JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

A CPU-GPU Hybrid System of Environment Perception and 3D Terrain Reconstruction for Unmanned Ground Vehicle

Wei Song*, Shuanghui Zou*, Yifei Tian**, Su Sun*,
Simon Fong**, Kyungeun Cho***, and Lvyang Qiu***

Abstract

Environment perception and three-dimensional (3D) reconstruction tasks are used to provide unmanned ground vehicle (UGV) with driving awareness interfaces. The speed of obstacle segmentation and surrounding terrain reconstruction crucially influences decision making in UGVs. To increase the processing speed of environment information analysis, we develop a CPU-GPU hybrid system of automatic environment perception and 3D terrain reconstruction based on the integration of multiple sensors. The system consists of three functional modules, namely, multi-sensor data collection and pre-processing, environment perception, and 3D reconstruction. To integrate individual datasets collected from different sensors, the pre-processing function registers the sensed LiDAR (light detection and ranging) point clouds, video sequences, and motion information into a global terrain model after filtering redundant and noise data according to the redundancy removal principle. In the environment perception module, the registered discrete points are clustered into ground surface and individual objects by using a ground segmentation method and a connected component labeling algorithm. The estimated ground surface and non-ground objects indicate the terrain to be traversed and obstacles in the environment, thus creating driving awareness. The 3D reconstruction module calibrates the projection matrix between the mounted LiDAR and cameras to map the local point clouds onto the captured video images. Texture meshes and color particle models are used to reconstruct the ground surface and objects of the 3D terrain model, respectively. To accelerate the proposed system, we apply the GPU parallel computation method to implement the applied computer graphics and image processing algorithms in parallel.

1. Introduction

Unmanned driving, which integrates visual computing, localization, and artificial intelligence technologies, allows computers to automatically operate motor vehicles safely and reliably. Through the mounted sensors, an unmanned ground vehicle (UGV) can perceive its surrounding environment and obtain the vehicle location, terrain, and obstacle information. This way, the vehicle maintains safe

Corresponding Author: Wei Song (sw@ncut.edu.cn)
*   Dept. of Digital Media Technology, North China University of Technology, Beijing, China (sw@ncut.edu.cn, shuanghui_zou@sina.com, sunsu71@foxmail.com)
** Dept. of Computer and Information Science, University of Macau, Macau, China (tianyifei0000@sina.com, ccfong@umac.mo)
***Dept. of Multimedia Engineering, Dongguk University, Seoul, Korea (cke@dongguk.edu, lvyang_Qiu@foxmail.com)

distance from obstacles and avoids traffic accidents to ensure the safety of urban residents.

Environment perception and reconstruction studies apply multi-sensor fusion technology to provide driving awareness for the navigation of mobile robots in unknown environments. Traditionally, researchers have used video sequences or stereo images to sense color and depth information of the surrounding environment for terrain perception. Because video cameras are easily influenced by changes in illumination and are limited by their resolution, the perception accuracy is unstable, especially in faint environments. Laser scanners, which are not affected by the illumination conditions in an environment, achieve higher accuracy in measuring three-dimensional (3D) distance and perceiving object shape, but they do not provide color information. Thus, the integration of laser scanning and video cameras has been widely researched in the domain of 3D terrain perception and reconstruction [1].

Owing to the large size of sensed 3D point clouds, it is difficult to achieve real-time modeling of the surrounding environment by using traditional CPU-based terrain reconstruction techniques [2]. Considering the real-time performance, accuracy, and intuition of 3D terrain reconstruction, in this paper, the authors propose a CPU-GPU hybrid system of environment perception and 3D reconstruction for UGV based on the integration of multiple sensors. After the original datasets, including 3D point clouds, image sequences, and motion information are collected, intrinsic and extrinsic calibration processes are implemented to estimate the sensors' measurement parameter, position, and rotation states. The local coordinates of the point clouds are transformed into global coordinates in the terrain model based on localization performed by an inertial measurement unit (IMU). A connected component labeling algorithm is applied to cluster objects as a driving awareness interface for path planning in urban environments [3]. The terrain model is composed of a texture mesh and color particles, which represent ground surface and non-ground objects, respectively. In addition, GPU-based parallel programming technology is utilized to increase the computing speed of the environment perception and reconstruction tasks.

The remainder of this paper is organized as follows. Section 2 provides an overview of related work. Section 3 introduces the proposed CPU-GPU hybrid system of environment perception and 3D reconstruction for UGV. Section 4 evaluates the performance of the proposed methods. Section 5 concludes the paper.

2. Related Works

Currently, multiple sensors are widely applied in 3D terrain reconstruction to collect environment information for UGVs. Sukumar et al. [4] proposed a 3D terrain reconstruction system that combined 3D laser scanners, two-dimensional (2D) cameras, global positioning system (GPS), and IMU to perceive and reconstruct the surrounding terrain. By separating raw datasets as ground surface and non-ground objects, Huber et al. [5] created a colored ground mesh and colored point clouds to represent a photorealistic terrain model in a virtual environment.

To register large-scale point clouds into limited memory, Elseberg et al. [6] utilized an octree to store and compress massive data. Redundant points between successive point clouds were removed effectively to enable large-scale terrain reconstruction with limited storage capacity. However, when mobile robots continue to explore in outdoor environments, the terrain model space will continue to

expand and exceed the performance space defined by the octree data structure. Gingras et al. [7] applied a mesh simplification algorithm to reduce the number of grids in large-scale ground models. Through reconstruction of unstructured terrain models by using an irregular triangular mesh, large-scale point clouds were registered into the terrain model effectively. However, these methods processed terrain reconstruction point-by-point, which made it very difficult to achieve real-time terrain reconstruction owing to low computational speed.

Object segmentation and tracking is the primary task of intelligent cognitive process, and this task provides a reliable basis for intelligent obstacle avoidance and path planning. Golovinskiy and Funkhouser [8] applied a k-nearest neighbor's graph algorithm to separate the foreground and background. By using a min-cut algorithm, the objects were segmented from the point clouds. However, it was difficult to update the k-nearest neighbors graph in real time when large-scale point clouds were sensed and registered into the terrain model. For actualizing real-time object segmentation, Douillard et al. [9] presented a segmentation strategy based on the voxel grids of 3D points. In this strategy, a terrain mesh was first constructed using 3D point clouds, in which ground points were extracted by computing the gradient field in the mesh model. Thereafter, a cluster-all method is applied to separate and classify the non-ground points. Himmelsbach et al. [10] proposed a bottom-up method to segment objects. In this method, fitted line segments searched from point clouds were utilized to determine the ground points. Subsequently, non-ground points were segmented into individual clusters based on the spatial connected relationship. Wang et al. [11] voxelized non-ground points and clustered interconnected voxels as target objects. They utilized the principal component analysis algorithm to determine the eigenvectors of voxel clusters instead of discrete non-ground points. The eigenvectors and eigenvalues described the shape features, thus reducing the memory consumption for object description. To achieve real-time object segmentation and tracking, Wang et al. [12] first projected LiDAR (light detection and ranging) point clouds collected by mobile unmanned vehicles onto a rasterized horizontal plane to conduct segmentation. Then, a support vector machine (SVM) was exploited to classify objects, and a Kalman filter algorithm was implemented to track different objects. Broggi et al. [13] used a stereo camera to capture point clouds surrounding an unmanned car, and these point clouds were grouped into several different clusters by using a flood fill method. A linear Kalman filter was employed to analyze the movement and posture of obstacles, which were classified as moving or stationary. The iterative and traversal computation processes required to search for neighboring points and analyze the large-scale dataset reduced the computational speeds of these methods [14].

In view of the insufficiency existing in 3D terrain perception and reconstruction, in the present work, we describe a CPU-GPU hybrid system of environment perception and 3D reconstruction to achieve the real-time and intuitive requirements of UGVs.

3. The Environment Perception and 3D Reconstruction System

Herein, we propose an environment perception and 3D reconstruction system based on multi-sensor integration. The multi-sensor datasets, including 3D point clouds, 2D images, motion information of mobile vehicles, are processed and integrated to realize environment perception and 3D terrain reconstruction. As shown in Fig. 1, the system contains three main modules: multi-sensor data collection and pre-processing, environment perception, and 3D reconstruction. The multi-sensor data

collection and pre-processing module implements the collection, processing, and fusion of multi-sensor terrain data. The environment perception module uses the LiDAR and gyroscope to filter the traversable ground surface and detect non-ground obstacles in the surrounding environment. The 3D reconstruction module performs real-time and high-definition reconstruction of ground and non-ground objects by using a texture mesh model and a color particle model to represent a real environment intuitively in a virtual environment.
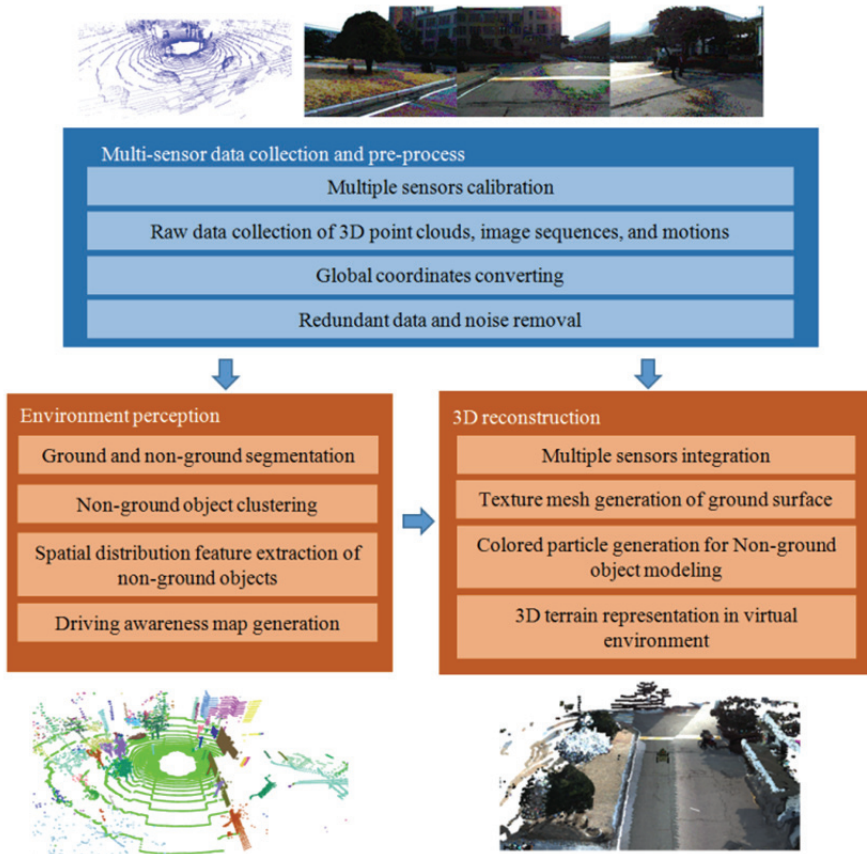

Fig. 1. The proposed system framework of environment perception and reconstruction.

3.1 Multi-Sensor Data Collection and Pre-process Module

In the multi-sensor data collection and pre-processing module, the system obtains global environment data of the unmanned vehicle. Owing to inconsistencies in the spatial position and rotation angle between the LiDAR and the cameras, direct projection of the point clouds onto the video images leads to distorted mapping. By using a calibration board, calibration between the cameras and the LiDAR adjusts the projection matrix to provide a 3D terrain model with accurate color information.

After calibrating multiple sensors, the system utilizes these sensors to collect raw datasets for real-time environment perception and large-scale 3D terrain reconstruction. The local coordinates of the point clouds are converted into global coordinates based on the position and motion datasets of the

UGV, which are sensed by the IMU. By using the translation and rotation matrices, the 3D point clouds in different coordinate systems are transformed into a global coordinate system. The repeated areas detected by LiDAR sensors, and cameras introduce redundant data leading to massive wastage of storage resources. Thus, redundant and noise data are removed following the deduplication principle. In a repeatedly covered voxel, only one point is registered in the terrain model, so that 3D reconstruction is realized with low memory consumption.

3.2 Multi-Sensor Data Collection and Pre-process Module

The environment perception module utilizes the data collected by the mounted LiDAR and IMU to realize environment perception, which involves the following functions: ground and non-ground segmentation, non-ground object clustering, spatial distribution feature extraction of non-ground objects, and driving awareness map generation. Fig. 2 presents a sequential diagram of the proposed environment perception module, in which the data flow and functions of the CPU and GPU procedures are shown. The corresponding variables are explained in Table 1.
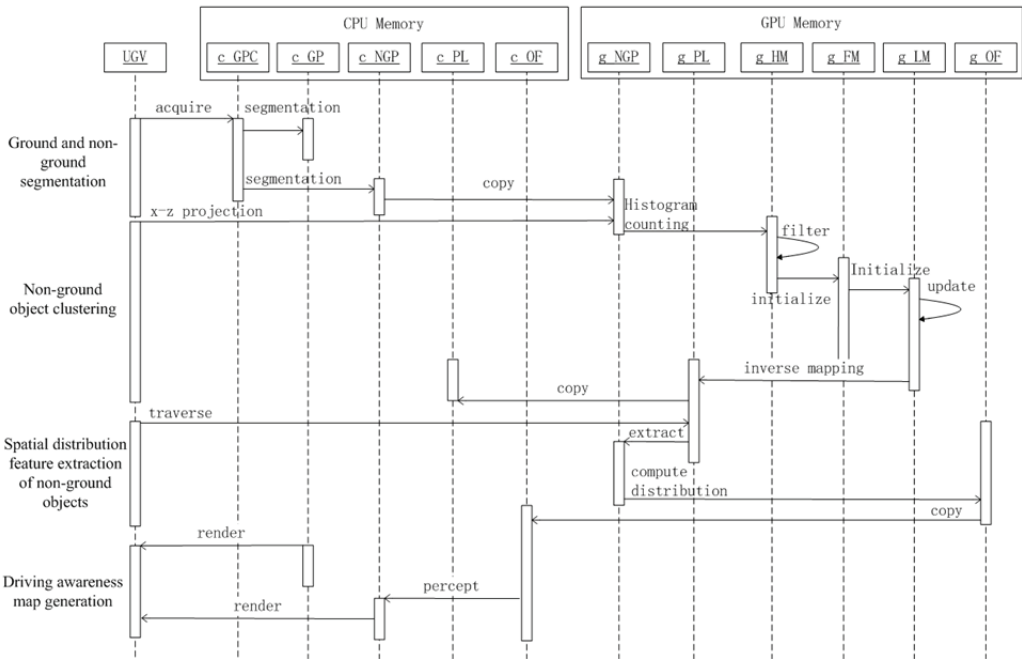

Fig. 2. Proposed system framework for environment perception and reconstruction.

Firstly, the converted global point clouds stored in CPU memory are acquired. By using a height threshold, the non-ground points are filtered out and copied to GPU memory for implementing object clustering in parallel. The non-ground points stored in GPU memory are projected to the x-z plane to generate a 2D histogram map with a rasterizing process. The histogram map is utilized to count the non-ground sensing points above the covered area of each grid cell in the map [15]. A flag map is initialized to mark the effectiveness of the point counts recorded in each cell of the histogram map. The label map is initialized and updated to indicate connected components of grid cells in the histogram

map through several iterations in GPU memory. Thus, from the clustering result in the label map, point labels are obtained by inverse mapping to identify individual objects in the environment. The point labels are copied from the GPU to the CPU memory as reference information about non-ground objects, which are rendered in the 3D reconstruction module. In the spatial distribution feature extraction function of non-ground objects, the point labels and non-ground points in GPU memory are traversed to extract and compute the distribution features of non-ground objects. The object features stored in GPU memory are copied into CPU memory. In the driving awareness map generation function, ground points and non-ground points are rendered based on the object features as the result of environment perception [16].

Table 1. Explanation of variables in Fig. 2

Abbreviation	Description
c_GPC	Global point cloud in CPU memory
c_GP	Ground points in CPU memory
c_NGP	Non-ground points in CPU memory
c_PL	Point labels in CPU memory
c_OF	Object feature in CPU memory
g_NGP	Non-ground point in GPU memory
g_PL	Point labels in GPU memory
g_HM	Histogram map in GPU memory
g_FM	Flag map in GPU memory
g_LM	Label map in GPU memory
G_OF	Object Feature In GPU Memory

3.3 3D Reconstruction Module

The 3D reconstruction module performs four main functions: multiple sensor integration, texture mesh generation of ground surface, colored particle generation for non-ground object modeling, and 3D terrain representation in virtual environment. GPU-based parallel programming technology is applied to accelerate the 3D reconstruction process. The CPU and GPU functions in the 3D reconstruction module are described in the sequential diagram shown in Fig. 3. The corresponding variables are explained in Table 2.

Under the multiple sensors integration function, global point clouds are acquired from CPU memory and segmented into non-ground and ground points, which are stored in GPU memory. The ground points are registered into a ground mesh so that the large-scale discrete ground points are removed and represented by a texture mesh. A calibration method is used to adjust the projection matrix between the global point cloud and the image sequences. In the texture mesh generation process, image sequences are copied from the CPU to the GPU memory, in which the ground mesh is mapped onto the image sequences determined using the calibrated projection matrix. The pixel image sequence mapped from the triangles in the ground mesh is registered to the texture datasets of the ground mesh. The generated ground texture mesh is copied from the GPU to the CPU memory to realize real-time rendering in the virtual environment. The non-ground objects are represented by colored particles, which are generated by mapping the non-ground points onto the image sequence to acquire color information. In the 3D

terrain representation process, the texture mesh and the color particles stored in CPU memory are rendered as ground surface and objects, respectively.
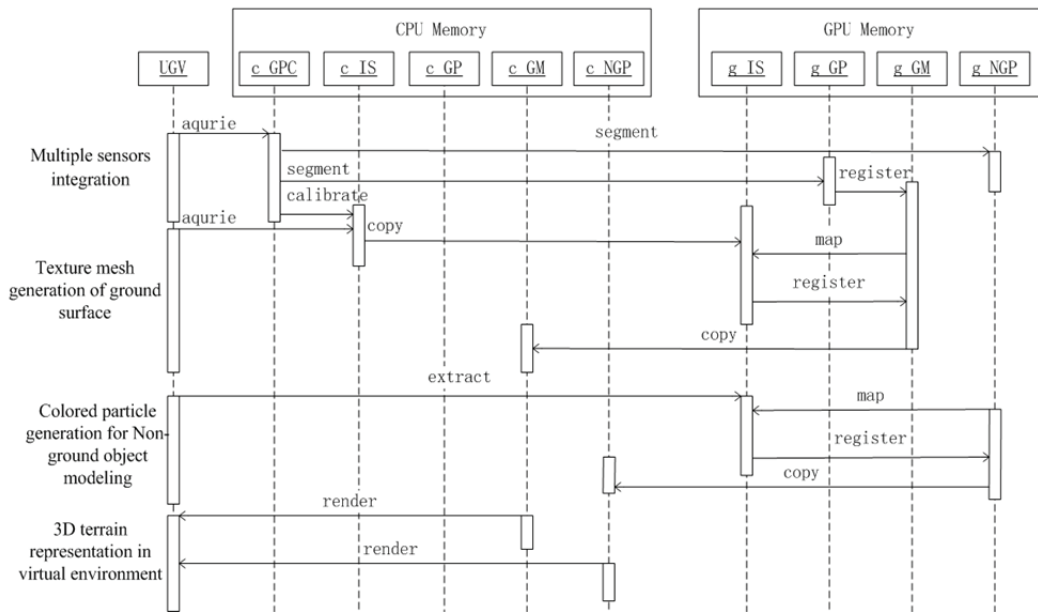

Fig. 3. CPU-GPU sequential diagram of proposed 3D reconstruction module.

Table 2. Explanation of variables in Fig. 3

Abbreviation	Description
c_GPC	Global point cloud in CPU memory
c_IS	Image sequence in CPU memory
c_GP	Ground points in CPU memory
c_GM	Ground mesh in CPU memory
c_NGP	Non-ground points in CPU memory
g_IS	Image sequence in GPU memory
g_GP	Ground points in GPU memory
g_GM	Ground mesh in GPU memory
g_NGP	Non-ground point in GPU memory

4. Experiments

As shown in Fig. 4, the UGV was equipped with a LiDAR to collect point cloud data, three GC655 VGA CCD cameras to photograph the surroundings, and an MTi-G-700 IMU to report the position and rotation information of the vehicle. The CUDA programming method for parallel computing was used in our experiment. The system was executed on a 3.20 GHz Intel Core Quad CPU computer with a GeForce GT 770 graphics card and 4 GB RAM. The system utilized Direct3D software development kits to visualize the object segmentation and 3D reconstruction results.
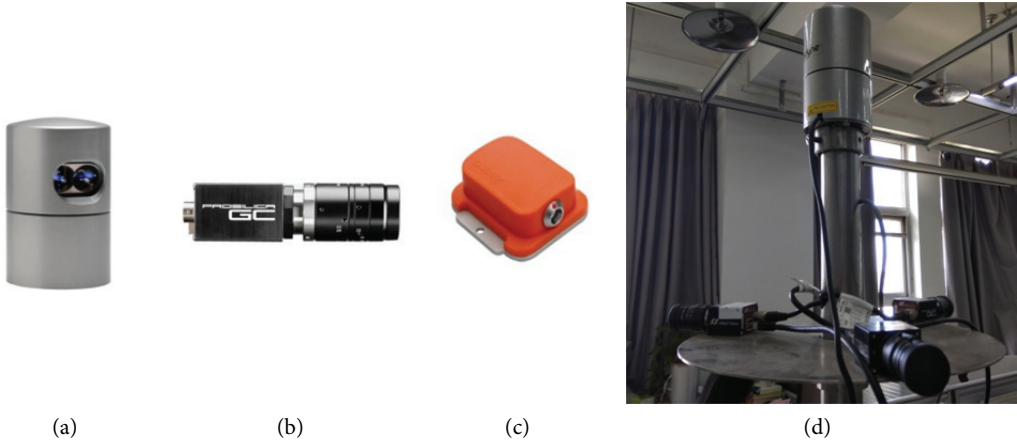
| (a) | (b) | (c) | (d) |

Fig. 4. Multiple sensors mounted on UGV. (a) LiDAR, (b) CCD camera, (c) IMU, and (d) multiple sensor integration.

Fig. 5 shows the surrounding point clouds perceived by the LiDAR mounted on the UGV. By using the proposed ground segmentation technique, the ground points were segmented, as shown in green, and these points were recognized as traversable regions. As shown in Fig. 6, by using the connected component labeling algorithm, non-ground points were divided into several distinguished objects shown in different colors, which are considered obstacles. This way, the UGV could plan its path effectively.

The system adopted the CUDA GPU programming technique to realize parallel computation in the three proposed modules. The object segmentation speed performances in the proposed CPU-GPU hybrid system and the CPU-based method were demonstrated in Fig. 7. In each frame, the reconstructed terrain model contained nearly 100,000 3D points, which were hard to be processed by the CPU sequential computation in real time. Our proposed method was able to achieve more than 31.3 frames per second (fps) on average, which satisfied the real-time processing requirement of UGV.
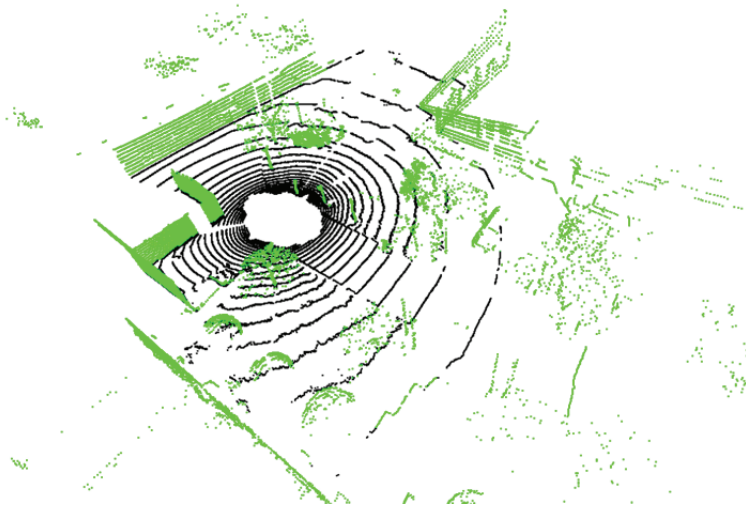

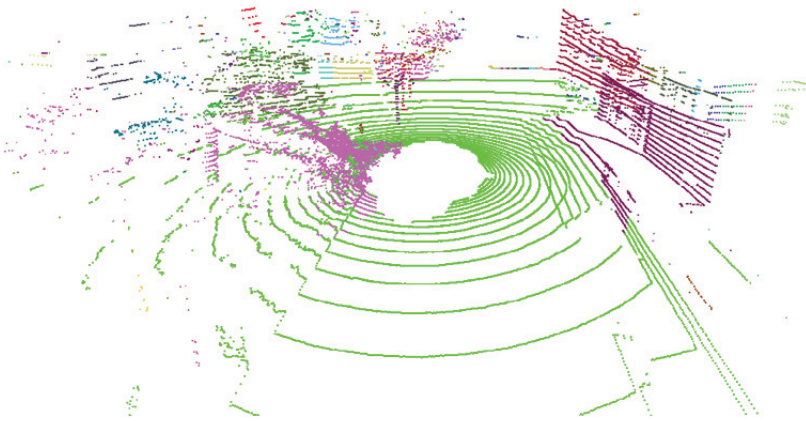**Fig. 5.** Segmentation result of ground and non-ground points.

Fig. 6. Ground segmentation and object clustering result in LiDAR point clouds.

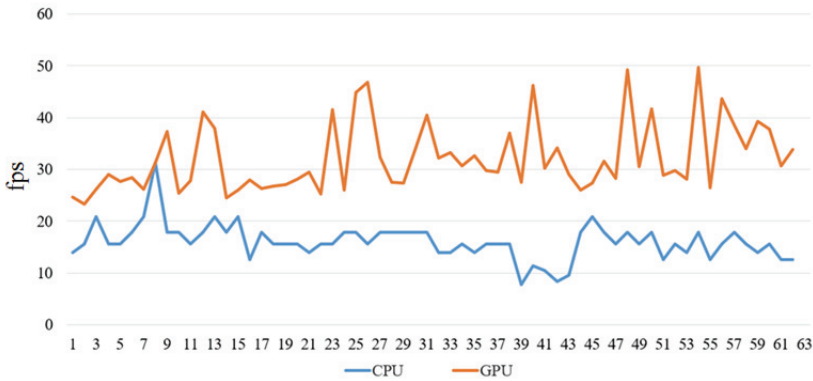**Fig. 7.** Object segmentation speed performances using the proposed CPU-GPU hybrid system and the CPU-based method.

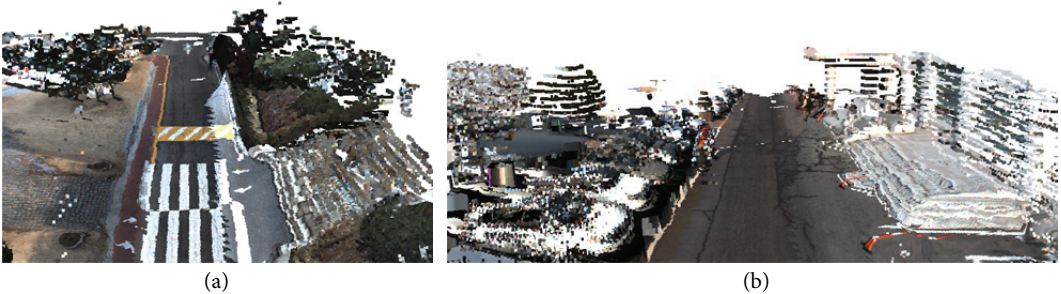<div align="center">(a)   (b)</div>

Fig. 8. High-resolution terrain reconstruction results using texture mesh (a) and colored particles (b).

In the system, the ground surface and non-ground objects were reconstructed using a high-resolution texture mesh and colored particles, respectively, as shown in Fig. 8. The speed of environment perception and high-resolution reconstruction was more than 23.0 fps. Thus, the proposed system can effectively solve the problem of limited computing speed and satisfy the requirements for real-time processing of large-scale datasets.

5. Conclusion

In this paper, we introduced a CPU-GPU hybrid system of environment perception and 3D reconstruction for UGVs, and the data flows and functions of this system were designed based on sequence diagrams. The system mainly contained three modules, namely, data collection and pre-processing, environment perception, and 3D terrain reconstruction. After removing redundant and noise data, the pre-processing function registered LiDAR point clouds and video sequences into a global terrain model. The environment perception module segmented the ground surface from the point clouds and clustered the non-ground points into individual objects by using a connected component labeling algorithm. In the 3D terrain reconstruction module, the projection matrices of the LiDAR and the three cameras were calibrated to project the local point clouds onto the video image sequences accurately. The terrain model was reconstructed with the texture mesh and the colored particles to represent the surrounding environment intuitively. Our proposed method is compatible with applications such as intelligent surveying and mapping, robot vision, and 3D modeling.

Acknowledgement

References

[1] Y. Matsushita and J. Miura, "On-line road boundary modeling with multiple sensory features, flexible road model, and particle filter," *Robotics and Autonomous Systems*, vol. 59, no. 5, pp. 274-284, 2011.

[2] J. M. Noguera, R. J. Segura, C. J. Ogayar, and R. Joan-Arinyo, "Navigating large terrains using commodity mobile devices," *Computers & Geosciences*, vol. 37, no. 9, pp. 1218-1233, 2011.

[3] W. Song, S. Zou, Y. Tian, S. Fong, and K. Cho, "Classifying 3D objects in LiDAR point clouds with a back-propagation neural network," *Human-centric Computing and Information Sciences*, vol. 8, article no. 29, 2018.

[4] S. R. Sukumar, S. Yu, D. L. Page, A. F. Koschan, and M. A. Abidi, "Multi-sensor integration for unmanned terrain modeling," in *Proceedings of SPIE 6230: Unmanned Systems Technology VIII*. Bellingham, WA: International Society for Optics and Photonics, 2006.

[5] D. Huber, H. Herman, A. Kelly, P. Rander, and J. Ziglar, "Real-time photo-realistic visualization of 3D environments for enhanced tele-operation of vehicles," in *Proceedings of 2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)*, Kyoto, Japan, 2009, pp. 1518-1525.

[6] J. Elseberg, D. Borrmann, and A. Nuchter, "One billion points in the cloud–an octree for efficient processing of 3D laser scans," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 76, pp. 76-88, 2013.

[7]   D. Gingras, T. Lamarche, J. L. Bedwani, and E. Dupuis, "Rough terrain reconstruction for rover motion planning," in *Proceedings of 2010 Canadian Conference on Computer and Robot Vision (CRV)*, Ottawa, Canada, 2010, pp. 191-198.

[8]   A. Golovinskiy and T. Funkhouser, "Min-cut based segmentation of point clouds," in *Proceedings of 2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops),* Kyoto, Japan, 2009, pp. 39-46.

[9]   B. Douillard, J. Underwood, V. Vlaskine, A. Quadros, and S. Singh, "A pipeline for the segmentation and classification of 3D point clouds," in *Experimental Robotics*. Heidelberg: Springer, 2014, pp. 585-600.

[10] M. Himmelsbach, F. V. Hundelshausen, and H. J. Wuensche, "Fast segmentation of 3D point clouds for ground vehicles," in *Proceedings of 2010 IEEE Intelligent Vehicles Symposium*, San Diego, CA, 2010, pp. 560-565.

[11] J. Wang, R. Lindenbergh, and M. Menenti, "SigVox: a 3D feature matching algorithm for automatic street object recognition in mobile laser scanning point clouds," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 128, pp. 111-129, 2017.

[12] H. Wang, B. Wang, B. Liu, X. Meng, and G. Yang, "Pedestrian recognition and tracking using 3D LiDAR for autonomous vehicle," *Robotics and Autonomous Systems*, vol. 88, pp. 71-78, 2017.

[13] A. Broggi, S. Cattani, M. Patander, M. Sabbatelli, and P. Zani, "A full-3D voxel-based dynamic obstacle detection for urban scenario using stereo vision," in *Proceedings of 2013 16th International IEEE Conference on Intelligent Transportation Systems-(ITSC)*, Hague, The Netherlands, 2013, pp. 71-76.

[14] A. Khatamian and H. R. Arabnia, "Survey on 3D surface reconstruction," *Journal of Information Processing Systems*, vol. 12, no. 3, pp. 338-357, 2016.

[15] W. Song, L. Liu, Y. Tian, G. Sun, S. Fong, and K. Cho, "A 3D localisation method in indoor environments for virtual reality applications," *Human-centric Computing and Information Sciences*, vol. 7, article no. 39, 2017.

[16] D. Zeng, Y. Dai, F. Li, R. S. Sherratt, and J. Wang, "Adversarial learning for distant supervised relation extraction," *Computers, Materials & Continua*, vol. 55, no. 1, pp. 121-136, 2018.

Wei Song https://orcid.org/0000-0002-5909-9661

He received his B.Eng. degree in Software Engineering from Northeastern University, Shenyang, China, in 2005, and his M.Eng. and Dr.Eng. degrees in the Department of Multimedia from Dongguk University, Seoul, Korea, in 2008 and 2013, respectively. Since September 2013, he has been an Associate Professor at the Department of Digital Media Technology of North China University of Technology. His current research interests are focused on parallel computation, 3D modeling, object recognition, virtual reality, and Internet of Things.

Shuanghui Zou https://orcid.org/0000-0002-2185-4078

She is a postgraduate student at the School of Computer Science and Technology of North China University of Technology, Beijing, China, since September 2017. Her current research interests are focused on machine learning, IoT, image processing, and 3D reconstruction.

Yifei Tian https://orcid.org/0000-0001-8884-2170

She is a postgraduate student at the Computer and Information Science Department of the University of Macau, Macau, China, since September 2018. Her current research interests are focused on IoT, big data mining, computer graphics, and 3D reconstruction.

Su Sun https://orcid.org/0000-0002-6246-2046

She is an undergraduate student at the Department of Digital Media Technology of North China University of Technology, Beijing, China, since September 2015. Her current research interests are focused on unmanned cars, virtual reality and computer graphics.

Simon Fong https://orcid.org/0000-0002-1848-7246

He graduated from La Trobe University, Australia, with a 1st Class Honours B.Eng. Computer Systems degree and a Ph.D. Computer Science degree in 1993 and 1998, respectively. Simon is now working as an Associate Professor at the Computer and Information Science Department of the University of Macau. He is also one of the founding members of the Data Analytics and Collaborative Computing Research Group in the Faculty of Science and Technology. Prior to his academic career, Simon took up various managerial and technical posts, such as systems engineer, IT consultant and e-commerce director in Australia and Asia.

Kyungeun Cho https://orcid.org/0000-0003-2219-0848

She is a full professor at the Department of Multimedia Engineering at the Dongguk University in Seoul, Korea since September 2003. She received her B.Eng. degree in Computer Science in 1993, her M.Eng. and Dr.Eng. degrees in Computer Engineering in 1995 and 2001, respectively, all from Dongguk University, Seoul, Korea. During 1997–1998, she was a research assistant at the Institute for Social Medicine at the Regensburg University, Germany, and a visiting researcher at the FORWISS Institute at TU-Muenchen University, Germany. Her current research interests are focused on the areas of intelligence of robot and virtual characters and real-time computer graphics technologies. She has led a number of projects on robotics and game engines and also has published many technical papers in these areas.

Lvyang Qiu https://orcid.org/0000-0002-6275-206X

He is a graduate student at Dongguk, Seoul, Korea, since September 2018. His current research interests are focused on 3D reconstruction, path planning, and image processing.