor and step-motor controller as shown in Fig. 2 b). The horizontal laser stripe located at right side of the sensor head is used for making 2D navigation map and localizing the mobile robot on the map.

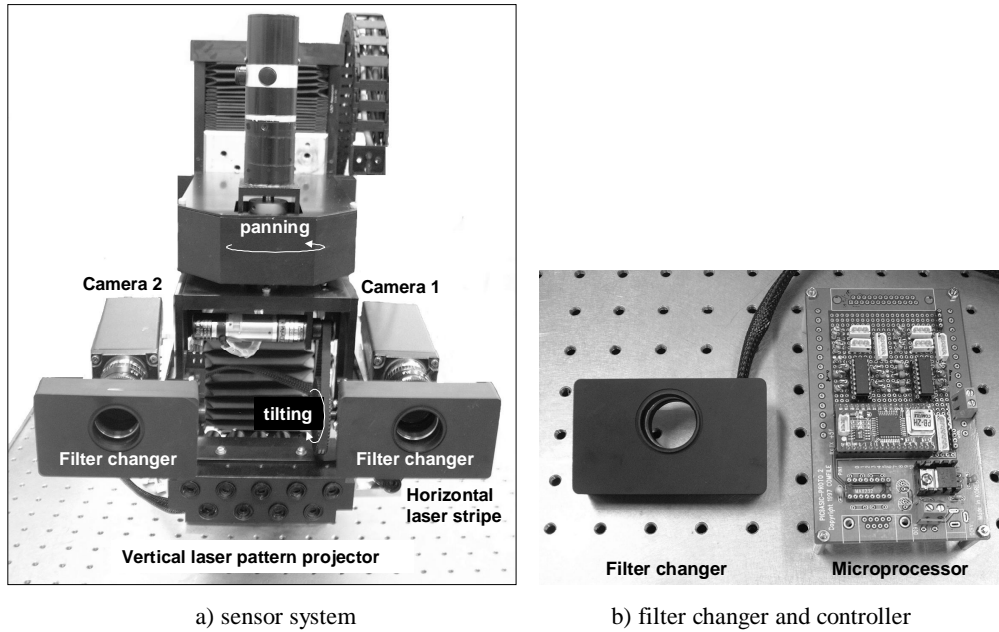


Fig. 2 Sensor head for 3D environment perception

2.1 Active trinocular system

The laser beam generated by a multi-stripe laser pattern projector is projected onto the scene. Then, two cameras composing a stereovision system take the distorted pattern images of the scene onto which the laser is projected. Generally, ambiguities occur during the process of identifying the corresponding pairs of image elements between stereo images. For the robust correspondence matching between laser line features acquired in two cameras, we proposed a novel correspondence matching algorithms in our previous paper[5]. Here, we explain it briefly. 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 developed 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. 3 shows an algorithmic flowchart for 3D information extraction of laser line features observed in the active trinocular vision. Firstly, for line features extracted in three input images, labeling procedure is performed. Then, for the pixel elements forming each line feature, trinocular epipolar constraint is checked. Whenever a trio of corresponding pixels passing this test is discovered, the labels of related line features are memorized, and the voting procedure is performed. After the voting is completed for all of the line features, the ballot counting procedure is followed.

Fig. 4 shows the three-dimensional polls for correspondence voting between line features, which is composed of so-called accumulator array with $l \times m \times n$ 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. 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:

$$A(i, j, k) = A(i, j, k) + 1. \quad (1)$$

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:

$$A(i, j, k) = A(i, j, k) + \frac{1}{N} \quad (2)$$

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:

$$C(i, j) = \{(A(i, j, k), k) | \text{Max}(A(i, j, k)) \text{ for } k = 1, \dots, l\} \quad (3)$$

where $\text{Max}(\cdot)$ is a maximum selection function.

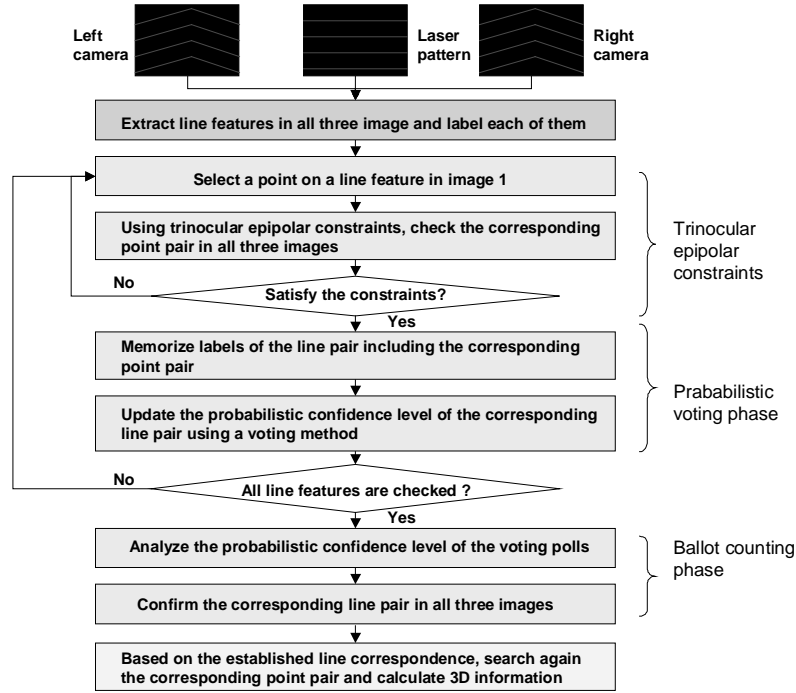


Fig. 3 Algorithmic flowchart for 3D information extraction of laser line features in the active trinocular vision

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.

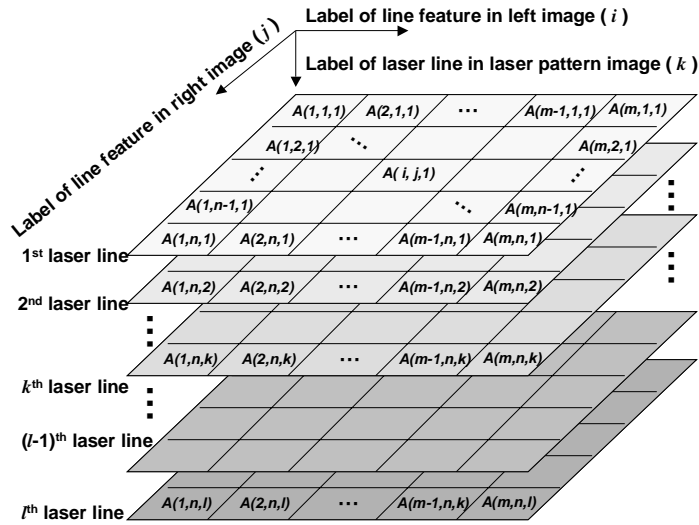


Fig. 4 Accumulator array used for correspondence matching between line features, with $l \times m \times n$ size

2.2 Stereo vision

When a normal stereo vision is utilized in this sensor system, a range information can be extracted from the acquired intensity images. However, the well-known correspondence problem still makes this process difficult. The search for the correct match of a point is called the correspondence problem and is one of the central and most difficult parts of the stereo problem. As a representative approach, the epipolar constraint between two cameras and the area-based intensity correlation method is combined as shown in Fig. 4. Among a variety of similarity measures between two windows specified in left and right image, a simplest way called SAD(sum of absolute differences) or SSD(sum of squared differences) has been used widely. A detail description for current trend of stereo vision research can be found in the reference [12].

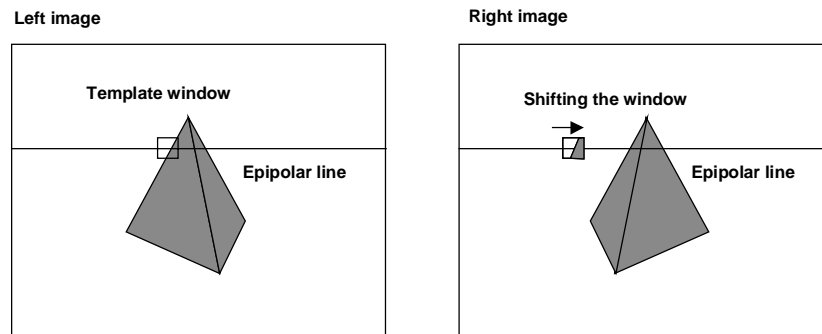


Fig. 4 Area-based correspondence matching with epipolar constraints in stereo vision

3. ALGORITHMS FOR SENSOR OPERATION

3.1 Fusion architecture

Fig. 5 shows the concept of one example of fusing information of laser line observed from active trinocular vision and intensity images acquired from stereo vision. Here, the information from stereo vision takes a role of solving possibility

of ambiguous correspondence in range information extraction from active trinocular vision. This information fusion architecture makes active trinocular vision part more robust. The cases in which ambiguous correspondences happen in active trinocular vision are as follow:

- 1) local maxima along vertical column in the corresponding accumulator cell, $A(i, j, k)$ have similar values within a boundary.
- 2) the corresponding feature pair is occluded in the other camera, and the incorrect feature pair has a high value at the relevant corresponding accumulator cell, $A(i, j, k)$.

For the confirmation of these ambiguous cases, the stereo vision can give additional constraints to the trinocular laser sensor.

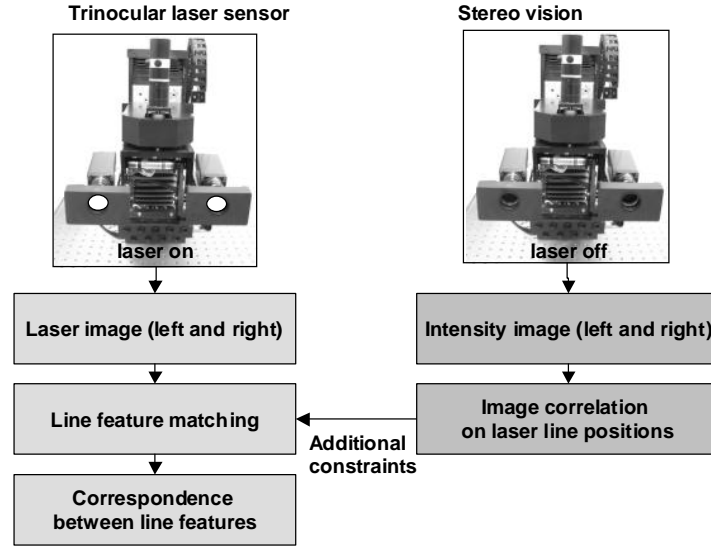


Fig. 5 Conceptual diagram of an approach fusing information from two sensor parts (for solving ambiguous correspondences between laser line features using stereo intensity images)

3.2 Fusion algorithm

For this purpose, another accumulator array, $B(i, j, k)$, is composed from the stereo vision, which has the same configuration as shown in Fig. 4. After initialization of the accumulator array, for all pixels consisting of line features in laser images, the trinocular epipolar constraint is checked. Whenever a pair of pixel points that pass successfully the correspondence check algorithm is found, the correlation cost of them in intensity image is calculated. The correlation accumulator cell, $B(i, j, k)$, is updated as follows:

$$B(i, j, k) = B(i, j, k) + \frac{\sum_{(u,v) \in W} (I_1(u, v) - \bar{I}_1)(I_2(x+u, y+v) - \bar{I}_2)}{\sqrt{\sum_{(u,v) \in W} (I_1(u, v) - \bar{I}_1)^2 \cdot \sum_{(u,v) \in W} (I_2(x+u, y+v) - \bar{I}_2)^2}} \quad (4)$$

where I_1 denotes intensity values in the template window, and \bar{I}_1 is its average intensity value. Similarly, I_2 is intensity values in the shifting window, and \bar{I}_2 is its average intensity value. $\sum_{(u,v) \in W}$ indicates summation over the window.

In this equation, the second term represents a zero mean normalized cross correlation. According to its definition, high value of the correlation accumulator array means that the related laser line features correspond under high possibility.

In fusion phase, decision-making based on two information can be summarized as:

If the essential accumulator array cells have higher confidence level than a predefined threshold value, T_{sub} , and the related correlation accumulator cells have higher correlation level than other predefined threshold value, T_{sub2} , then, the relevant pair of laser line features is confirmed finally.

$$\begin{cases} \text{if } C(i, j) > T_{sub} \\ \text{if } B(i, j, k) > T_{sub2} \end{cases}, \quad \text{Then, line pair } (i, j) \text{ is confirmed.} \quad (5)$$

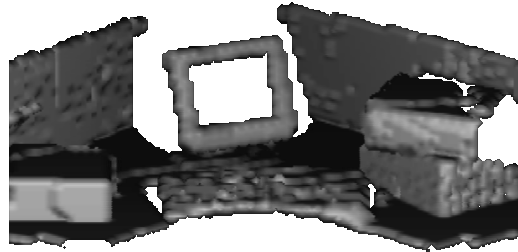
This decision-making rule can be easily implemented by fuzzy rule table, and then membership variables are the correlation accumulator cell value, $B(i, j, k)$ and the essential accumulator cell value, $C(i, j)$. The membership function is designed to be an exponential sigmoid function, which is smoothly differentiable[13].

4. EXPERIMENTS FOR THE PROPOSED SENSOR SYSTEM

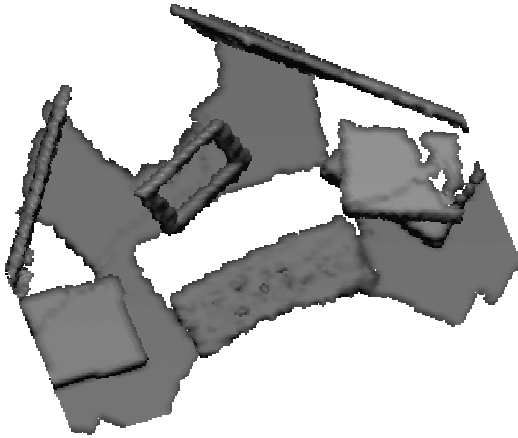
For showing the ability of the sensor system that stereo vision supports active trinocular vision, we tested this operation program on several environment configurations. The representative experimental results are shown in Fig. 6 and 7. In front of the mobile robot with the proposed sensor system, various objects are placed and measured. As shown in figures, the shape of objects, especially, legs of fallen chair, monitor, and keyboard, is well represented in the reconstruction space. The object shape was reconstructed in 100x100x100 voxel space, and the figures are rendered images. The volumetric size of a cubic cell is 1cm³.



a) measured scene



b) side view (rendered from voxel modeling)

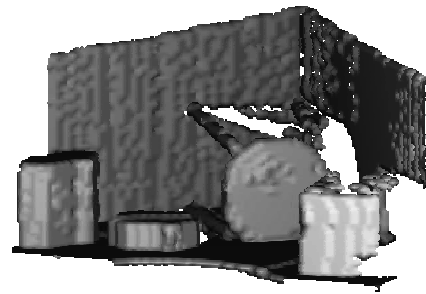


c) bird's eye view (rendered from voxel modeling)

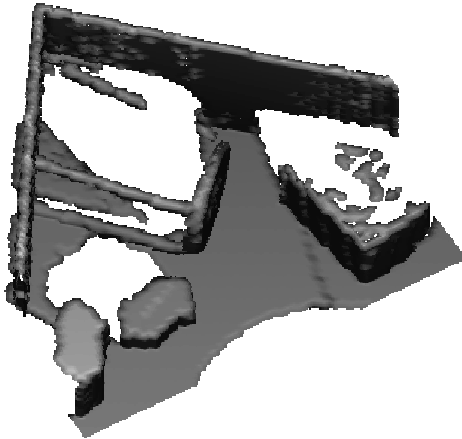
Fig. 6 Measurement on a scene using the sensory system implemented on mobile robots (a LCD monitor, a box with silver-coated paper, a book, and a normal box)



a) measured scene



b) side view (rendered from voxel modeling)



c) bird's eye view (rendered from voxel modeling)

Fig. 7 Measurement on a scene using the sensory system implemented on mobile robots (a box, stacked two bricks, one brick, and a fallen chair)

5. CONCLUSION

A major research issue associated with mobile service robots is the creation of autonomous environment sensing and recognition system for navigation and task execution that is robust and efficient. In this paper, we proposed a novel sensor system consisting of a stereo vision and an active trinocular vision, and did an operation algorithm combining information from these sensors. In the algorithms, intensity information from stereo vision supports imperfect 3D information from active trinocular vision. Fusion of two information is performed in the voting space known as 3D accumulator array, and can be easily implemented by fuzzy decision-making method using simple fuzzy rule-table. The performed experiments show the feasibility of successful perception of the mobile robot with the proposed sensor system on several environments. The proposed sensor operation method is a form that stereo vision supports active trinocular vision. Conversely, the acquired information from active trinocular vision can be used for dense stereo-matching in stereo vision. Now, this research is going on progress, and we are focused on the bi-directional information exchange and supports between two sensors.

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. M. Y. Kim, H. S. Cho and H. K. Lee, "An Active Trinocular Vision System for Sensing Mobile Robot Navigation Environments," IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004 (to be published).

6. P. Dias, V. Sequeira, J.G.M. Gongalves, and F. Vaz, "Automatic registration of laser reflectance and colour intensity images for 3D reconstruction," *Robotics and Autonomous Systems*, Vol. 39, pp. 157-168, 2002.
7. V. Sequeira, K. Ng, E. Wolfart, J.G.M. Gongalves, and D. Hogg, "Automated reconstruction of 3D models from real environment," *ISPRS Journal of Photogrammetry & Remote Sensing*, Vol. 54, pp. 1-22, 1999.
8. M. A. Abidi, M. Abdulghafour and T. Chandra, "Fusion of Visual and Range Features Using Fuzzy Logic," *Control Engineering Practice*, Vol. 2, No. 5, pp. 833-847, 1994.
9. I. S. Chang and R. H. Park, "Segmentation based on fusion of range and intensity images using robust trimmed methods," *Pattern Recognition*, Vol. 34, pp. 1951-1962, 2001.
10. K. Umeda, K. Ikushima, and T. Arai, "Fusion of range image and intensity image for 3D shape recognition," *Proceedings of IEEE Int. Conf. Robotics and Automation*, pp. 680-685, 1996.
11. K. Tate and Z. N. Li, "Depth Map Construction from Range-guided Multiresolution Stereo Matching," *IEEE Trans. Systems, Man, and Cybernetics*, Vol. 24, No. 1, 1994.
12. D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *International Journal of Computer Vision*, Vol. 47, pp. 7-42, 2002.
13. H.J. Zimmermann, *Fuzzy set theory and its applications 2nd ed.*, Kluwer Academic Publishers, 1991.