A Novel Robot Sensor System Utilizing the Combination Of Stereo Image Intensity And Laser Structured Light Image Information

Hyun-Ki Lee*, Song Xingyong*, Min-Young Kim** and Hyungsuck Cho*

* Department of Mechanical Engineering, KAIST, Daejeon, Korea

(Tel : +82-42-869-3253; E-mail: {hklee, songxy, hscho}@lca.kaist.ac.kr)

**Department of Surface Mounting Technology, Mirae Co., Baekseok-dong, Cheonan, Korea

(Tel : +81-42-869-3253; E-mail: mykim@lca.kaist.ac.kr)

Abstract: One of the important research issues in mobile robot is how to detect the 3D environment fast and accurately, and recognize it. Sensing methods of utilizing laser structured light and/or stereo vision are representatively used among a number of methodologies developed to date. However, the methods are still in need of achieving high accuracy and reliability to be used for real world environments. In this paper to implement a new robotic environmental sensing algorithm is presented by combining the information between intensity image and that of laser structured light image. To see how effectively the algorithm applied to real environments, we developed a sensor system that can be mounted on a mobile robot and tested performance for a series of environments.

Keywords: Stereo Vision, Laser Structured Light, Information Combination, Mobile Robot

1. INTRODUCTION

A major research issue for mobile robots is to develop an effective environment sensing and recognizing system. Among them the binocular or multi vision sensor system has been widely used as representative ones of passive visual sensors. But they still have some problems due to the illumination noise, insufficient feature information in environment composed of plain surfaces, and correspondence problem between multiple images. These reasons have led 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 can be observed clearly 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 to get more reliable range information, and have a promising alternative proposal which includes 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 as the laser visual sensor and the laser range finder measuring the time-of-flight. The laser range finder has more advantages in views of measuring range and accuracy, but it still has some problems such as high hardware cost, high power consumption, heavy weight, and low accuracy in near range. In addition, it needs many scanning procedures. This scanning procedure is a time consuming task to limit the sensing time, and also 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 was necessary to develop a new visual sensor system, which combined the laser structured light method and stereo vision method mentioned above.

In our previous work, we proposed 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 [5, 14]. In case when 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. However, in case when 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 other 3D range information based on stereo vision theory or for supporting the robustness of the active trinocular vision sensor.

Recently, several researchers have performed researches of combining laser range data and intensity images: (1) texture mapping on the range data and registration for realistic 3D modeling [6, 7], (2) efficient edge searching 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 the depth map acquisition with high resolution from a low-resolution laser range image and a stereo pair of high-resolution intensity images.

In this paper, we propose a novel information combination method improving the correspondence matching problem of the stereo vision via the information from the active vision. The mobile robot used in this paper is shown in Figure 1. For autonomous navigation for unknown space, it is equipped with a number of sensors, e.g., ultrasonic sensors, infrared sensors, and the proposed sensor head shown in Figure 2. During the navigation process, it perceives the navigation environment by using these sensors. To obtain the detailed 3D range information on environment, the proposed sensor head mounted on a pan-tilt unit is utilized, which enables the cameras to change the viewing direction freely.

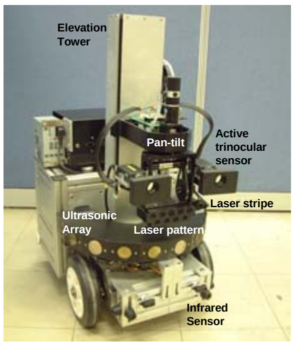


Fig. 1 An autonomous mobile robot, LCARIII

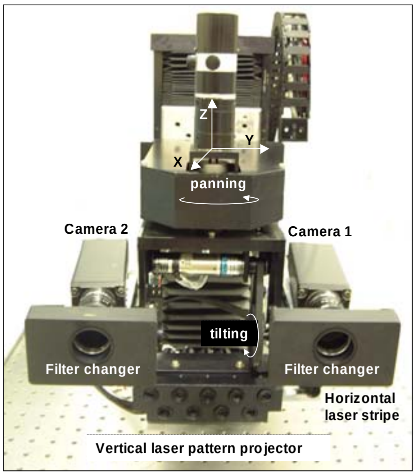


Fig. 2 Sensor head for 3D environment perception

2. STEREO VISION

When a stereo vision is utilized in this sensor system among various local area methods, a range information can be extracted from the acquired intensity images. However, the well-known correspondence problem still makes this range extracting process difficult. The related researches can be roughly classified into local method and global method [15]. For local methods a representative one utilizes the area-based intensity correlation between the right and left images. Among a variety of similarity measures between two windows specified in the left and right images, a simple measure called SAD (sum of absolute differences), SSD (sum of squared differences), NCC (normalized cross correlation), or SMW (symmetric multiple window) has been widely used. Although they are efficient, they are sensitive to local intensity noise (occlusion or uniform textured region). In contrast, global methods are less sensitive to these problems since global constraints such as depth continuity constraint provide additional support for these ambiguous regions. However, they may cause in expensive computational costs. As a global method, dynamic programming and graph cuts have been frequently used. A detailed description for current trend of stereo vision research can be found in the references [12, 15].

In the work related to this paper, a dynamic programming method [16] is used, which focuses on the range information acquisition. This method has an advantage of being able to handle large untextured regions which are frequently encountered in navigation environments. It performs an optimization process at each corresponding epipolar lines of two left and right images respectively (same rows at each intensity image). Before the optimization, the rectification needs to be applied to make the corresponding epipolar lines parallel within each image frame. By scanning each pixel in the epipolar lines located in the same row, the optimization process is performed with the following cost function:

$$J(m) = N_{oc}\lambda_{oc} - N_m\lambda_r + \sum_{i=1}^{N_m} d(x, y) \tag{1}$$

where *J(m)* is the cost function of the m^{th} row, λ_{oc} is the constant occlusion penalty, λ_r is a constant match reward, $d(x, y) = |I_L(x) - I_R(y)|$ is the intensity dissimilarity between candidate pixel x in the left image and candidate pixel y in the right image at each m^{th} row scan line, and N_{oc} and N_m are the number of occlusions and matches, respectively. It is noted that if occlusion and reward are not considered J contains only the dissimilarity $d(x, y)$, representing local area methods as indicated in the third term.

To carry out the minimization of J defined in the above, the following procedure is taken:

1. We form an array composed of row pixels of the right image × row pixels of the left image.
2. We select a set of pixels that may be indicative of correspondence candidates lying within a designer-defined maximum disparity.
3. Finally, we choose a set of cells that yield a minimum value of *J*.

In this way, we obtain the depth information on the epipolar line. By repeating the whole epipolar line within the image frame, we obtain the dense depth map over the captured image.

3. THE SENSOR INTEGRATION ALGORITHM

3.1 The architecture of combination algorithm

Figure 3 shows the concept of proposed sensor integration algorithm The information from laser structured light image takes a role of acquiring 3D dense depth map from stereo intensity images with assistance of relatively accurate but sparse range data. This architecture makes stereo vision more robust for highly ambiguous regions due to geometric occlusion and untextured object surface.

The DP approach [16] mentioned above can handle a few untextured regions, but it may not be effective for large untextured regions. Inherently local area method of stereo vision algorithm has this weak point, because it just considers the intensity difference between each candidate pixel in left and right images during matching process. To overcome this difficulty more reliable range data from active vision play a role of constraint which can give a relationship between neighborhood rows for optimization method of dynamic programming.

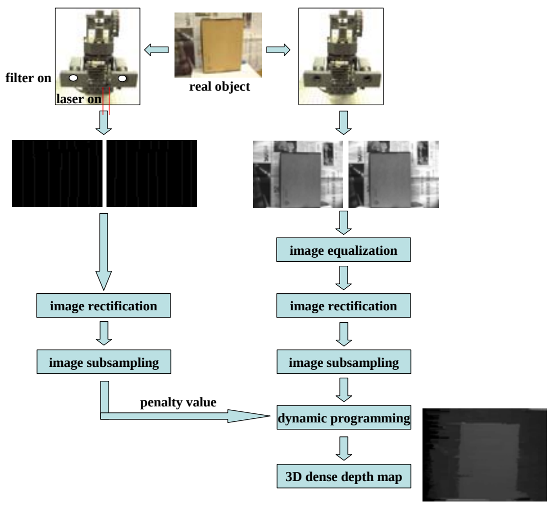
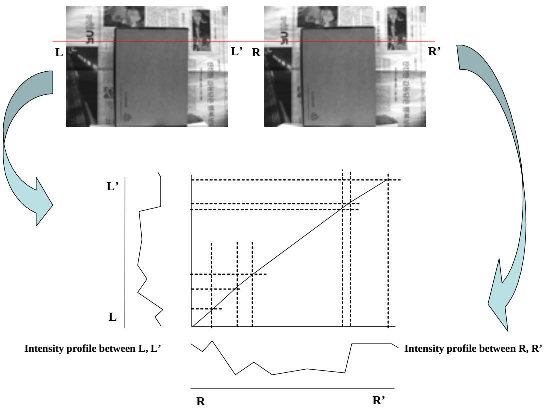


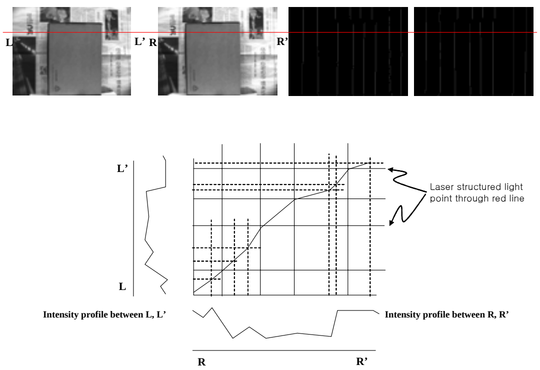
Fig. 3 Conceptual diagram of a proposed algorithm

3.2 The proposed integration algorithm

The stereo vision algorithm using intensity images can make a dense depth map at once, but the accuracy is generally lower than that obtained by the active vision sensor. In other words, active vision algorithms using laser structured light can make a depth map more accurately. But active vision algorithms can get a sparse depth information due to the feature of laser structured line. If we want to make a dense depth map, it needs some scan process using moving mechanism. According to integration of these two sensors we can make a more accurate and fast range information.



(a) Original DP algorithm



(b) Proposed integration algorithm

Fig. 4 Original DP algorithm vs. proposed integration algorithm

First, we choose one row which has correspondence at each left and right intensity image as shown in Figure 4. Original algorithm is dealing with whole intensity information as shown in Figure 4, but here for simplification we assume each row of each intensity image consists of ten pixels and disparity maximum value is four. Although originally stereo vision algorithm deals with at the whole rows of each intensity image, here we just explain the algorithm by taking one row for example because the algorithm for one row can be adapted to other rows as the same way.

Second, we get laser structured light information at the same row which is chosen in the previous step. We assume laser structured line position at the left laser structured light image is the fifth pixel and one at the right laser structured light image is the third as shown in top figures of Figure 5. In this figures the white grids indicate the laser structured light position.

Next, we can make a 10 by 10 matrix, in which x axis indicates pixels of the right row, and y axis indicates pixels of the left row chosen at the first time. The matrix shape is shown in Figure 4, where white pixels indicate the candidate pixels to be matched and black pixels indicate the non-candidate pixels. As we assume the maximum disparity is 4 initially, we can see that in the figure the number of white pixels in each column of the matrix is 4.

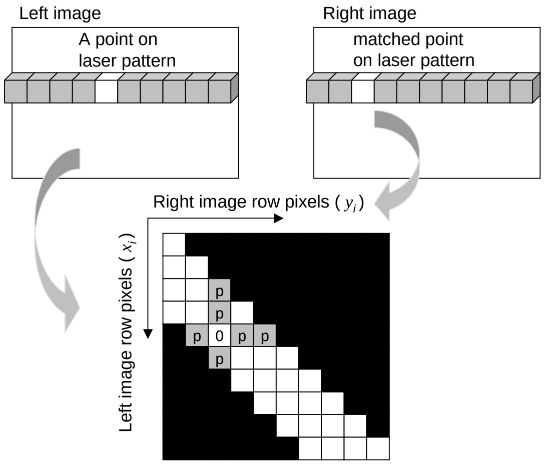


Fig. 5 A proposed method of assigning penalty values in dynamic programming in case of combining laser information to stereo information

Next, we define modified cost function for proposed integration algorithm from Eq (1).

$$J(m) = N_{oc}\lambda_{oc} - N_m\lambda_r + \sum_{i=1}^{N_m} d\left(x,\ y\right) + P(x,y) \qquad (2)$$

where $P(x,y)$ penalty value between candidate pixel x in the left image and candidate pixel y in the right image at each m^{th} row scan line

Finally, we try to find the depth information from the candidate pixel using dynamic programming. Originally in the dynamic programming process, at each column and row we choose one candidate pixel (i.e. (y, x) = <(1,1) (2,3) (3,5) · · >) from white pixel in the figure, which satisfies the modified cost function defined in Eq. (2). In the integration algorithm, the pixel which locates in the intersection between right laser structured light point and left laser structured light point (right laser structured light point, left laser structured light point) assigns reward value for the cost function, and other pixels which locate (*, left laser structured light point), or (right laser structured light point, *) give big penalty value. For example, at the previous situation of (right laser structured light point, left laser structured light point)=(3,5), the pixel (3,5) takes 0 value as reward value, other pixels<(3,3) (3,4) (3,6) (2,5) (4,5) (5,5)> have a big penalty value. By doing this, the pixel (right laser structured light point, left laser structured light point) always include range data set which satisfy the cost function and it influences the range data sets for neighborhood pixels.

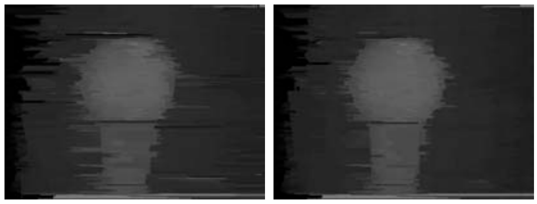
4. EXPERIMENTS FOR THE PROPOSED ALGORITHMS

In order to demonstrate the performance of the proposed algorithm, we conducted a series of experiments under several environment configurations. In this work, the environments are classified into those obtained by combinations of untextured or textured background, and untextured or textured objects. Objects and background planes are placed at the distance range of about 0.7m in front of the sensor system mounted on the mobile robot. The size of the acquired dense map was fixed within 640x480 pixels. The dynamic programming (DP), we used for integration purpose as described in Figure 3, is loaded from Open CV Library developed by Intel[17] was merged into our algorithms.

Figure 6 shows the results obtained from environment 1 composed of an object (foot ball) and a randomly textured background, which is shown in Figure 6(a). The image containing laser structured stripes is depicted in Figure 6(b). The two results in Figure 6(c) and (d), one obtained by DP and the other by the proposed algorithm show a considerable difference in depth map. This kind of trend can also be observed from the experimental results obtained from environment 2 which consists of a rectangular block and the same textured background as used in the above experiment. The backgrounds were composed differently from the previous one in order to investigate the effectiveness of the algorithm case relatively to the DP algorithm. Figures 8 and 9 illustrate such backgrounds where two objects are placed in the vicinity of each other in the same untextured background rather than the textured one. As we can see from both figures, the proposed algorithm works much better than DP algorithm. In the case of the proposed, we obtained rather accurate depth map whereas DP cannot produce recognizable depth map. The reason may be due to the fact that the algorithm may not work well in the environment having no appreciable gray difference. In Table 1, standard deviation of error has more important meaning than average of error. The error value is defined as difference between real depth value and reconstructed depth value. While the average of error shifts its position slightly, the standard deviation tends to decrease with helps of the proposed algorithms. Experimental results indicate almost 30~70% improvement in the standard deviation of the error.

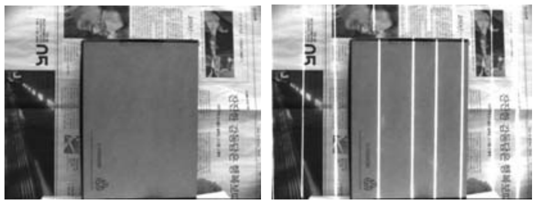
In order to observe the error characteristics due to algorithm parameter, two experiments were performed. First, Figure 10 shows how the penalty value assigned to unmatched cells in the search grid, which is a key parameter in the proposed fusion algorithms, affects the final disparity map. From above-mentioned environments, with variations of the penalty value, statistical error characteristics were investigated. Obtained experimental results show that a value between 500 and 900 is proper for the penalty value. It is noticeable that higher penalty value always does not lead lower depth error and there is a certain band of the penalty value that makes the depth error lower. Second, Figure 11 shows how variations of the occlusion penalty and the match reward values in dynamic programming affect final disparity maps. On this contour map, statistical error characteristics are represented as the summation of error standard deviations for the whole experimental results with their variations. In λ_{oc} and λ_r coordinates, it is found that (10, 15) and (5, 25) shows two separated local minimums.

a) left image from stereo vision b) left image with laser line

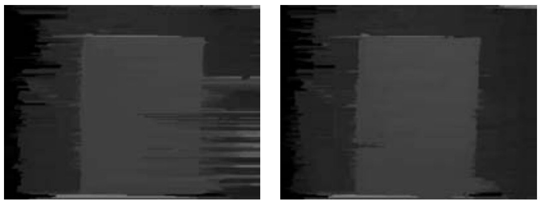


c) range image by dynamic programming d) range image by the proposed integration algorithm

Fig. 6 Experimental result for test environment 1

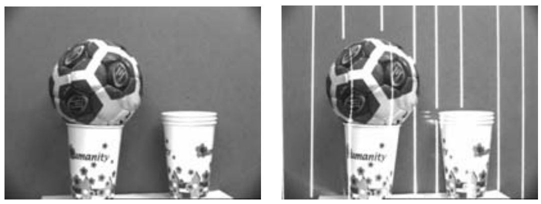


a) left image from stereo vision b) left image with laser line

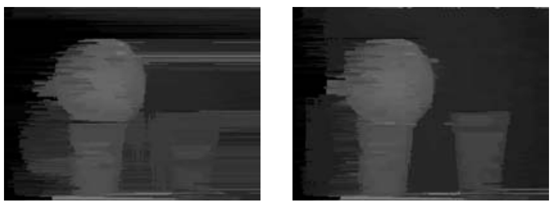


c) range image by dynamic programming d) range image by the proposed integration algorithm

Fig. 7 Experimental result for test environment 2

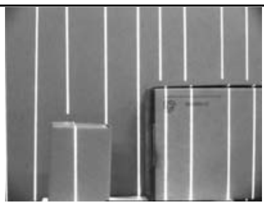


a) left image from stereo vision b) left image with laser line



c) range image by dynamic programming d) range image by the proposed integration algorithm

Fig. 8 Experimental result for test environment 3



a) left image from stereo vision b) left image with laser line

c) range image by dynamic programming d) range image by the proposed integration algorithm

Fig. 9 Experimental result for test environment 4

Table 1 Error statistics between ground truth and experimental result.

	dynamic programming [16]		Proposed fusion algorithms	
	average error (pixel)	standard error deviation	average error (pixel)	standard error deviation
Experiment 1 (Fig. 6)	0.7	114.2	1.3	83.2 27%improvement
Experiment 2 (Fig. 7)	1.2	148.9	2.8	52.9 64%improvement
Experiment 3 (Fig. 8)	0.8	336.9	1.8	86.7 74%improvement
Experiment 4 (Fig. 9)	14.5	391.1	2.5	104.3 73%improvement

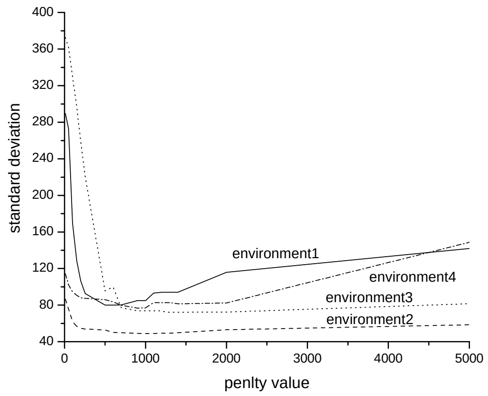


Fig. 10 Error standard deviation with variation of the penalty value in the proposed algorithm

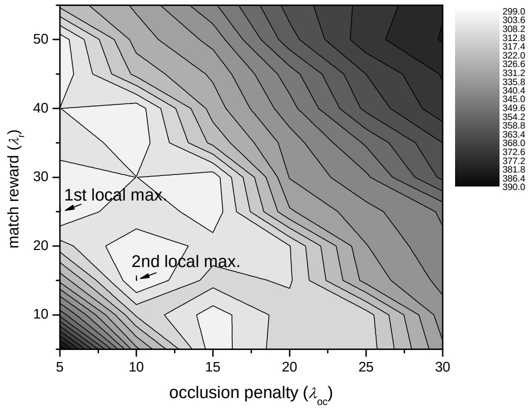


Fig. 11 Summation of error standard deviations for all experimental cases with variations of the occlusion penalty and the match reward in dynamic programming (contour plot)

5. CONCLUSION

For the mobile robot to perform a task and navigation accurately and robustly environment recognition is very important. In this paper by introducing a novel sensor combination method for this visual sensor system acquiring more reliable range information, using active and passive information simultaneously, we can get a accurate and robust result. Since each of two sensors has its own advantages and disadvantages on the measurements of various environments, the proposed sensor combination algorithm is essentially needed for robot sensing system. The acquired information from active vision sensor can be used for dense stereo-matching in stereo vision.

To see how the proposed algorithms can be applied to real applications, we applied it to the sensor system mounted on a mobile robot. The performed experiments show that the proposed sensing algorithm can successfully and robustly make a depth map at any environments. The contribution of this paper is summarized as follows:

1. A simple and efficient combination algorithm was proposed for laser structured light image and intensity image information.
2. Due to the proposed combination algorithms that apply additional constraint of vertical constancy to the general dynamic programming method, it is shown that horizontal streaks in general dynamic programming methods are reduced largely in each disparity image.
3. In case when objects are placed in front of untextured planes, sandwiched regions between objects frequently fall into serious disparity ambiguity with DP. But using the proposed algorithm we can get a almost reliable result.

Now, this research is going on progress, and our research is focused on three-dimensional map building system using proposed algorithm for mobile robots.

ACKNOWLEDGMENTS

This work was supported (in part) by the Ministry of Information & Communications, Korea, under the Information Technology Research Center (ITRC) Support Program.

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.
[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, pp. 134-144, 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 Adademic Publishers, 1991.
[14] M. Y. Kim, H. S. Cho and H. K. Lee, ““Novel visual sensor system fusing an active trinocular vision and a stereo vision for 3D range data acquisition,” Proceedings of SPIE Vo. 5603, Machine Vision and its Optomechatronic Applications, Philadelphia, Pennsylvania, USA, pp. 31-39, 2004.
[15] M. Z. Brown, D. Burschka, and G. D. Hager, “Advances in computational stereo,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 25, No. 8, pp. 993-1008, 2003.
[16] S. Birchfield and C. Tomasi, “Depth discontinuities by pixel-to-pixel stereo,” International Journal of Computer Vision, Vol. 35, No. 3, pp. 269-293, 1999.
[17] http://www.intel.com/research/mrl/research/opencv