Elemental Technologies for Supporting the Integration of Real Space and Cyberspace

Jun Shimamura, Shingo Ando, and Ryuichi Tanida

Abstract

To create a human-centered society in which everyone can lead a comfortable, vibrant, and high-quality life, we hope that a system that integrates real space and cyberspace will become a reality. In this article, a mobile mapping system (MMS) point cloud data processing technology called geoNebula™ is introduced, which analyzes data measured in real space, recognizes space and objects, and compresses the data into a compact form suitable for constructing cyberspace. This technology is being researched and developed to establish a mechanism for creating value through the sophisticated blending of real space and cyberspace.

Keywords: cyberspace, real space, MMS point cloud data processing technology

1. Building cyberspace

With the development of information and communication technologies, it is becoming possible to collect and analyze vast amounts of data from the Internet of Things. In response, the Japanese government and various companies are pursuing research and development to create a system that integrates real space and cyberspace in the manner advocated in Society 5.0. NTT has started research and development of the 4D digital platform™, which can integrate four-dimensional (4D) information (latitude, longitude, altitude, and time) as precisely as possible in real time and make future predictions from those data to implement the above system in society.

The 4D digital platform integrates sensing data about people and objects on the Advanced Geospatial Information Database with high accuracy and containing rich semantic information in a manner that enables advanced prediction, analysis, simulation, and optimization in various industrial fields (**Fig. 1**). We believe that this platform has many possible uses in areas such as alleviation of road traffic, optimal use of urban assets, and maintenance and management of social infrastructure. To develop the Advanced Geo-

spatial Information Database, it is necessary to convert real space into data by using technology such as laser-imaging detection and ranging (LiDAR) and construct cyberspace with those data. This process is generally composed of the following three steps.

Step 1 (data collection): Real space measured using LiDAR, etc. is converted into data.
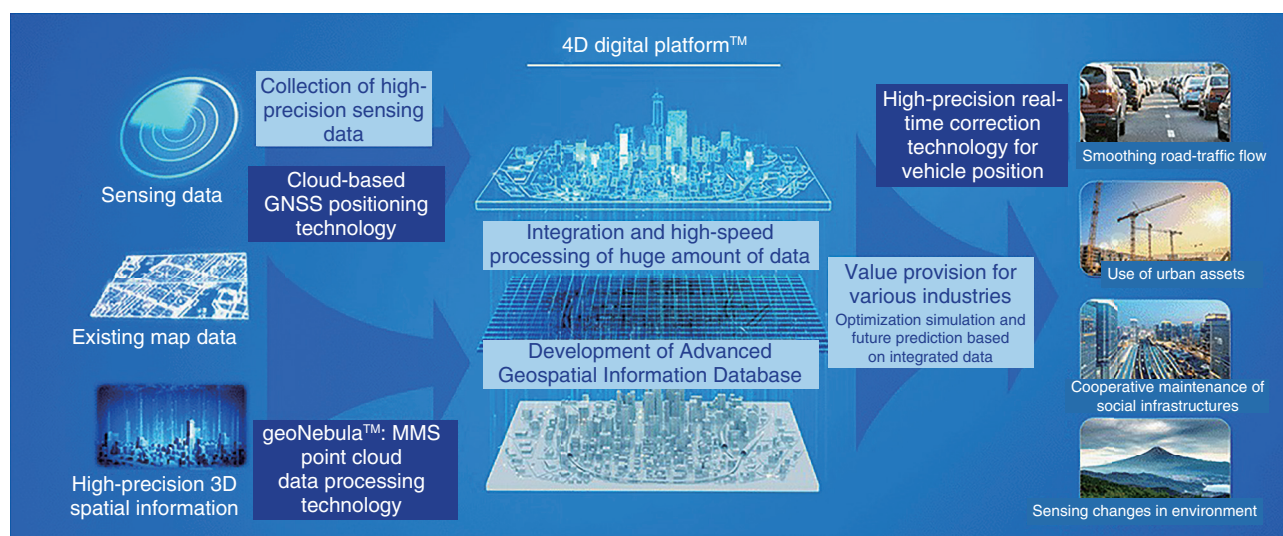
Step 2 (space and object recognition): Recognize space and objects from the measurement data and obtain their exact positions.

Step 3 (data storage): Store the finished data and original data for updating or other uses.

The latest research results concerning geoNebula™—a mobile mapping system (MMS) point cloud data processing technology we are developing to save labor and improve the efficiency of each step—are introduced hereafter.

2. Real-space structuring technology that enables labor saving and efficiency improvement of data collection and space/object recognition

With our research and development, we aim to establish the following two technologies for structuring real space from 3D point clouds measured using

GNSS: global navigation satellite system

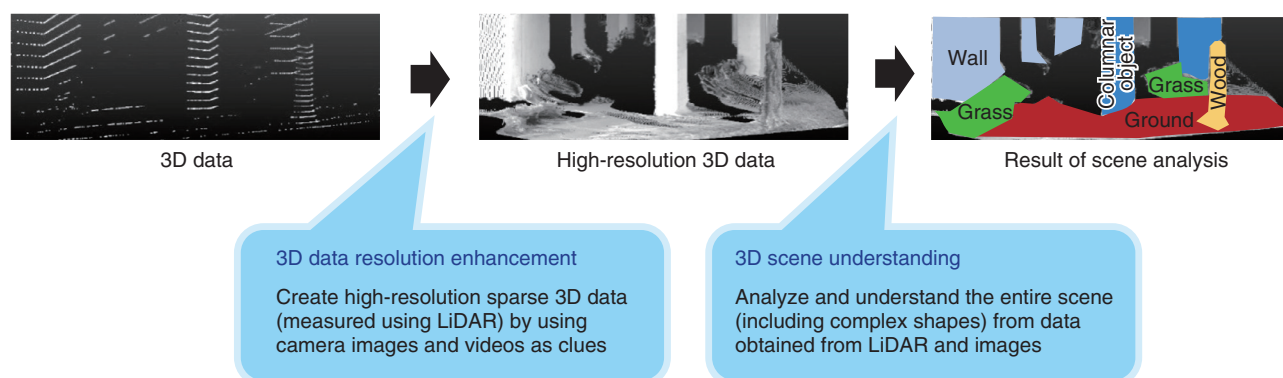Fig. 1. Overview of 4D digital platform™.

Fig. 2. Real-space structuring technology.

LiDAR, etc. (**Fig. 2**).

(1) 3D data resolution enhancement
(2) 3D scene understanding

3D data resolution enhancement is positioned as step 1 (data collection) in the three steps for constructing cyberspace listed at the end of the previous section. Due to the structure of a 3D point cloud, vacant spaces (gaps) of a certain width are created between points that are spatially adjacent to each other. This technology (3D data resolution enhancement) can acquire higher-definition 3D point clouds by generating multiple new points to fill these gaps and interpolating them [1]. 3D scene understanding is

positioned as step 2 (space and object recognition). It is capable of automatically assigning semantic labels (road surface, building, etc.) to all points in a dense point cloud on a large scale and with high accuracy on a point-by-point basis [2]. These two technologies are explained in detail as follows.

2.1 3D data resolution enhancement

Less expensive LiDAR equipment tends to produce larger gaps between points. In such a case, much of the 3D-shape information of an object is lost, and the object's detailed features are obscured. The resolution of camera images has significantly increased, so

it is now possible to acquire relatively dense 2D data even with inexpensive cameras. Therefore, if the information from the camera images taken separately is used appropriately, it would be possible to compensate for a sparse 3D point cloud. Although several technologies for this purpose are on the market, it is necessary to strictly align (calibrate) the LiDAR point cloud with the camera image as a prerequisite. Aligning the two sensors (LiDAR and camera) is generally difficult and requires a high level of skill. We therefore proposed a method of generating a stable and dense point cloud even with a relatively simple rough calibration. What sets our method apart from other methods is not only installation cost (low cost of equipment) but also operation cost (simple calibration) are accounted for.

2.2　3D scene understanding

Several deep neural networks (DNNs) that can be applied to non-grid data such as 3D point clouds have been proposed. 3D scene understanding is one such DNN for non-grid data, and we proposed using it to solve a task called semantic segmentation [2]. Semantic segmentation of a 3D point cloud is the process of separating all points in the point cloud into multiple segments by assigning meaningful labels to each point. By pre-training with a DNN, it is possible to estimate the label that should be assigned to each point in an unknown 3D point cloud. However, most conventional methods target point cloud data within a limited space such as indoors. Accordingly, to handle a wide area such as an entire city with a high density of point clouds, a high-spec computer is required. With that issue in mind, we devised a method for efficient semantic segmentation of large point cloud data even on a relatively inexpensive computer [2]. Combining this method with the aforementioned 3D data resolution enhancement technology makes it possible to acquire high-density point clouds inexpensively and understand an entire 3D scene in detail.

3.　4D point cloud coding technology that enables highly compressed storage of 3D point cloud data

4D point cloud coding is a technology that highly compresses and stores 3D point cloud data obtained from LiDAR and stereo cameras [3]. This technology targets 3D point cloud data concerning, for example, roads and cities collected using the Advanced Geospatial Information Database, and we are researching and developing technology for highly efficient compression of such data and their convenient utilization.

The first problem regarding accumulating and using 3D point cloud data of a wide area of a city is the size of the data. Efficiently storing a large amount of point cloud data necessitates a highly efficient compression technique. Moreover, 3D point cloud data collected moment by moment by using in-vehicle LiDAR, etc. are generated in fragments along the trajectory of a moving vehicle. Therefore, to accumulate and manage those data as a 3D point cloud database covering a wide area of a city, it is necessary to integrate the fragmented 3D point cloud data and detect temporal changes in the latest data from past data. To use such integrated "urban" data, it is necessary to implement a random-access function to retrieve 3D point cloud data at arbitrary locations. In addition to developing the compression function, we are researching and developing three main functions: data integration, temporal-change detection, and random access.

The compression function is based on the point cloud coding technology called geometry-based point cloud compression (G-PCC), which is being standardized by MPEG (Motion Picture Experts Group) and enables highly efficient compression of 3D point cloud data with a large data size. By efficiently representing the spatial location of points as an octree structure while dividing the space hierarchically, G-PCC can compress the data size up to about a tenth of the original size (**Fig. 3**).

Integration of 3D point cloud data and detection of temporal changes are explained as follows. In addition to developing the function for compressing the data of individual files, we are investigating functions for integrating the data in the compressed files and easily comparing the differences between the data. As mentioned above, 3D point cloud data compressed by G-PCC maintain the spatial location of points hierarchically in an octree structure. This hierarchical structure makes it easy to extract a rough outline of the objects in the data from the encoded data. By extracting this outline, it is possible to, for example, detect changes in 3D point cloud data at different acquisition times and combine fragmented data while taking object boundaries into account (**Figs. 4** and **5**).

Regarding the random-access function, by using the hierarchical structure of G-PCC, we are also investigating (i) a scalable function for extracting sparse summary data from dense point cloud data and (ii) a function for extracting only a part of the space. With these functions, it becomes easy to not only extract partial data but also obtain a general idea of
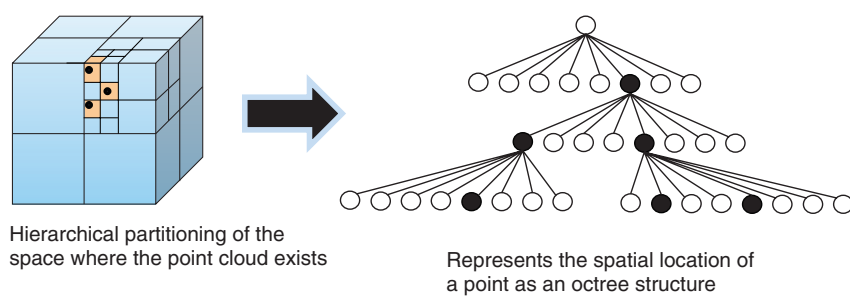
Hierarchical partitioning of the space where the point cloud exists

Represents the spatial location of a point as an octree structure

Fig. 3.   Layering of point cloud data using an octree structure.

Fragment data of 3D point cloud (1)

Fragment data of 3D point cloud (2)

High-speed integration by detecting the boundary between both types of data while gradually refining the accuracy by using a hierarchical structure
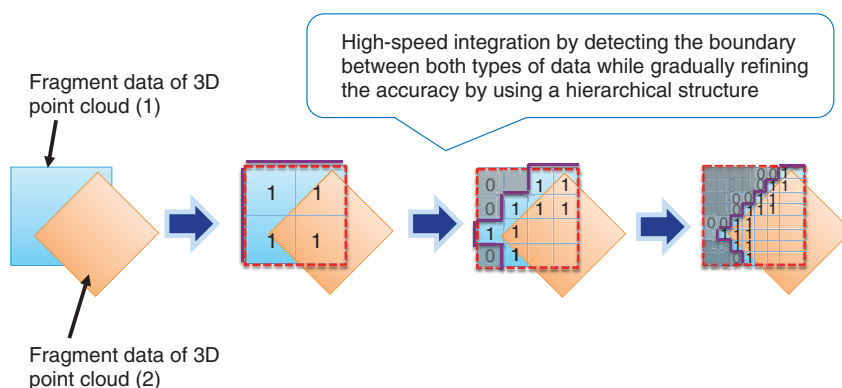
Fig. 4.   Combined processing of 3D point cloud data.

the data over a wider area.

We are currently researching and developing a software engine that can handle these three functions in a unified manner. Using this engine will make it possible to store 3D point cloud data of a vast city in a compact size and easily retrieve the data for any arbitrary position and time. We plan to continue our research and development to enable faster and more efficient encoding while taking into account the fact that the volume and distribution of 3D point cloud data will further increase in the future.

4.   Future developments

Using the technologies described above, we will establish technologies for constructing cyberspace that can precisely reproduce real space and embody a mechanism for value creation through the advanced integration of real space and cyberspace. Therefore, we will contribute to the creation of a human-centered society in which everyone can lead a comfortable, energetic, and high-quality life.
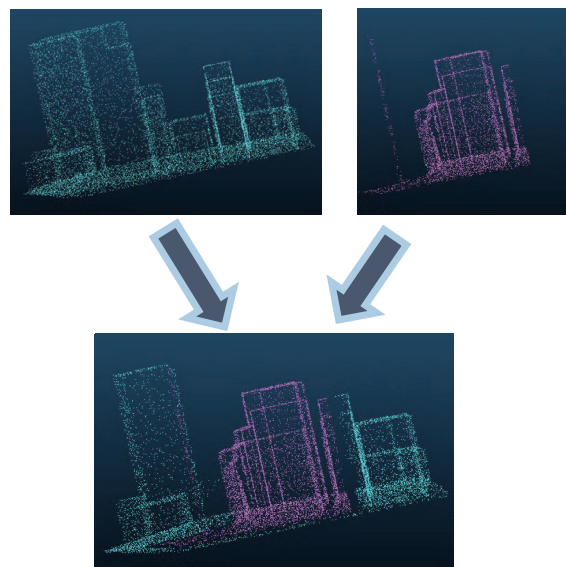

Fig. 5.   Result of integrating 3D point cloud data.

References

[1] Y. Yao, R. Ishikawa, S. Ando, K. Kurata, N. Ito, J. Shimamura, and T. Oishi, "Non-learning Stereo-aided Depth Completion Under Misprojection via Selective Stereo Matching," IEEE Access, Vol. 9, pp. 136674–136686, 2021.

[2] K. Kurata, Y. Yao, S. Ando, N. Ito, and J. Shimamura, "Aggregative Input Convolution for Large-scale Point Cloud Semantic Segmentation," The 7th IIEEJ International Conference on Image Electronics and Visual Computing (IEVC 2021), Sept. 2021.

[3] M. Watanabe, I. Igarashi, R. Tanida, and H. Kimata, "To That Place That Day: An Efficient Method for Merging Compressed Point Clouds to Create Point Cloud Data at Any Time and Place," Proc. of DICOMO 2021, pp. 357–362, June 2021 (in Japanese).

Jun Shimamura
Senior Research Engineer, Supervisor, Cyber-Physical System Group Leader, NTT Human Informatics Laboratories.
He received a B.E. in engineering science from Osaka University in 1998 and M.E. and Ph.D. from Nara Institute of Science and Technology in 2000 and 2006. He joined NTT Cyber Space Laboratories in 2000. His research interests include computer vision and mixed reality. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Ryuichi Tanida
Senior Research Engineer, NTT Human Informatics Laboratories.
He received a B.E. and M.E. in electrical engineering from Kyoto University in 2001 and 2003. Since joining NTT in 2003, he has been engaged in research on video coding and its application technology. His research interests include image processing, video coding, and sensing data coding technology. He is a member of IEICE.

Shingo Ando
Senior Research Engineer, Supervisor, NTT Human Informatics Laboratories.
He received a B.E. in electrical engineering from Keio University, Kanagawa, in 1998 and Ph.D. in engineering from Keio University in 2003. He joined NTT in 2003, where he has been engaged in research and practical application development in the fields of image processing, pattern recognition, and digital watermarks. He is a member of IEICE, the Institute of Image Electronics Engineers of Japan (IIEEJ), and the Institute of Image Information and Television Engineers (ITE).