

# Spatial-information Processing Technology for Establishing a 4D Digital Platform

*Yasuhiro Yao, Kana Kurata, Naoki Ito, Shingo Ando, Jun Shimamura, Mayuko Watanabe, Ryuichi Tanida, and Hideaki Kimata*

## Abstract

The four-dimensional (4D) digital platform integrates various sensing data about humans, things, and environments in real time into high-precision spatial information, enabling fusion with various industries' platforms and future predictions. The platform achieves these by matching and integrating 4D information such as latitude, longitude, altitude, and time with high accuracy on an advanced geospatial-information database. For maintaining an advanced geospatial-information database with high accuracy and abundant semantic information, this article introduces two spatial-information processing technologies: (i) real-world digitalization to detect features from images and sparse/low-density 3D point clouds and (ii) 4D-point-cloud coding that efficiently stores and uses 3D data including time variations.

*Keywords: 4D digital platform, real-world digitalization, 4D-point-cloud coding*

## 1. What is a 4D digital platform?

The four-dimensional (4D) digital platform collects various sensing data concerning people, things, and environments in real time, and matches and integrates 4D information such as latitude, longitude, altitude, and time with high accuracy. In other words, it is a platform that enables data fusion with various industrial platforms as well as future prediction (**Fig. 1**). By combining the 4D digital platform and various data collected from Internet of Things, the platform makes it possible to determine the exact position of various mobile objects in geographic space, and various future predictions based on that data will become possible. We thus believe that it can be used in various applications such as smoothing road traffic, optimizing the use of urban assets, and maintaining social infrastructures.

Regarding the elemental technologies that make up the 4D digital platform, namely, spatial-information

processing technologies necessary for constructing an advanced geospatial-information database with high accuracy and abundant semantic information, we are promoting the research and development of (i) real-world digitalization technologies for detecting features from images and sparse and low-density 3D point clouds and (ii) 4D-point-cloud coding technology for efficiently storing and using 3D data including time changes. We introduce both technologies and discuss the status of our efforts concerning them.

## 2. Real-world digitalization technologies

To construct an advanced geospatial-information database, it is necessary to maintain highly accurate 3D spatial information centered on roads; however, satisfying this requires enormous cost and effort. Specifically, a costly dedicated vehicle equipped with a sensing device called laser-imaging detection and ranging (LiDAR) and a manual process for generating

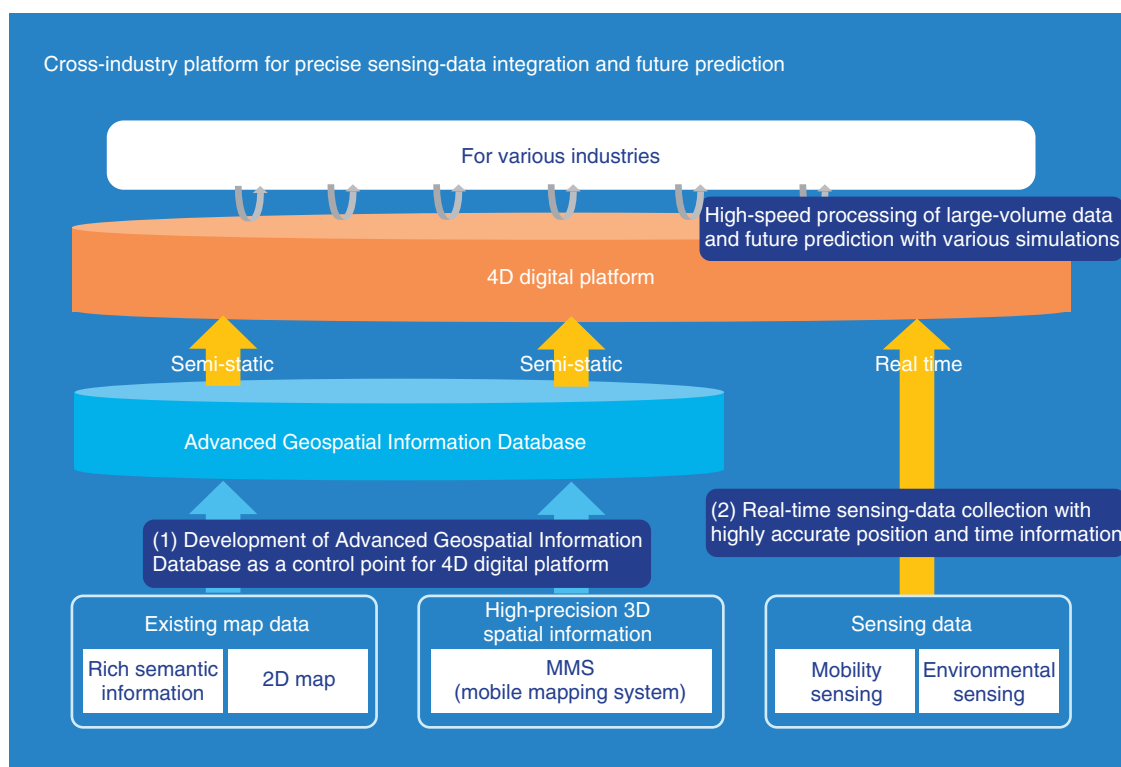


Fig. 1. Conceptual diagram of 4D digital platform.

maps are required.

To efficiently collect high-precision 3D spatial information, we are researching and developing real-world digitalization technologies that automatically and highly accurately detect various natural and artificial objects near roads from a combination of sparse and low-density 3D point clouds (measured with low-cost LiDAR) and images taken with a camera (**Fig. 2**).

Our real-world digitalization technologies include: (i) 3D-data high-resolution enhancement technology for generating high-resolution 3D data from sparse 3D point clouds based on images and videos; (ii) 3D-scene understanding and state-estimation technology for recognizing entire scenes (including complex shapes) and understanding states; and (iii) 3D-moving-object detection technology for using the sensing data obtained from moving-body sensors (in addition to 3D point clouds and images) to determine the position and orientation of vehicles and people. We also discuss our latest research achievements concerning our 3D-data high-resolution enhancement and 3D-scene understanding and state-estimation technologies.

## 2.1 3D-data high-resolution enhancement technology

The 3D-data high-resolution enhancement technology increases the resolution of 3D data, namely, a 3D point cloud with texture, from a combination of sparse and low-density 3D point clouds (measured by low-cost LiDAR) and images taken with a camera. Three-dimensional measurement with low-cost LiDAR provides sparse measurement results, and although 3D measurement is possible regardless of distance, the measurement results include noise. On the contrary, images taken with a camera contain dense data; however, stereoscopic 3D measurement using multiple images is not highly accurate for distant objects. We believe it is possible to generate 3D data with the same measurement accuracy as LiDAR and the same density as images taken with a camera while removing noise by processing the information from both LiDAR and camera data in an integrated manner.

We are working in stepwise on the research and development of the 3D-data high-resolution enhancement technology. Under the assumption that data are measured while a vehicle with in-vehicle sensors is

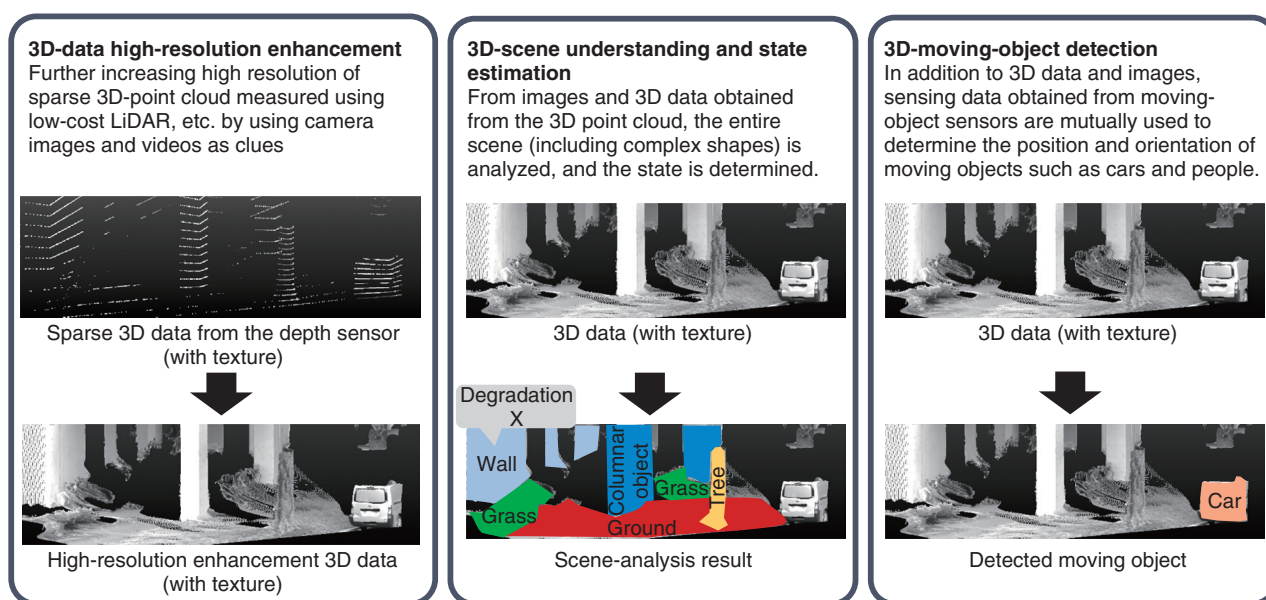


Fig. 2. Real-world digitalization technologies.

moving, the amount of information that we will integrate increases as follows for improving accuracy: (i) one image and one frame of LiDAR measurement data, (ii) multiple images and one frame of LiDAR measurement data, and (iii) multiple images and multiple frames of LiDAR measurement data that are continuous in a time series. Note that “one frame” of LiDAR measurement means data for one measurement over 360 degrees. Although it depends on the LiDAR equipment, LiDAR measures 360 degrees of the surroundings while rotating about 10 times per second.

We now introduce the 3D-data high-resolution enhancement technology, which derives high-density 3D data in real time without supervision from a single image and a sparse 3D point cloud measured by LiDAR.

First, a 3D point cloud measured by LiDAR is projected onto an image to generate an image that holds depth information called a depth map. The depth map created in this manner is referred to as a sparse depth map with many pixels that do not have depth values (Fig. 3).

This sparse depth map is processed using the input image as a clue to generate a dense depth map in which all pixels have a depth value. Such a method can be called *depth completion*. Even though depth completion had been developed, the conventional method generates a dense depth map for pixels that

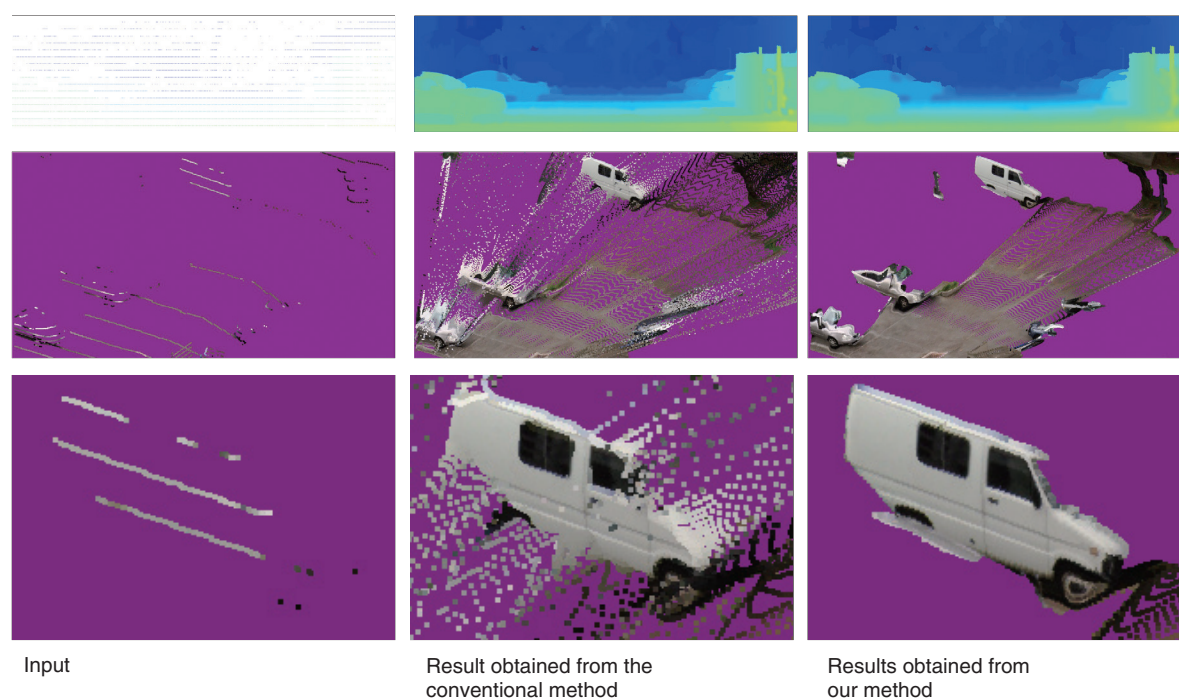
do not have depth by smoothly connecting the observed depth values [1]. This conventional depth completion is effective for complementing continuous observations between sparse observations; however, it smoothly interpolates the depth between different objects (Fig. 3).

To solve this problem, we previously proposed a method for smoothly connecting the observed depth values while adding a constraint that the change in depth becomes discontinuous when it straddles an object [2]. Compared with the conventional depth completion, our method not only improves accuracy but also obtains natural results when the depth map is visualized as 3D data (Fig. 3). In the future, we will further improve accuracy by increasing the amount of information to be integrated in the manner described above.

## 2.2 3D-scene understanding and state-estimation technology

The 3D-scene understanding and state-estimation technology is used for analyzing an entire scene (including complex shapes) and understanding its state. The aim with this technology is to automatically identify object areas such as buildings and roads from 3D data obtained from LiDAR and cameras and estimating states such as position and orientation.

The construction of an advanced geospatial information database faces two technical challenges: (i)



(Upper row: depth map; middle row: 3D data representation of “upper row”; lower row: enlarged view of “middle row”)

Fig. 3. 3D-data high-resolution enhancement technology using one image and one frame of LiDAR measurement.

being able to identify various natural and artificial objects near roads and (ii) keeping high classification accuracy while using data splitting and sampling to efficiently process large-scale (large-area and high-density) 3D data. To address the first challenge, we have been researching deep-learning-based point-cloud classification methods for making it easier to identify various objects. To address the second challenge, the problem that the identification accuracy decreases due to data splitting and sampling in deep-learning based methods needs to be solved.

We are working on solving this tradeoff between efficiency and classification accuracy. We experimentally identified that random sampling, often used to improve processing efficiency, deteriorates classification accuracy. Therefore, we are developing a deep-learning-based object classification model with sampling method that takes into account local shape around each 3D point of point clouds.

Specifically, high classification accuracy is achieved by executing the deep-learning-based classification process with a sampling method that preferentially leaves the points that have high distinguishability with respect to other points and do not change due to rotation or translation of the object instead of random

sampling [3]. This research is just beginning; accordingly, we would like to improve the performance of object identification through technical improvement and application in real environments.

### 3. 4D-point-cloud coding technology

In the real world, each person has a different purpose, and we use objects that have substance and take actions that suit our purposes. In this world, things change with the passage of time. There are various scales for purposes, objects, and units concerning the people involved, and the scales differ according to the action. We expect that the ability to acquire and reuse the states of things in the real world, which have different spatial and temporal scales, can provide a diversity of value to people. Aiming to use point clouds for various purposes, we are researching and developing 4D-point-cloud coding technology.

The method called LASzip has been applied for compressing a 3D point cloud. And the ISO/IEC<sup>\*1</sup> international standardization is currently promoting

<sup>\*1</sup> ISO/IEC: The International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC)



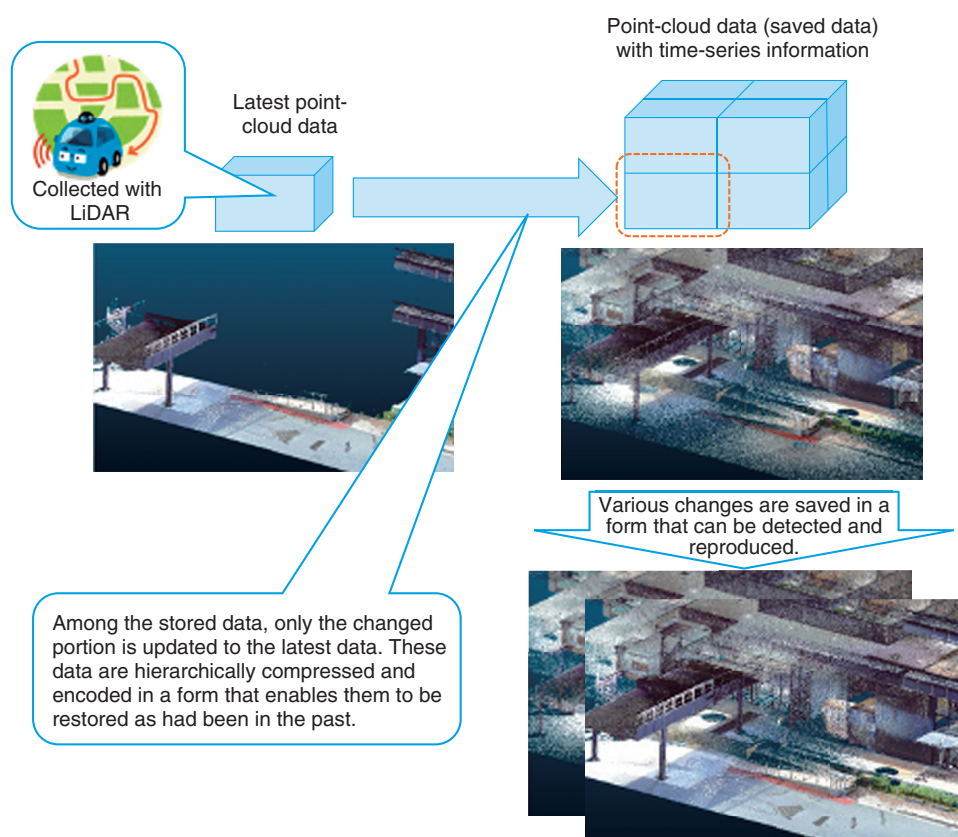


Fig. 4. Overview of 4D-point-cloud coding technology.

international standardization of a point-cloud coding method under the name MPEG<sup>\*2</sup> geometry-based point cloud compression (G-PCC). Neither method has a mechanism for preserving temporal changes, so they are insufficient for our purpose. We are researching and developing a representation and compression-coding method for point-cloud data that is based on our knowledge of video coding that expresses changes over time as differences in the latest point-cloud data.

The 4D-point-cloud coding technology is illustrated in **Fig. 4**. Since some of the spatial information can be obtained when acquiring point-cloud data, the entire space to be expressed is divided into a grid. When the grid is made, it has a structure of multiple sizes (spatial layers) in the form of recursive inclusion, just like a matryoshka doll. The grid of each minimum unit can possess point-cloud data, and the grid of the point-cloud data is expressed by a grid of the intermediate hierarchy. As a result, point-cloud data in the entire space can be compactly represented by a hierarchical grid. Also, if it is desirable to replace

some data with the latest point cloud data, point-cloud data partially included in the grid is replaced. While the point-cloud data are updated and the entire space is re-encoded, some of the past data are compressed and encoded as a difference in point-cloud data. As a result, the latest data can be decoded and expressed at any time, and the functionality that partially traces back to the past can be achieved. The functionality of compression coding as a difference in point-cloud data can be applied to quickly detect an object that was not there in the past. It is assumed that MPEG G-PCC will be used for compression encoding of point-cloud coordinate data.

We believe that the following use cases can be applied by compressing, encoding, and storing a point cloud by using this method. By acquiring point clouds from vehicles traveling on the same roads in a city every day, it becomes possible to obtain information

<sup>\*2</sup> MPEG: The Moving Picture Experts Group (MPEG) is a working group of ISO/IEC that develops standards for coded representation of digital audio and video and related data.

in real time about things that do not usually exist. Moreover, when a change is found, it becomes possible to reproduce the state before the change occurred retroactively in the past and simulate a secular change.

To apply these use cases, 4D-point-cloud coding technology, which can be used to store the point cloud (including temporal changes), will be indispensable. We will continue to promote research and development of 4D-point-cloud coding to make the real world more convenient.

#### 4. Future developments

For constructing an advanced geospatial information database with high accuracy and abundant semantic information, the following technologies are being developed: (i) real-world digitalization for detecting features with high accuracy from images and sparse and low-density 3D point clouds and (ii)

4D-point-cloud coding that efficiently stores and uses 3D data (including temporal changes).

We will continue to study the means of establishing each of these technologies and evaluate their performances with actual data through demonstration experiments, and in doing so, we will contribute to establishing a 4D digital platform.

#### References

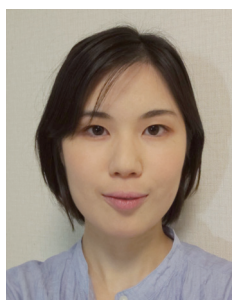
- [1] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruether, and H. Bischof, "Image Guided Depth Upsampling Using Anisotropic Total Generalized Variation," ICCV 2013, pp. 993–1000, Sydney, Australia, Dec. 2013.
- [2] Y. Yao, M. Roxas, R. Ishikawa, S. Ando, J. Shimamura, and T. Oishi, "Discontinuous and Smooth Depth Completion with Binary Anisotropic Diffusion Tensor," IEEE Robotics and Automation Letters, Vol. 5, No. 4, pp. 5128–5135, Oct. 2020.
- [3] K. Kurata, Y. Yao, S. Ando, and J. Shimamura, "Point Cloud Classification by Using Sampling Method for Complex Shapes," IPSJ SIG Technical Reports Computer Vision and Image Media, 2020-CVIM-220, pp. 1–6, Jan. 2020.



#### Yasuhiro Yao

Research Engineer, Universe Data Handling Laboratory, NTT Media Intelligence Laboratories.

He received a B.S. in pharmaceutical sciences from the University of Tokyo in 2007 and an M.E. in quantum engineering and system sciences from the University of Tokyo in 2010. He joined NTT in 2010, where he has been engaged in research and practical application development in computer vision and pattern recognition.



#### Kana Kurata

Staff, Universe Data Handling Laboratory, NTT Media Intelligence Laboratories.

She received a B.S. in earth and planetary sciences from Nagoya University in 2016 and an M.E. in environmental studies from Nagoya University in 2018. She joined NTT in 2018, where she has been engaged in research in computer vision and pattern recognition.



#### Naoki Ito

Senior Research Engineer, Universe Data Handling Laboratory, NTT Media Intelligence Laboratories.

He received an M.E. from Toyohashi University of Technology, Aichi, in 2001. He joined NTT the same year and engaged in research on character recognition. He moved to NTT EAST in 2004 and engaged in the development of security systems. He then moved to NTT Cyber Space Laboratories (now NTT Media Intelligence Laboratories) in 2008. He has been working on the development of real-world digitalization technologies.



#### Shingo Ando

Senior Research Engineer, Supervisor, Visual Media Project, NTT Media Intelligence Laboratories.

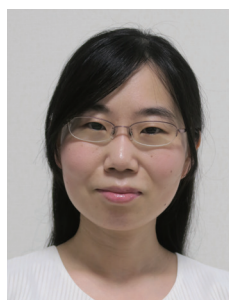
He received a B.E. in electrical engineering from Keio University, Kanagawa, in 1998 and a Ph.D. in engineering from Keio University, Kanagawa, in 2003. He joined NTT in 2003, where he has been engaged in research and practical application development in the image processing, pattern recognition, and digital watermarks. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Institute of Image Information and Television Engineers.



#### Jun Shimamura

Senior Research Engineer, Supervisor, Scene Analysis Technology Group Leader, NTT Media Intelligence Laboratories.

He received a B.E. in engineering science from Osaka University in 1998 and an M.E. and Ph.D. from Nara Institute of Science and Technology in 2000 and 2006. He joined NTT Cyber Space Laboratories (now NTT Media Intelligence Laboratories) in 2000. His research interests include computer vision and mixed reality. He is a member of IEICE and the Institute of Image Information and Television Engineers.



#### Mayuko Watanabe

Research Engineer, Signal Modeling Technology Group, NTT Media Intelligence Laboratories.

She received a B.S. and M.S. in mathematics from Tohoku University, Miyagi, in 2007 and 2009. She joined NTT in 2009, where she has been engaged in research and practical application development in video coding.



#### Ryuichi Tanida

Research Engineer, Visual Media Coding Group, Visual Media Project, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. in electrical engineering from Kyoto University in 2001 and 2003. He joined NTT to study video coding to develop a real-time H.264 software encoder application. He moved to NTT Advanced Technology Corporation in 2009 to manage various software development projects related to video compression, distribution, and TV phones. He has been at NTT Media Intelligence Laboratories since 2012 as a research engineer in the visual media coding group. His current research focuses on the development of HEVC (High Efficiency Video Coding) software encoders. He is a member of IEICE.



#### Hideaki Kimata

Senior Research Engineer, Supervisor, Signal Modeling Technology Group Leader, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. in applied physics, and a Ph.D. in electrical engineering from Nagoya University in 1993, 1995, and 2006. He joined NTT in 1995 and has been involved in the research and development of video coding, realistic communication, computer vision, and video recognition based on machine learning (deep learning). He is a chair of the IEICE Technical Committee on Image Engineering.