REALISTIC 3D MODELING SYSTEM FOR REMOTE CONTROLLED CONSTRUCTION MACHINES: MULTI-SENSOR FUSION BASED APPROACH

Hyojoo Son
Chung-Ang University, Seoul, Korea
hjson0908@wm.cau.ac.kr

Changwan Kim
Chung-Ang University, Seoul, Korea
changwan@cau.ac.kr

Abstract

An important feature of remote controlled machines is the realistic three-dimensional (3D) modeling system for providing information feedback to the remote operator and facilitating remote operation. This paper describes a system for creating realistic 3D models of actual construction environments for the purpose of improving and extending the capabilities of interactive visual feedback between the operator and remote controlled construction machines during earthmoving operations. The proposed approach utilizes the fused data from a flash LADAR and a stereo vision system. The modeling process consists of three main steps: data filtering for noise reduction, data fusion for data enhancement, and realistic 3D model generation for visualization purposes. Field experiments were conducted on a construction site to validate the proposed system. Preliminary research results show that the resulting realistic 3D model can be successfully employed for the development of remote controlled construction machines that require interactive visual feedback.

KEYWORDS: construction machine, flash ladar, interactive visual feedback, realistic 3D modeling, remote control, sensor fusion, stereo vision system

INTRODUCTION

During remote operation, which is based on the concept of shared autonomy that distributes intelligence between human and machine, the operator operates the machine while viewing the visual feedback in the monitor using a joystick or manual controller (Hinzinger et al., 1992; Brunner et al., 1994). If the remote operator fails to understand the construction environment, judgment errors may result during operation of the machine. Therefore, to provide better spatial and situational awareness of the remote construction environment to the operator, the real-world environments can be represented in three-dimensional (3D) format on the monitor, thus having the same dimension as the actual environment (Oloufa et al., 2003). This can be accomplished via a realistic 3D modeling system that visualizes the geometry and objects in the operator field of view as realistic 3D models, provides interactive visual feedback to the operator, and enhances the performance of remote machine operation.

There are several requirements that a realistic 3D modeling system must meet to improve the operation displays that are used in remote controlled construction machines. First, the data acquisition method should be applicable to the outdoor environment and during the night. Second, the acquired range data should be accurate and have sufficient resolution and texture information so it will be useful for realistic visualization purposes. Third, the representation form should effectively represent an unstructured environment with complex geometry and objects. Finally, the system should be fully automated and updated within few seconds to react to the operations and account for the speed of the construction machines.

There have been numerous research efforts in the last few decades on remote controlled construction machines (e.g., Singh, 1997; Yamada and Doi, 2008). Despite the increased research activities and remarkable progress that has been achieved, systems are still far from being operational or active. The technology is still at the stages where a pre-defined CAD model is simulated or a typical two-dimensional video monitor is used; there are likely some significant difficulties on providing effective interactive visual feedback where a real 3D work environment is displayed.

To provide the operator with rich information feedback, multi-sensor fusion has traditionally been used to support remote controlled machine applications such as for vehicle teleportation (Meier et al., 1999; Huber et al., 2009). It has been proven as an effective method to provide the operator with information feedback by the synergistic use of the information provided by multiple sensors to compensate other limitations (Csathó et al., 1999). By adopting a multi-sensor fusion approach, it is expected that all the constraints in the development of realistic 3D modeling systems can be satisfied.

The aim of this study is to develop a multi-sensor fusion based realistic 3D modeling system to create a graphical representation of the work environment for enhanced equipment control by providing improved spatial perception to the operator. For this purpose, a unified framework for multi-sensor fusion and realistic 3D modeling is proposed. The proposed framework consists of three steps: data filtering, data fusion, and realistic 3D model generation. Field experiments were undertaken to test the performance of the proposed system.

OVERVIEW OF SYSTEM

The main focus of this paper lies in the development of a realistic 3D modeling system that will increase an operator's situational awareness of the remote construction environment during control of an excavator, thereby facilitating operator performance. Such a realistic modeling system is a prerequisite for the development of remote controlled or teleoperated excavators. Sensors for data acquisition from the work environment are mounted on the top of the cockpit of the machine, thereby providing a realistic 3D model of the work environment, which will provide the same perspective view as if the operator was controlling the machine. Figure 1 shows the excavator task-specific field of view provided by the sensors in the horizontal and vertical directions. The resulting task-oriented 3D model improves remote operation by assisting the operator, as does the graphical control interface. In addition, the model will be utilized in the development of a fully automated excavator that can operate independently.

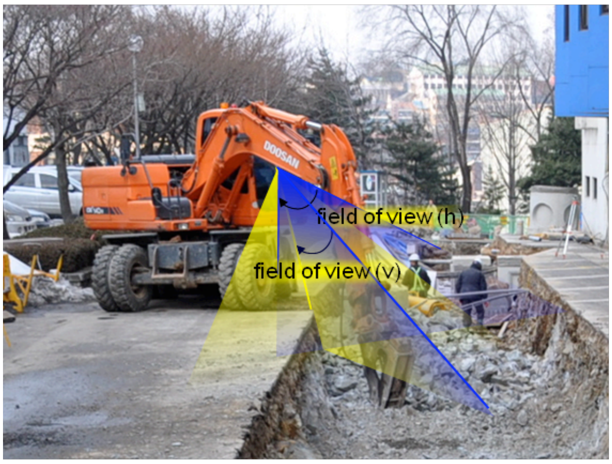


Figure 1: Positions of the sensors equipped within the excavator and the covered work area

The realistic 3D modeling system proposed herein is based on a multi-sensor fusion approach that consists of a laser-based SwissRanger SR-3000 flash LADAR system (Mesa Imaging AG, 2010) and a vision-based Bumblebee® XB3 stereo vision system (Point Grey Research, Inc., 2010). The advances in the relatively inexpensive flash LADAR has enabled the accurate geometric modeling of 3D environments since it is capable of capturing highly accurate 3D data sets of the terrain (Hedge and Ye, 2008). Moreover, it provides range data at a maximum rate of 30 frames per second in real-time (Habbit et al., 2003). However, although the SwissRanger SR-3000 flash LADAR system is the most advanced system, it has a low resolution of 176*144. The result is that it is difficult to accurately represent the work environment from the raw data because the low resolution increases data ambiguity. The vision-based Bumblebee® XB3 stereo vision system provides rectified 1280*1024 pixels color images along with range data in real-time. However, the accuracy of the automated stereo vision system such as the Bumblebee® XB3 stereo vision system is not as good as laser-based sensors such as flash LADAR (Huber et al., 2009), and large amounts of range data can be missing when there are objects without distinctive features (Hedge and Ye, 2008).

To overcome the limitations of each of these two sensors, many studies have focused on a multi-sensor data fusion approach (e.g., Meier et al., 1999; Anderson et al., 2005). The addition of high-resolution imagery from vision-based sensors such as a stereo vision system, CCD sensors, or digital camera with range data from laser-based sensors such as flash LADAR enables enhancement of the resolution of data and sub-pixel accuracy (Yang et al., 2007). In addition, it enables realistic modeling for visualization by providing image and textural information that are not possible using only laser-based sensors (Huber, 2009). Therefore, the range data acquired from stereo vision systems are not used in this study, instead rectified color images are used to enhance the resolution of range data and sub-pixel accuracy acquired from flash LADAR and provide image and textural information for realistic visualization purposes. To accomplish this, pre-calibration is first performed to verify the intrinsic parameters of sensors and to determine the relative position between the flash LADAR and stereo vision system. This calibration process syncs the perspectives of the flash LADAR and stereo vision systems, allowing the range data from the flash LADAR to be projected into the image to determine the corresponding image pixel. An additional calibration process that fuses or enhances the data is conducted, which will be explained in the next section along with the description of the proposed realistic 3D modeling process.

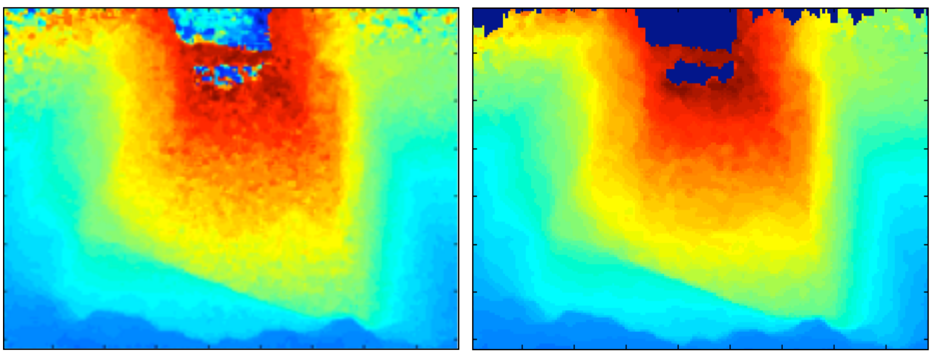
3D TERRAIN MODELING PROCESS

This section describes the process of proposed realistic 3D modeling system. The proposed modeling system consists of the following sequential steps: data filtering, data enhancement, and realistic 3D model generation. The proposed 3D model is validated by conducting field experiments on an actual construction site, which is a water pipe replacement process.

Data Filtering

In the modeling process, the input data are pre-calibrated range data and rectified color images from the flash LADAR and stereo vision systems. The range data obtained from flash LADAR contain considerable noise (Frome et al., 2004; Hedge and Ye, 2008). Figure 2(a) shows range data in which fluctuations in the level of gray indicate false range values. In an unstructured outdoor environment, range data acquired via flash LADAR suffer from 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 mainly caused by environmental conditions such as dust and rough surfaces of objects (Esselman and Verly, 1987; Wang et al., 2009). Such noise may have a negative effect on the quality of a 3D geometric model and data fusion results; therefore, data filtering must be done to eliminate the unnecessary noise, as the first step in the processing. To this end, a filtering method—one that uses an average-difference-value—was used.

The average difference value (ADV) employed in this research is the average of the differences between the value of a given pixel and those of its eight neighbors in the window centered at that pixel. If the ADV is larger than some predefined threshold value, the central pixel is assumed to be corrupted by dropout and is eliminated; otherwise, the central pixel is left unchanged. Throughout this process, points with an ADV above the threshold value were weeded out. After average-difference filtering is applied, there is still speckle noise, especially in the object region. In this study, a median filter was used to remove speckle noise and render the surfaces of objects more uniformly. A median filter is useful in eliminating speckle noise in a range image while preserving edge information (Doss, 2004). In a median filter, the range values of the pixels in the window centered on a given pixel are sorted in numerical order, and then the range value of the central pixel is replaced by the median value. Figure 2(b) shows the result of the data filtering process.



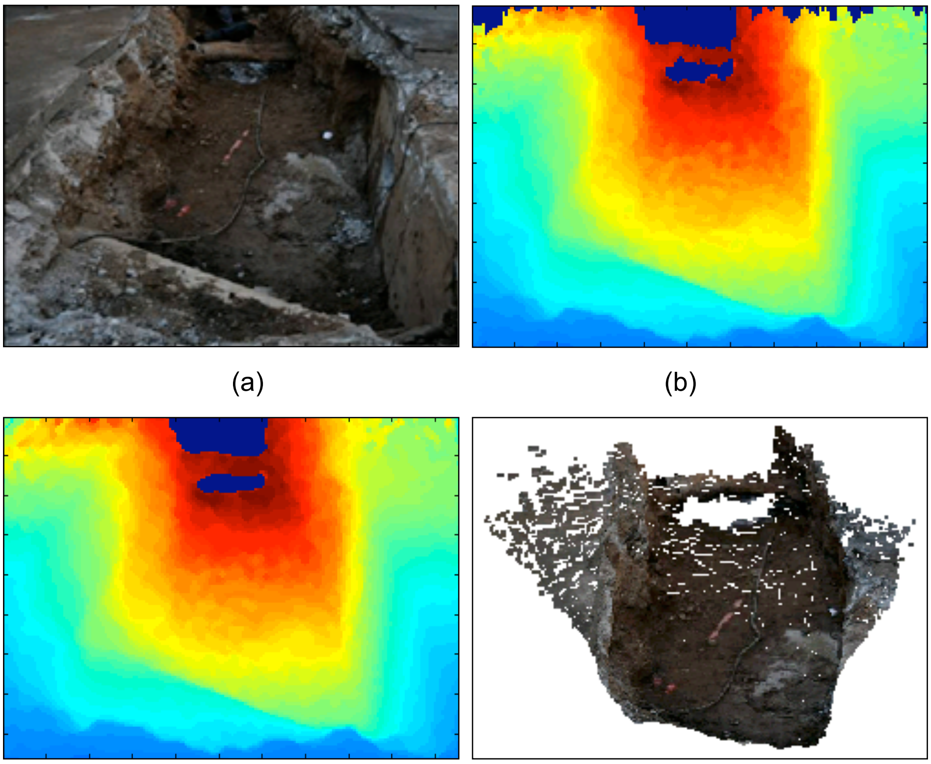
(a) (b)

Figure 2: Results of the data filtering process (a) Range data, (b) Result of the data filtering process

Data Fusion

After the data filtering process, although the noise in the original range data is reduced, the low resolution of 176*144 can corrupt the environment or object feature information such as edges or surfaces. In addition, the range data has no color or texture information, which is required for readily assimilating the real-world environment. Therefore, the data fusion process is proposed.

In this study, the algorithm adopts a bilateral filter to enhance the resolution and sub-pixel accuracy of the range data (Yang et al., 2007). By using two registered high-resolution color images as a reference, the low-resolution range data can be refined. First, the range data are up-sampled to the same size as the two color images. Second, cost volume is calculated according to the up-sampled range data. In this process, a stereo matching algorithm is performed together with the range data to calculate a more accurate cost volume. A bilateral filter is then used throughout each of the cost volume to calculate the new cost volume (Tomasi and Manduchi, 1998). Finally, the enhanced range data are obtained by taking the depth hypothesis. In Figure 3, results of the data fusion process are presented. Figure 3(a) and 3(b) shows the rectified color image and range data, respectively, after up-sampling from the low resolution of range data. Both the color image and range data have the resolution of 1,056*864. Figure 3(c) shows the enhanced range data, and Figure 3(d) shows the synthesized view of Figure 3(c). The quality has been improved significantly, and by visual comparison, the difference between Figure 3(b) and Figure 3(c) is clear.



(a) (b)

(c) (d)

Figure 3: Results of the data fusion process (a) Rectified color image, (b) Range data after data filtering, (c) Range data after data fusion, (d) Synthesized views using (c)

Realistic 3D Model Generation

In the data fusion process, in addition to resolution and accuracy enhancement, 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 (as shown in Figure 3(d)). Therefore, a realistic 3D 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, in order to generate a geometric 3D model of the entire scene from the usable range data. A mesh representation has the benefit of providing highly detailed representation of the environments or objects from a set of range data. A number of different types of mesh-generation methods are available in computer graphics. Among them, the 3D triangulation mesh is one of the most popular types employed to represent surfaces and generate models from 3D point clouds (Schroeder et al., 1996; Liu, 2007). One of the most commonly used triangulation algorithms is the Delaunay triangulation algorithm (Attali, 1997; Amenta and Bern, 1999), which generates a triangular mesh that can be used to represent complex environments. Thus it appears to be a good choice for representation of rough 3D environments such as those present during earthmoving operations on construction sites. Therefore, the Delaunay triangulation algorithm was employed to generate an initial surface mesh, and a geometric model of the entire scene from the resulting range data is performed. The surface texture is computed for every triangle in the geometric model, which are in the form of meshes. The initial mesh model is then painted for the purpose of achieving a realistic visualization model. The resulting realistic 3D model is presented in Figure 4.

Figure 4: Final realistic 3D model

CONCLUSIONS

As long as humans use displays of data generated for the control remote construction machines, a realistic three-dimensional (3D) modeling system is necessary to provide rich information feedback to the operator and facilitate remote operation. While single sensors such as a video camera are common, the multi-sensor usage for visualization and autonomy is already well established. This paper proposed and described a system used to visualize the construction environment in order to assist the operator of a remote controlled construction machine with a multi-sensor data fusion approach. The 3D modeling system created a realistic representation of the real-construction environments using the fusion of range data from flash LADAR and 2D imagery from a stereo vision system. The preliminary field experimental results show that combining data from multiple, complementary sensors increases the quality of the information available to the operator and makes human-machine interaction more efficient. In short, sensor fusion offers the potential to significantly improve remote operation.

The capability to visualize the geometric characteristics of the modeled system will greatly improve remote operation as an interactive visual feedback tool in earthmoving operations. Moreover, the realistic 3D modeling system proposed herein is essential technology for semi-automated and fully-automated construction machines. Therefore, it is believed to be extendable and applicable not only to earthmoving machines, such as excavators, but also to many other construction machines.

Future work will focus on optimization of the algorithms for effective data filtering, data fusion, and realistic 3D model generation during data conversion from both range data and color images to the realistic 3D model. In addition, experiments will be conducted on other construction sites and workspaces having different characteristics to validate the feasibility of the proposed method.

REFERENCES

Amenta, N., and Bern, M (1999) Surface reconstruction by voronoi filtering. Discrete and Computational Geometry, 22, 481–504.

Anderson, D., Herman, H., and Kelly, A (2005) Experimental characterization of commercial flash ladar devices. Proceedings of the 1st International Conference on Sensing Technologies, Palmerston North, New Zealand, 2005.

Attali, D (1997) R-regular shape reconstruction from unorganized points. Proceedings of the 13rd Annual Symposium on Computational Geometry, Nice, France.

Brunner, B., Arbter, K., and Hirzinger, G (1994) Task directed programming of sensor based robots. Proceedings of the IEEE International Conference on Intelligent Robots and Systems, Munich, Germany.

Csathó, B., Schenk, T., Lee, D., and Filin, S (1999) Inclusion of multispectral data into object recognition. International Archives of Photogrammetry and Remote Sensing, 32(7-4-3W6), 53–61.

Doss, N (2004) 3D modeling of biological structures. BS Thesis, Univ., Western Australia, Nedlands, WA 6907, Australia.

Esselman, T.R., and Verly, J.G (1987) Some applications of mathematical morphology to range imagery. Proceedings of the IEEE International Conference on International Conference on Acoustics, Speech, and Signal Processing 1987, Dallas, USA.

Frome, A., Huber, D., Kolluri, R., Bulow, T., and Malik, K (2004) Recognizing objects in range data using regional point descriptors. Proceedings of the European Conference on Computer Vision 2004, Prague, Czech Republic.

Habbit, R.D., Nellums, R.O., Niese, A.D., and Rodriguez, J.L (2003) Utilization of flash ladar for cooperative and uncooperative rendezvous and capture. Proceedings of the Society of Photographic Instrumentation Engineers, Orlando, USA.

Hedge, G.M., and Ye, C (2008) SwissRanger SR-3000 range images enhancement by a singular value decomposition filter. Proceedings of IEEE International Conference on Information and Automation, Zhangjiajie, China.

Hinzinger, G., Heindl, J., Landzettel, K., and Brunner, B (1992) Multisensory shared autonomy – a key issue in the space robot technology experiment ROTEX. Proceedings of the 1992 IEEE/RSJ International Conference on Intelligent Robots and Systems, Raleigh, USA.

Huber, D., Herman, H., Kelly, A., Rander, P., and Ziglar, J (2009) Real-time photo-realistic visualization of 3D environments for enhanced tele-operation of vehicles. Proceedings of the International Conference on 3-D Digital Imaging and Modeling, Kyoto, Japan.

Liu, L (2007) Scouting algorithms for field robots using triangular mesh maps. Dissertation, University of Saskatchewan.

Meier, R., Fong, T., Thorpe, C., and Baur, C (1999) A sensor fusion based user interface for vehicle teleoperation. Field and Service Robots Conference, Pittsburgh, USA.

Mesa Imaging AG, http://www.mesa-imaging.ch/, last accessed on March 10, 2010.

Oloufa, A. A., Ikeda, M., and Oda, H (2003) Situational awareness of construction equipment using GPS, wireless and web technologies. Automation in Construction, 12(6), 737-748.

Point Grey Research, Inc., http://www.ptgrey.com/, last accessed on March 10, 2010.

Schroeder, W., Martin, K., and Lorensen, B (1996) The visualization toolkit: an object-oriented approach to 3D graphics. New Jersey: Prentice Hall PTR.

Singh, S (1997) The state of the art in automation of earthmoving. Journal of Aerospace Engineering, 10(4), 179-188.

Tomasi, C., and Manduchi, R (1998) Bilateral filtering for gray and color images. Proceedings of the International Conference on Computer Vision, Bombay, India.

Wang, Q., Li, Q., Chen, Z., Sun, J., and Yao, R (2009) Range image noise suppression in laser imaging system. Optics and Laser Technology, 41(2), 140-147.

Yamada, H., and Doi, T (2008) Teleoperation of hydraulic construction robot using virtual reality. Proceedings of the 7th JFPS International Symposium on Fluid Power, Toyama, Japan.

Yang, Q.X., Yang, R.G., Davis, J., and Nister, D (2007) Spatial-depth super resolution for range images. Proceedings of the Computer Vision and Pattern Recognition 2007, Minneapolis, USA.