FUSION OF TIME-OF-FLIGHT CAMERA AND STEREO CAMERA DATA FOR ENHANCED 3D WORK ENVIRONMENT REPRESENTATION FOR CONSTRUCTION EQUIPMENT AUTOMATION

Hyojoo Son, Gyomin Lim, and Changwan Kim*

Department of Architectural Engineering, Chung-Ang University, Seoul, Korea
** Corresponding author (changwan@cau.ac.kr)*

ABSTRACT: As the need to develop semiautomated and automated construction equipment increases in the construction industry, 3D work environment representation becomes increasingly important. Accurate, realistic, and rapid 3D representation of the work environment can be utilized for obstacle detection and collision avoidance by providing information about the state of the work environment. This paper describes a system to create a 3D graphical representation of the work environment for safe and efficient equipment operation that has desirable properties and satisfies the constraints on the development of semiautomated and automated construction equipment. For this purpose, a unified framework for a multi-sensor data fusion-based 3D work environment representation system is proposed. The proposed framework consists of four main steps: data acquisition, data filtering, data fusion, and 3D graphical model generation. A field experiment was undertaken on a construction site to validate the proposed system. The preliminary research results show that the resulting 3D graphical representation of the work environment can be successfully employed in the development of construction equipment that assists the operator in identifying obstacles or operates independently.

Keywords*: 3D Work Environment Representation, Construction Eequipment, Flash LADAR, Semiautomated and Automated Equipment, Multi-sensor Data Fusion, Stereo Camera*

1. INTRODUCTION

Construction relies to a large extent upon a wide range of heavy excavation and earthmoving equipment [1]. The most construction equipment is hand-operated at construction sites, where numerous construction workspace hazards and potentially dangerous situations occur. In addition, a high number of those incidents are directly caused by operator error [2]. On the other hand, semiautomated and automated equipment can improve the safety and efficiency of construction equipment operation by allowing operators to avoid physical exposure to the dangerous and hazardous construction environment or function without human involvement. Given these advantages, semiautomated and automated equipment have great potential for applications not only at typical construction sites, but also in hazardous environments into which it is difficult or impossible for humans to enter, such as mining environments, nuclear waste facilities, restoration work after a catastrophe, and space.

For the operation of semiautomated and automated equipment, a three-dimensional (3D) representation of the terrain and obstacles that might be in close proximity to the equipment is indispensable for planning to perform tasks in dynamically changing environments on their own. During the operation of semiautomated and automated equipment, the equipment is partially or fully planned and controlled by itself. Thus, the work environment needs to be represented accurately and sent to the perception system of the equipment rapidly in order to that the system understands and analyze the surrounding environment.

There are several requirements that a 3D work environment representation system must meet to improve the operation of semiautomated and automated equipment. First, the acquired data should be accurate and have sufficient resolution. A high degree of accuracy and resolution allows for a precise understanding and analysis of the work environment. Second, the acquired data should have color information. The use of color information can be incorporated to recognize and analyze terrain and obstacles with 3D information. Third, the acquisition of data should be fully automated and updated quickly enough to permit the planning and execution of the tasks by accounting for the speed of the construction machines.

Recently, researchers have adopted a flash laser distance and ranging (LADAR) to represent the state of the work environment [3,4,5]. The flash LADAR is a new generation of active 3D sensors. It enables the acquisition of 3D data from an entire scene at 30 frames per second [6]. Although the previous research works validated that flash LADAR could be used to capture moving objects and provide both static and dynamic information in laboratory experiments, test beds, or construction sites, these studies have limitations. The resolution of 3D data acquired from flash LADAR is limited. In addition, the captured data are typically contaminated by noise, especially in outdoor environments [7]. These disadvantages have not been fully considered in previous research in the construction industry, and it has limited its widespread use in representing the work environment for construction equipment operation.

To overcome these limitations, researchers in 3D computer vision, computer graphics, and human-computer interactions have investigated multi-sensor data fusion method [8,9,10,11,12,13]. Multi-sensor data fusion is defined as the process of combining data acquired from two or more sensors into a single unified data set. In particular, the fusion of data acquired from different types of sensors, such as laser-based sensors and image-based sensors, eliminates the disadvantages of using sensors alone by complementing each other. By integrating the low-resolution and noisy range data from flash LADAR with one or more high-resolution color images from image-based sensors, such as a charged-coupled device (CCD) camera, stereo camera, or video camera, it is possible to produce high resolution and accurate range data. Moreover, it has the advantage of relatively low-cost and compact combination of sensors. Multi-sensor data fusion is the most promising method at achieving accurate, realistic, and rapid 3D representation of the surrounding environment, which is required for the development of semiautomated and automated equipment.

The aim of this study is to develop a multi-sensor data fusion-based 3D work environment representation system to create a 3D graphical representation of the work environment for enhanced equipment operation that has desirable properties and satisfies the constraints on the development of construction equipment. For this purpose, a unified framework for multi-sensor data fusion-based 3D work environment representation system is proposed. The proposed framework consists of four steps: data acquisition, data filtering, data fusion, and 3D graphical model generation. A field experiment was undertaken on a construction site to validate the proposed system.

2. MULTI-SENSOR DATA FUSION-BASED 3D WORK ENVIRONEMENT REPRESENTATION

The main focus of this paper lies in the development of an accurate, realistic, and rapid 3D work environment representation system that will increase the efficiency and safety of semiautomated and automated construction equipment operation. The proposed framework consists of four steps: data acquisition, data filtering, data fusion, and 3D graphical model generation. In this section, a detailed description of the proposed process will be given.

2.1 Multi-Sensor Data Acquisition System

In this study, the laser-based sensor used was a SwissRanger SR-3000 flash LADAR [14]. It was mounted with an image-based sensor, a Bumblebee® XB3 stereo camera [15]. Although the SwissRanger SR-3000 is an advanced system, it has a low-resolution (176*144). As a result, it is difficult to accurately represent the work environment from the raw data because the low-resolution increases data ambiguity. The image-based Bumblebee® XB3 provides rectified high-resolution (1,280*1,024) color images. In this study, the rectified color images were used to enhance the resolution and improve the sub-pixel accuracy of range data acquired from the SwissRanger SR-3000. Moreover, it was possible to obtain rich information about the work environment because the color information acquired from the stereo camera could be assigned to each range data.

In order to combine the data from different sensors, the calibration process that estimates and verifies intrinsic and extrinsic parameters of different sensors is essential to sync the perspectives of the sensors [16,10]. In this study, for both types of calibration, an algorithm developed by Zhang [17] was used. First, the estimation of intrinsic parameters, such as focal length, radial and tangential lens distortion coefficients, and the projection center was performed. Then, the compensation of radial and tangential distortions was achieved using the estimated intrinsic parameters. Second, the estimation of extrinsic parameters was conducted to determine the relative rotation and translation parameters of the three sensors. This calibration process synced the perspectives of the SwissRanger SR-3000 and the Bumblebee® XB3.

2.2 Data Filtering

Range data obtained from the flash LADAR contain considerable noise [19]. In an unstructured outdoor environment, range data acquired via flash LADAR include at least two types of noise: dropout, which occurs when no return signal is received because there is no object, and speckle noise, which is primarily caused by environmental conditions, such as dust and rough surfaces of objects [20,21]. Such noise might have a negative effect on the quality of a 3D geometric model and data fusion results; consequently, the data must be filtered to eliminate the unnecessary noise as the first step in processing. To this end, an average difference value filtering and a median filtering were used [5].

2.3 Data Fusion

Although the noise in the original range data is reduced after the data filtering process, the low-resolution (176*144) can corrupt the environment or object feature information such as edges or surfaces. In addition, the range data have no color or texture information, which is required to readily assimilate the real-world environment. Therefore, a data fusion process that employed joint bilateral upsampling algorithm [22] was proposed.

Joint bilateral upsampling algorithm is a variation of the bilateral filter originally proposed by Tomasi and Manduchi [23]. It has gained popularity as the sensor data fusion technique since it preserves the boundary of up-sampled range data accurately [10,11]. Before describing the joint bilateral upsampling algorithm, a brief description of the bilateral filter will be given here.

The bilateral filter is an edge-preserving smoothing filter for intensity. The basic idea of how to preserve edges is that it takes a weighted average of local pixels. The weight of each pixel is calculated by its distance in both the square window and the range space. The bilateral filter can be expressed as follows:

$$J_p = \frac{1}{k_p} \sum_{q \in \Omega} I_p f(\|p - q\|) g(\|I_p - I_q\|)$$

where I_p is the intensity of pixel p in the input image I, Ω is a spatial neighborhood surrounding p, J_p is the value of pixel p in the filtered image J, and k_p is a normalizing factor. In this expression, $f(\|p-q\|)$ is a data term, whereas $g(\|I_p - I_q\|)$ is a range term. In addition, f is the spatial filter kernel, and g is the range filter kernel. In this study, f is taken to be the Gaussian function centered at p with standard deviations σ_s. Further, g is taken to be the Gaussian function with standard deviations σ_r.

Joint bilateral upsampling assumes that objects with similar intensity often have similar depths in a scene. It can be used to enhance a low-resolution range image I to the resolution of a high-resolution intensity image $\tilde{I}$ by considering the high-resolution guidance image taken from the same scene. Instead of evaluating both components of the filter on I, the high-resolution reference image is used in the range term, which enables the edge-preserving upsampling of I to the resolution of $\tilde{I}$. Crabb et al. [24] extended the use of joint bilateral upsampling to high-resolution color image. The distance between colors is calculated as a Euclidean distances in the RGB color space. To achieve this goal, the upsmapled image $\tilde{S}_p$ is computed as follows:

$$\tilde{S}_p = \frac{1}{k_p} \sum_{q_\downarrow \in \Omega} I_p f(\|p_\downarrow - q_\downarrow\|) g(\|\tilde{I}_p - \tilde{I}_q\|)$$

where $p_\downarrow$ and $q_\downarrow$ are pixels in the low-resolution image p and q is the corresponding pixels in the high-resolution intensity or color image.

2.4 3D Graphical Model Generation

In the data fusion process, in addition to resolution enhancement and accuracy improvement, the data fusion result produces range data for every color pixel in color images by combining the range data and color images at the pixel level. Therefore, a realistic 3D graphical model can be generated by creating a geometric 3D model using range data and painting the model using the color information.

The resulting enhanced range data is subjected to a mesh generation process to generate a geometric 3D model of the entire scene. A mesh representation has the benefit of providing a highly detailed representation of the environments or objects from a set of range data. In this study, the Delaunay triangulation algorithm was employed to generate an initial surface mesh. It generates a triangular mesh that can be used to represent complex environment [25,26]. A geometric model generation of the entire scene from the resulting range data is performed. The surface texture is computed for every triangle in the geometric model in the form of meshes. The initial mesh model is then painted for the purpose of achieving a realistic visualization model.

3. FIELD EXPERIMENTAL RESULTS

The proposed multi-sensor data fusion-based 3D work environment representation system was validated by conducting a field experiment on an actual construction site, which is an excavation process. The multi-sensor data acquisition system was mounted on the top of the cockpit of the machine. Fig. 1(a) shows a construction site scene and Fig. 1(b) shows the operator's view of the work environment.

(a) (b)

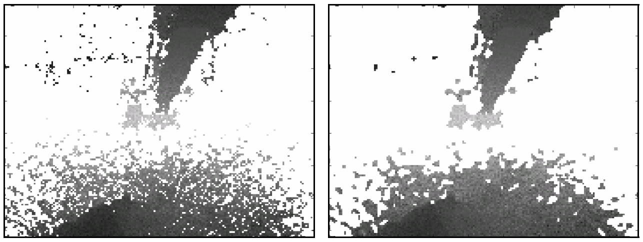
Fig. 1 (a) Construction site scene, (b) Operator's view of the work environment

Fig. 2 shows the data acquired from the calibrated multi-sensor data acquisition system. Fig. 2(a) shows a low-range data and Fig. 2(b) shows a rectified high-resolution color image.

(a) (b)

Fig. 2 (a) Range data (176*144), (b) Rectified color image (1,024*768)

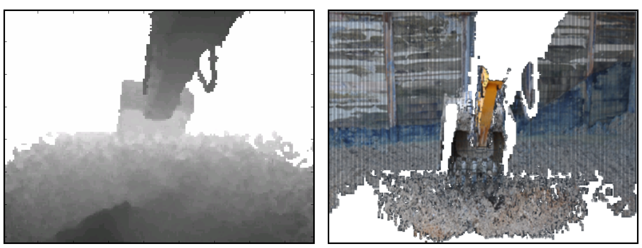
Fig. 3 shows the result of the data filtering process. Throughout the process, dropout and speckle noises in the range data were effectively eliminated and smooth surfaces were formed. However, some data corresponding to objects are partially removed together with noise in the data filtering process since that data are corrupted by noise.



(a) (b)

Fig. 3 Result of the data filtering process (a) Result of dropout noise reduction using average difference values, (b) Results of speckle noise reduction by median filtering

Fig. 4 shows the result of the data fusion process. After upsampling from the low-resolution range data, the range data have a resolution of 1,024*768. Fig. 4(a) shows the enhanced range data and Fig. 4(b) shows the synthesized view of the colorized range data. The quality has been improved significantly and, by visual comparison, the difference between Fig. 2(a) and Fig. 4(a) is clear.



(a) (b)

Fig. 4 Results of the data fusion process (a) Enhanced range data (1,024*768), (b) Synthesized view of colorized range data

As the final step, the 3D work environment was graphically represented, as shown in Fig. 5. The resulting model shows how the proposed method can be applied to accurately and realistically represent the real-world environment.

Fig. 5 Final 3D work environment representation

4. CONCLUSIONS

This study proposed and described a system to accurately, realistically, and rapidly represent the construction work environments in a 3D format. For this purpose, the multi-sensor data fusion system comprised of the flash LADAR and the stereo camera was designed. The proposed system created an accurate and realistic representation of the real construction environment using the fused range data from the flash LADAR and color images from the stereo camera. The preliminary field experimental results show that combining data from multiple, complementary sensors enhances the quality of the information available to the perception system of the construction equipment and makes human-machine interaction more efficient. The resulting 3D graphical representation of the work environment can be successfully employed in the development of construction equipment that assists the operator in identifying obstacles or operates independently. In short, sensor fusion offers the possibility of significantly improving the efficiency and safety of construction equipment operation.

Future work will focus on optimization of the algorithms for effective data filtering, data fusion, and 3D graphical model generation. In addition, experiments will be conducted on other construction sites and work environments having different characteristics to validate the feasibility of the proposed method.

ACKNOWLEDGEMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0023229).

REFERENCES

[1] Tichon, J. and Diver, P., “Plant Operator Simulation: Benefits and Drawbacks for a Construction Training

Organization”, *Cognition, Technology & Work*, Vol. 12(3), pp. 219–-229, 2010.
[2] Kim, C., Haas, C.T., and Liapi, K.A., “Rapid, On-Site Spatial Information Acquisition and Its Use for Infrastructure Operation and Maintenance”, *Automation in Construction*, Vol. 14(5), pp. 666–684, 2005.
[3] Teizer, J., Caldas, C.H., and Haas, C.T., “Real-Time Three-Dimensional Occupancy Grid Modeling for the Detection and Tracking of Construction Resources”, *Journal of Construction Engineering and Management*, Vol. 133(11), pp. 880–888, 2007.
[4] Gong, J. and Caldas, C., “Data Processing for Real-Time Construction Site Spatial Modeling”, *Automation in Construction*, Vol. 17(5), pp. 526–535, 2008.
[5] Son, H., Kim, C., and Choi, K., “Rapid 3D Object Detection and Modeling Using Range Data from 3D Range Imaging Camera for Heavy Equipment Operation”, *Automation in Construction*, Vol. 19(7), pp. 898–906, 2010.
[6] Chan, D., Buisman, H., Theobalt, C., and Thrun, S., “A Noise-Aware Filter for Real-Time Depth Upsampling”, *Proc., ECCV Worksh. on Multi-Camera and Multi-Modal Sensor Fusion Algorithms and Applications*, Marseille, France, 2008.
[7] Xiang, X., Li, G., Tong, J., and Pan, Z., “Fast and Simple Super Resolution for Range Data”, *Proc., Int. Conf. on Cyberworlds*, Singapore, Singapore, 2010.
[8 Beder, C., Bartczak, B., and Koch, R., “A Combined Approach for Estimating Patchlets from PMD Depth Images and Stereo Intensity Images”, *Lecture Notes in Computer Science*, Vol. 4713, pp. 11–20, 2007.
[9] Zhu, J., Wang, L., Gao, J., and Yang, R., “Spatial-Temporal Fusion for High Accuracy Depth Maps Using Dynamic MRFs”, *IEEE Transactions on Consumer Electronics*, Vol. 32(5), pp. 899–909, 2010.
[10] Yang, Q.X., Yang, R.G., Davis, J., and Nister, D., “Spatial-Depth Super Resolution for Range Images”, *Proc., Computer Vision and Pattern Recognition*, Minneapolis, MS, 2007.
[11] Choi, J., Min, D., and Sohn, K., “2D-Plus-Depth Based Resolution and Frame-Rate Up-Conversion Technique for Depth Video”, *IEEE Transactions on Consumer Electronics*, Vol. 56(4), pp. 2489–2497, 2010.
[12] Huhle, B., Schairer, T., Jenke, P., and Straßer, W., “Fusion of Range and Color Images for Denoising and Resolution Enhancement with a Non-Local Filter”, *Computer Vision and Image Understanding*, Vol. 114(12), pp. 1336–1345, 2010.
[13] Yang, Q., Tan, K., Culbertson, B., and Apostolopoulos, J., “Fusion of Active and Passive Sensors for Fast 3D Capture”, *HP Laboratories Technical Report*, Issue 97, 2010.
[14] Mesa Imaging AG, http://www.mesa-imaging.ch/, last accessed on March 31, 2011.
[15] Point Grey Research, Inc., http://www.ptgrey.com/, last accessed on March 31, 2011.
[16] Hahne, U. and Alexa, M., “Depth Imaging by Combining Time-of-Flight and On-Demand Stereo”, *Proc., DAGM Worksh. on Dynamic 3D Imaging*, Jena, Germany, 2009.
[17] Zhang, Z., “A Flexible New Technique for Camera Calibration”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22(11), pp. 1330–1334, 2000.
[19] Hedge, G.M., and Ye, C., “SwissRanger SR-3000 Range Images Enhancement by a Singular Value Decomposition Filter”, *Proc., IEEE International Conference on Information and Automation*, Zhangjiajie, China, 2008.
[20] Esselman, T.R., and Verly, J.G., “Some Applications of Mathematical Morphology to Range Imagery”, *Proc., IEEE International Conference on International Conference on Acoustics, Speech, and Signal Processing*, Dallas, TX, 1987.
[21] Wang, Q., Li, Q., Chen, Z., Sun, J., and Yao, R., “Range Image Noise Suppression in Laser Imaging System”, *Optics and Laser Technology*, Vol. 41(2), pp. 140–147, 2009.
[22] Kopf, J., Cohen, M.F., Lischinski, D., and Uyttendaele, M., “Joint Bilateral Upsampling”, *ACM Transactions on Graphics*, Vol. 26(3), pp. 96-1–96-5, 2007.
[23] Tomasi, C., and Manduchi, R., “Bilateral Filtering for Gray and Color images”, *Proc., Int. Conf. on Computer Vision*, Bombay, India, 1998.
[24] Crabb, R., Tracey, C., Purank, A., and Davis, J., “Real-Time Foreground Segmentation via Range and Color Imaging”, *Proc., CVPR Worksh. on Time-of-flight Computer Vision*, Anchorage, Alaska, 2008.
[25] Attali, D., “R-Regular Shape Reconstruction from Unorganized Points”, *Proc., 13rd Annual Symp. On Computational Geometry*, Nice, France, 1997.
[26] Amenta, N. and Bern, M., “Surface Reconstruction by Voronoi Filtering”, Discrete and Computational Geometry, Vol. 22(4), pp. 481–504, 1999.