

## Respond to Any Requests without Expecting Something in Return to Build a Trust Relationship with Others



*Seishi Takamura*  
*Senior Distinguished Researcher,*  
*NTT Media Intelligence Laboratories*

### Overview

With the spread of the Internet, video data are becoming ubiquitous in our daily lives. Video-coding technology supports the stress-free situation in which we can send, receive, and use high-quality video via television broadcasting and the Internet. The amount of data is expected to increase to 40 YB (yottabytes) around 2040 due to the diffusion and sophistication of Internet-of-Things sensors, and technologies that support such huge amounts of data are attracting much attention worldwide. We interviewed Seishi Takamura, a senior distinguished researcher at NTT Media Intelligence Laboratories, about a coding technology for fully using a huge amount of data and his attitude as a researcher.

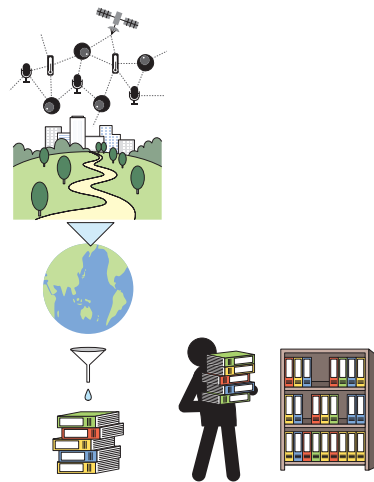
*Keywords: video coding, multimodal data, omni-ambient data*

### Being at the forefront of video-coding technology by taking a different approach

*—Please tell us about your current research.*

I'm currently researching omni-ambient data-organizing technology (**Fig. 1**). This technology is a kind of ultra-high-compression coding for fully using the huge amount of globally generated data (i.e., omni-ambient data), which is said to increase to 40 YB (yottabytes; 1 YB =  $10^{24}$  bytes) around 2040, without having to discard any. For storing and distributing such a huge amount of data, an approach based on information triage (i.e., selection of information)—

which reduces the amount of data to fit into the available storage capacity and transmission capacity—is currently the mainstream. With this approach, low-priority data are discarded in accordance with the rules of triage, but valuable data may be included in the discarded data. In contrast, our omni-ambient data-organizing technology stores and distributes a huge amount of data without the need to discard any by enhancing noise removal and information compression. It enables compression by 100 to 1000 times while maintaining higher quality than the current technology. If we use this technology to change the focus of the data-distribution infrastructure from data volume to data quality, we believe that we can



A technology that (i) processes and organizes an enormous amount of information acquired from multimodal sensors distributed on a global scale and (ii) records that information in an easy-to-use state.

Fig. 1. Concept of omni-ambient data-organizing technology.

develop various applications and create new business. In particular, omni-ambient data-organizing technology focuses on video data, which account for more than 80% of the total amount of data. Since the amount of video data used is increasing annually, organizing such data is a major advantage of this technology.

*—It seems that this technology will make our lives more comfortable and enjoyable. Please tell us about the key components.*

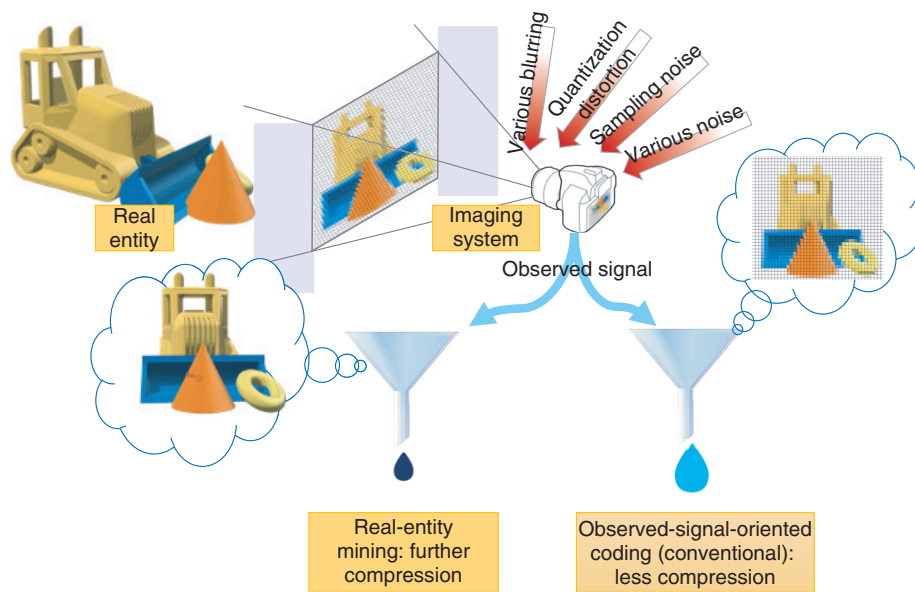
Omni-ambient data-organizing technology is composed of many technical elements, such as point-cloud processing and coding, ray-space processing and coding, error-control coding, 360° video processing and coding, and coding-oriented video generation. The key technology is real-entity mining (**Fig. 2**). Conventional coding technology encodes the captured video as is. However, real-entity mining removes disturbances such as noise, distortion, blurring, and flaws due to the lack of information from a captured video, infers the original image of the object, and encodes the inferred image. Since the state and information of materials and objects are extracted (i.e., inferred) as faithfully as possible to the original ones then encoded, the quality of decoded data exceeds that of the captured video, and an image that approaches the real object is reproduced. Of

course, the video data can be further compressed compared with conventional coding.

Although real-entity mining has several applications, we developed water-bottom video coding, which reproduces an image of the real entity of an object seen through fluctuation of the water surface (**Fig. 3**). A scene containing water is an important element that gives the video a sense of presence, however, it is particularly difficult to encode such a scene by using conventional coding. Our water-bottom video coding is highly regarded, for example, we received an award at an international conference, and we have also been invited to submit a paper for a prestigious journal. To further evolve this technology and compare its performance with that of the latest version of international-standard reference software, we are conducting an extremely time-consuming simulation that takes four to five months for one frame<sup>\*1</sup>.

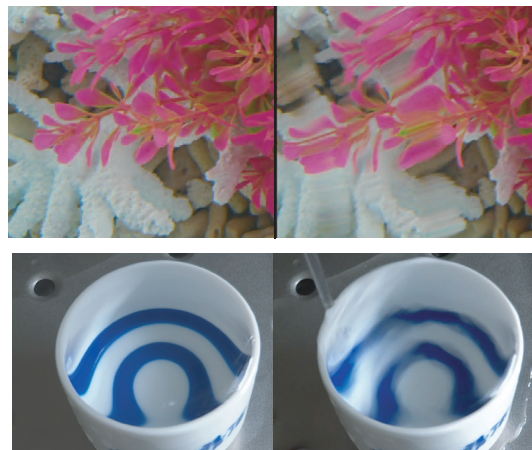
We are also working on point-cloud coding for video. Point clouds are a collection of points in a three-dimensional space that do not follow a grid pattern (i.e., the points exist at irregular positions), and they have a very high degree of freedom in terms of data-point dimension, color in accordance with view

\*1 Frame: The unit (i.e., one image) that composes a video. For example, Japanese broadcast video consists of 30 frames per second, and general movies consist of 24 frames per second.



We aim for higher-quality coding through data processing that focuses on the existence of an original object (real entity), instead of conventional coding that is faithful to the observed signal.

Fig. 2. Real-entity mining.



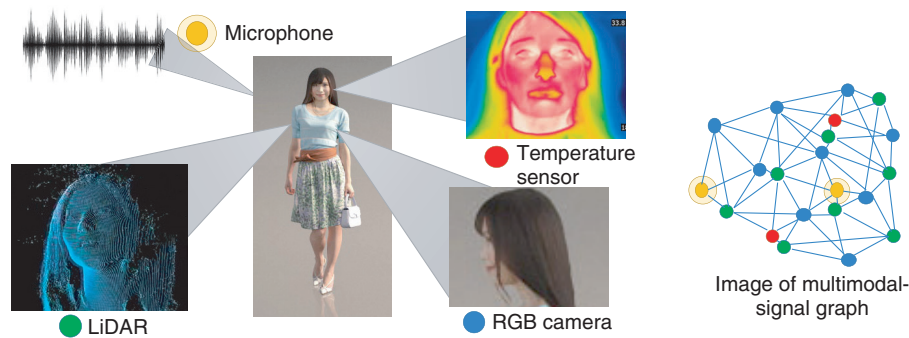
When water is still      When water is moving  
The shape of the bottom object changes drastically.

Fig. 3. Example images of water-bottom video.

angle, surface normal, etc., and represented by a large amount of data. Point-cloud coding is already an internationally standardized coding technology. However, taking a different approach, namely, lowering the degrees of freedom to increase compression rate and integrating information between multiple frames

of video, I'm wondering if it is possible to achieve the contradictory goals of improving quality and increasing compression ratio at the same time.

I'm also investigating multi-modal data compression. The information carried by multimodal signals acquired from an object to be observed from various



Multimodal signals obtained by measuring the same object with various sensors are diverse and have different sampling times and positions but are fundamentally correlated by the laws of physics.

LiDAR: light detection and ranging  
RGB: red, green, blue

Fig. 4. Graph signal processing.

locations with various sensors does not contain the temporal-sampling position or spatial-sampling position in a grid pattern, and the dimensions (temperature, coordinates, color, and sound pressure) also differ. Since there is only one real object to be observed, I think that there should be a correlation between these pieces of information, and finding it can be useful for reducing noise, predicting the future, and compressing the amount of data to make the data easier to use. However, a basic theory for handling such signals in an integrated manner is currently not available. Research on the framework called graph signal processing<sup>\*2</sup> (Fig. 4) is being actively conducted for a single-modal signal. The challenge is to develop a framework for multimodal signals.

The sensor, which is the source of data generation, might have sensed too much data for the object to be observed. Speaking of image sensors, I should mention a technology called compressed sensing, which makes it possible to obtain a fairly accurate image signal by calculation processing in the subsequent process even if some pixels are thinned out. I believe that reducing the amount of data before processing will become more important to reduce power consumption as Internet-of-Things sensors become ubiquitous. The key is to restore the signal in the later calculation process, and the cleverer we are at doing this calculation, the more accurate the observation data will be.

I believe that there are two challenges: improving the accuracy of signal restoration on the basis of a

new calculation principle and developing a cooperative restoration between different modal signals in a multimodal-sensor environment. We want to establish these technologies as international standards such as JPEG<sup>\*3</sup> and MPEG<sup>\*4</sup>. However, we have a long way to go and the hurdles are high, but that's why I find these challenges rewarding.

### Besides great successes, there are medium and small success.

—Can you tell us the lessons you have learned while working as a researcher?

For several years after joining NTT, I repeated trial and error without worrying about time. That approach is a privilege for young people, and I think it should be taken if time permits. However, as I get older, I realize that time is not infinite. In 2006, a former NTT researcher, Professor Hiroshi Ishii of the Massachusetts

\*2 Graph signal processing: Signal processing using graph, which is one of the methods for expressing the structure of information with nodes (points) and edges (lines or edges). It is a relatively new signal-processing technology used to analyze signals from social networks and sensor networks appropriately and efficiently.

\*3 JPEG: Standards for still image coding developed by the Joint Photographic Experts Group, a working group of International Organization for Standardization/International Electrotechnical Commission (ISO/IEC).

\*4 MPEG: Standards for audio and video coding data developed by the Moving Picture Experts Group, a working group of ISO/IEC.

Institute of Technology, visited our Yokosuka Research and Development Center and talked about his research strategy. Although Professor Ishii told us to plan our research by superposing our life span on a research timeline, I didn't immediately understand what he was saying. Now that I'm busy and have limited time for trial and error, I'm keenly aware that time is a finite resource. Since I have to carefully examine my research activities and act accordingly, I began to narrow my research themes. Fortunately, the direction of my selected themes did not deviate much from my research activities of the past; in fact, I think that lack of deviation is because I practiced trial and error when I was younger. The water-bottom video coding that I mentioned above is an example of how that approach worked. To investigate this technology, I had the rather twisted idea of "daring to increase the number of frames by one" to compress a given video. Then, after thinking from the standpoint of an encoder, I pondered the question of what kind of frame would be easier to encode, and I subsequently came up with the idea of letting the encoder generate (increase number of) frames and not stopping the generation of frames after one time, but repeating it many times.

It sounds easy to put it in words, but it took about 15 hours to generate a frame once, so it took a lot of patience and effort to repeatedly generate frames. Although this example is the story after narrowing down the theme, I think I wouldn't have come up with the idea without my experience of repeating trial and error.

I have hosted an international competition specializing in water-bottom video coding for two years, and our coding method surpassed the data produced by the winner of the competition. When I submitted a paper to an international conference, I had my first experience in which all three reviewers gave the highest evaluation and was commended as an "excellent paper" of the conference. I'm particularly honored that one of the authorities in this field, a respected Australian professor said, "It's an outstanding paper about a technology that can be used not only for the bottom of water but also for other purposes."

I think it is normal in the research world to have medium and small successes in addition to great successes. The wider the world you can see, the less attachments you will have, and the more challenges you can take on.

What I mean by "success" as a researcher is success when you obtain the results you originally intended or it may be success if you get unexpected by-prod-

ucts. On the contrary, I think that "failure" means that you cannot obtain results even if you spend a long time trying, the difficult situation continues, or you stop your research due to external pressure.

*—Could you tell us what you have cherished as a researcher?*

My idea of researching with an attitude of enhancing strengths rather than overcoming weaknesses that I talked about in my previous interview has not changed. Researchers aim to be "one of a kind," and if we are not prominent, we cannot survive, so this attitude is important.

In addition, I now feel the importance of give and give and not refusing requests as much as possible. When planning projects or events, you must make a request to various related parties. It is necessary to build relationships on a regular basis so that people will listen to your requests. For that reason, I try to respond to requests in a sincere manner. You will be able to understand the feelings of the person being asked and be less likely to make unreasonable requests. When building such relationships based on trust, I think it is very important not to think about getting something in return. If I receive a request when I am busy, even though it will take up my time, I try to respond to the request with an open mind.

On top of what I said above, I think that connecting with others is everything. I have undertaken quite a lot of external positions, including at academic societies, such as IEEE (Institute of Electrical and Electronics Engineers), and have developed relationships through those positions. Working as the secretary of an academic society led me to be invited by a professor whom I invited to the conference as a visiting researcher at his university.

Also, when talking to people outside the company, I am often surprised to hear that people become unexpectedly connected. Once I found out that the influential figure who was grateful to me at an academic society and a younger person in the company who was helpful to me at work were father and son, and the relationship with that younger person became closer. Having said that, I am not perfect, so there are some people who I don't get on with very well; however, I can learn by observing them. Many senior researchers are still working even as they approach 70, and I want to be like them. At the same time, I become inspired by communicating with people younger than me.

## The world is always looking toward “What to do next?”

*—You are nurturing not only research activities but also connections that transcend generations.*

I think personal connections are a truly precious asset. I also think it is important to covet achievements from research activities. In other words, “Only those who run after two hares will catch both hares.” For example, if the deadline for a paper overlaps the schedule of an event such as a get-together, most people would prioritize the paper, but I think it is better to start the paper after attending the event. You will never catch two hares if you act like as the proverb actually goes “If you run after two hares you will catch neither.” Perhaps I came up with this idea because I am surrounded by many talented people, and I realize that their achievements cannot be measured on the scale for ordinary people.

Any achievement will be a thing of the past from the moment you achieve it. Making great inventions, writing groundbreaking papers, and winning prestigious awards are achievements to be proud of. However, people are always looking at what you will do next. Therefore, I always keep in mind that achievements are things of the past to aim for the next target.

*—Please say a few words to our junior researchers.*

You can choose a role model or a person you admire and emulate them as you move forward; however, since each person is different, it is quite natural if you don’t end up like them. Although I was also aiming to emulate a certain person, when I look at my own footsteps, I see that I’m taking a different approach and producing different results from those of that person.

I think it’s OK to modify your goals from time to time in a variety of situations and after different experiences; in fact, you may obtain better results if you don’t stick to them. However, I feel that if you keep thinking about what you really want to achieve or want to be, it will come true unexpectedly, so I think it’s important to continue. You may be delighted when that happens, but it’s important to think that it’s a waypoint, not a goal. They say fortune is unpredictable and changeable, right?

I also think that many young people these days are too serious. In the past, there were more off-the-wall people in the laboratory. Being too serious seems to go against the present trend towards increasing diver-

sity. Even though you may cower when you get caught up in the results and evaluations, I think it is necessary to be a bit “off the wall” to overcome the difficult situations that you will one day encounter in your long research life. Research won’t necessarily go better even if you think about it logically. If you are looking for a “hit” among various possibilities, it might be better to do research while thinking as flexibly as possible. I want you to take it easy while making sure to keep your attitude sincere.

### ■ Interviewee profile

#### Seishi Takamura

Senior Distinguished Researcher, Signal Modeling Technology Group, Universe Data Handling Laboratory, NTT Media Intelligence Laboratories.

He received a B.E., M.E., and Ph.D. from the Department of Electronic Engineering, Faculty of Engineering, the University of Tokyo, in 1991, 1993, and 1996. His current research interests include efficient video coding and ultrahigh-quality video processing. He has fulfilled various duties in the research and academic community in current and prior roles, including serving as associate editor of IEEE Transactions on Circuits and Systems for Video Technology (2006–2014), editor-in-chief of the Institute of Image Information and Television Engineers (ITE), executive committee member of the IEEE Region 10 and Japan Council, and director-general of ITE affairs. He has also served as chair of ISO/IEC Joint Technical Committee (JTC) 1/ Subcommittee (SC) 29 Japan National Body, Japan head of delegation of ISO/IEC JTC 1/SC 29, and as an international steering committee member of the Picture Coding Symposium. From 2005 to 2006, he was a visiting scientist at Stanford University, CA, USA.

He has received 57 academic awards including ITE Niwa-Takayanagi Awards (Best Paper in 2002, Achievement in 2017), the Information Processing Society of Japan (IPSJ) Nagao Special Researcher Award in 2006, Picture Coding Symposium of Japan (PCSJ) Frontier Awards in 2004, 2008, 2015, and 2018, the ITE Fujio Frontier Award in 2014, and the Telecommunications Advancement Foundation (TAF) Telecom System Technology Awards in 2004, 2008, and in 2015 with highest honors, the Institute of Electronics,

Information and Communication Engineers (IEICE) 100-Year Memorial Best Paper Award in 2017, the Kenjiro Takayanagi Achievement Award in 2019, and Industrial Standardization Merit Award from Ministry of Economy, Trade and Industry of Japan in 2019 (as an individual) and in 2020 (as NTT team).

He is an IEEE Fellow, IEICE Fellow, senior member of IPSJ, and member of Japan Mensa, the Society for Information Display, the Asia-Pacific Signal and Information Processing Association, and ITE.