Next-generation Media Transport MMT for 4K/8K Video Transmission

Takayuki Nakachi, Takahiro Yamaguchi, Yoshihide Tonomura, and Tatsuya Fujii

Abstract

MPEG Media Transport (MMT) for heterogeneous environments is being developed as part of the ISO/IEC (International Organization for Standardization/International Electrotechnical Commission) 23008 standard. In this article, we overview the MMT standard and explain in detail the MMT LDGM (low-density generator matrix) FEC (forward error correction) codes proposed by NTT Network Innovation Laboratories. We also explain remote collaboration for content creation as a use case of the MMT standard.

Keywords: MPEG Media Transport, Low-density Generator Matrix Codes, ISO/IEC

1. Introduction

It has been about 20 years since the widely used MPEG-2 TS (MPEG-2 Transport Stream) standard[*1] was developed by ISO/IEC (International Organization for Standardization/International Electrotechnical Commission) MPEG (Moving Picture Experts Group). Since then, media content delivery environments have changed. While video signals have been diversified, 4K/8K video systems have been developed. Moreover, there are a wide variety of fixed and mobile networks and client terminals displaying multi-format signals. The MPEG Media Transport (MMT) standard [1] is being developed as part 1 of ISO/IEC 23008 for heterogeneous environments. MMT specifies technologies for the delivery of coded media data for multimedia services over concatenated heterogeneous packet-based network segments including bidirectional IP networks and unidirectional digital broadcasting networks. NTT Network Innovation Laboratories has developed 4K transmission technologies for digital cinema, ODS (other digital stuff)[*2], super telepresence, and other applications. We have contributed to the standardization of MMT in order to diffuse 4K transmission core technologies such as forward error correction (FEC) codes and layered signal processing.

2. MMT overview

MPEG-2 TS provides efficient mechanisms for multiplexing multiple audio-visual data streams into one delivery stream. Audio-visual data streams are packetized into small fixed-size packets and interleaved to form a single stream. This design principle means that MPEG-2 TS is effective for streaming multimedia content to a large number of users.

In recent years, it has become clear that the MPEG standard is facing several technical challenges due to the emerging changes in multimedia service environments. One particular example is that the pre-multiplexed stream of 188-byte fixed size MPEG-2 TS packets is not quite suitable for IP (Internet Protocol)-based delivery of emerging 4K/8K video services due to the small and fixed packet size and the rigid packetization and multiplexing rules. In addition, it might be difficult for MPEG-2 TS to deliver multilayer coding data of scalable video coding (SVC) or multiview coding (MVC) via multi-delivery channels. Error resiliency is also insufficient.

To address these technical weaknesses of existing

---

*1 MPEG-2 TS: A standard format for transmission developed by ISO/IEC MPEG. It is used for the current digital terrestrial broadcasting.
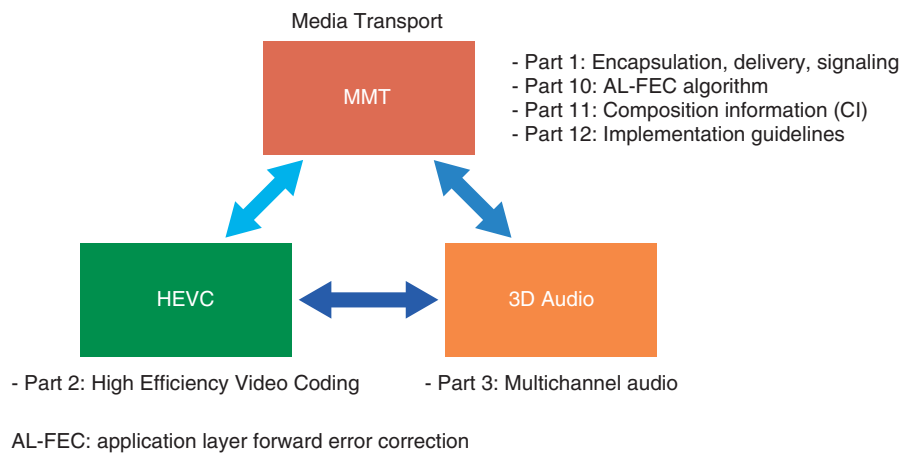*2 ODS: A live-streaming application for theaters.

Media Transport

MMT

- Part 1: Encapsulation, delivery, signaling
- Part 10: AL-FEC algorithm
- Part 11: Composition information (CI)
- Part 12: Implementation guidelines

HEVC

3D Audio

- Part 2: High Efficiency Video Coding   - Part 3: Multichannel audio

AL-FEC: application layer forward error correction

Fig. 1.   Overview of MPEG-H.

① Encapsulation

Legacy media   File   Media CODEC

Media fragment unit (MFU)

Media processing unit (MPU)

③ Signaling

MMT signaling

MMT payload

② Delivery

MMT packet   RTP, HTTP

UDP, TCP

IP packet

Delivery format

Package

Storage format

⬛ : Outside the scope of MMT

HTTP: Hypertext Transfer Protocol
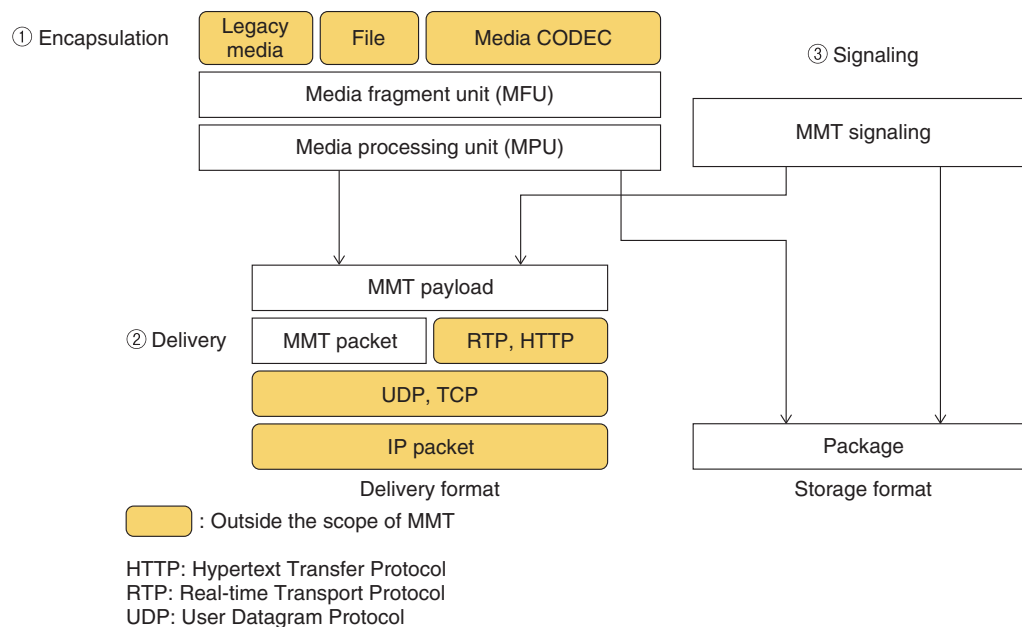RTP: Real-time Transport Protocol
UDP: User Datagram Protocol

Fig. 2.   MMT protocol stack.

standards and to support the wider needs of network-friendly transport of multimedia over heterogeneous network environments including next-generation broadcasting system, MPEG has been developing transport and synchronization technologies for a new international standard, namely MMT, as part of the ISO/IEC 23008 High Efficiency Coding and Media Delivery in Heterogeneous Environments (MPEG-H) standard suite. An overview of the MPEG-H standard is shown in **Fig. 1**. The MMT standard consists of

Parts 1, 10, 11, and 12.

The protocol stack of MMT, which is specified in MPEG-H Part 1 [2] is shown in **Fig. 2**. The white boxes indicate the areas in the scope of the MMT specifications. MMT adopts technologies from three major functional areas. 1) Encapsulation: coded audio and video signals are encapsulated into media fragment units (MFUs)/media processing units (MPUs). 2) Delivery: an MMT payload is constructed by aggregating MPUs or fragmenting one MFU. The
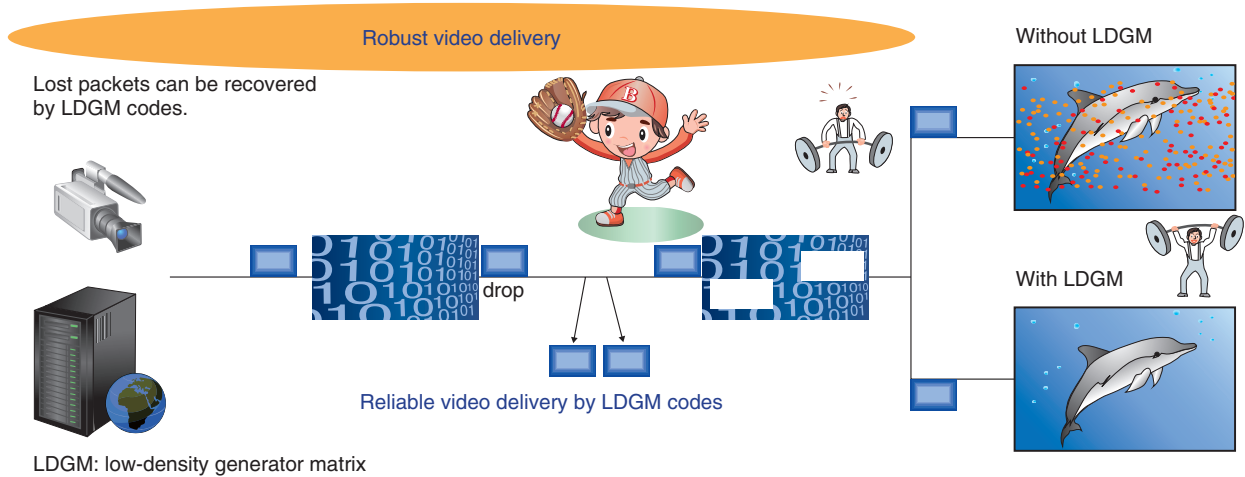
Fig. 3.   Reliable video transmission.

size of an MMT payload needs to be appropriate for delivery. An MMT packet is carried on IP-based protocols such as the User Datagram Protocol (UDP) or Transmission Control Protocol (TCP). It is a variable-length packet appropriate for delivery in IP packets. Each packet contains one MMT payload, and MMT packets containing different types of data and signaling messages can be transferred in one IP data flow. 3) Signaling: MMT signaling messages provide information for media consumption and delivery to the MMT client.

A client terminal identifies MPUs constituting the content and their presentation time by processing the signaling message. The presentation time is described on the basis of Coordinated Universal Time (UTC)[*3]. Therefore, the terminal can consume MPUs in a synchronized manner even if they are delivered on different channels from different senders. MMT defines application layer forward error correction (AL-FEC) codes in order to recover lost packets, as shown in **Fig. 3**. MPEG-H Part 1 specifies an AL-FEC framework, and MPEG-H Part 10 [3] specifies AL-FEC algorithms. Furthermore, MPEG-H part 11 defines composition information (CI), which identifies the scene to be displayed using both spatial and temporal relationships among content. These functions mentioned above are the novel features of MMT, and are not shared by MPEG-2 TS.

3.   Powerful FEC codes

MPEG-H Part 10 defines several AL-FEC algo-

rithms, including Reed-Solomon (RS) codes[*4] and low-density generator matrix (LDGM) codes proposed by NTT Network Innovation Laboratories. Each AL-FEC code has advantages and disadvantages in terms of error recovery performance and computation complexity. RS codes are based on algebraic structures. They give an optimal solution in terms of MDS (minimum distance separation) criteria. However, an RS code makes it difficult to increase the block length because the decoding process features polynomial-time decoding. It cannot reach Shannon's capacity[*5], which gives the theoretical limit. This can result in low coding efficiency and excessive computation overhead. Unlike RS codes, LDGM codes can handle block sizes of over several thousand packets because of their very low computation complexity. In particular, LDGM codes are suitable for the transmission of huge data sets such as 4K/8K video (**Fig. 4**).

LDGM codes are one type of linear code; the party check matrix H[*6] consists of a lower triangular matrix (LDGM structure) that contains mostly 0's and only

---

*3   UTC: UTC is the time standard commonly used across the world. It is used to synchronize time across Internet networks. In the MMT standard, it can be used as a timestamp.

*4   Reed-Solomon (RS) codes: RS codes are error-correcting codes in which redundant information is added to data so that it can be reliably recovered despite errors in transmission. Decoding is time-consuming when the block length is increased.

*5   Shannon's capacity: Error-correcting codes can alleviate some errors, but cannot alleviate all the errors introduced by the channel in a digital communications system. Shannon's capacity theorem states that error-free transmission is possible as long as the transmitter does not exceed the channel's capacity.
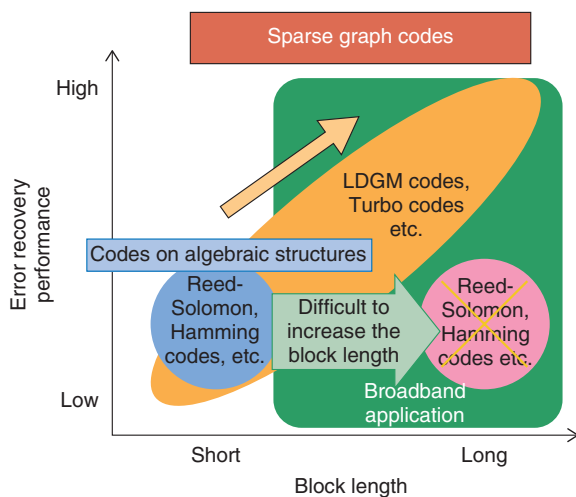
Fig. 4.   FEC codes.

a small number of 1's. An example of LDGM encoding and decoding procedures is shown in **Fig. 5**. Using a sparse parity generator matrix makes it possible to process large code blocks of over a thousand IP packets as a single code block, which offers robustness to packet erasure error in networks. Our proposed LDGM codes provide good error recovery performance while keeping the computational complexity low because the sub-optimal parity generator matrix is used when applied to the message passing algorithm (MPA). Furthermore, the proposed LDGM codes can use irregular matrices that provide good error recovery performance for both MPA and MLD (maximum likelihood decoding).

Furthermore, the specified LDGM codes can support the following two schemes. One is a sub-packet division and interleaving method [4]. Generally, LDGM codes provide superior error recovery performance for large code-block sizes. However, the error recovery performance for short code-blocks is inferior to that of the RS codes. The proposed sub-packet division and interleaving LDGM scheme solves this problem. This scheme increases the number of symbols in one block size and increases the error recovery performance.

The other is a Layer-Aware LDGM (LA-LDGM) scheme [5]. The structure of conventional LDGM codes does not support partial decoding. When conventional LDGM codes are used for scalable video data such as JPEG2000 created by JPEG (Joint Photographic Experts Group) or SVC code streams, performance is low and scalability is lost. The LA-

LDGM scheme maintains the corresponding relationship of each layer. The resulting structure supports partial decoding. Furthermore, the LA-LDGM codes create highly efficient parity data by considering the relationships of each layer. Therefore, LA-LDGM codes raise the probability of recovering lost packets.

4.   Use cases of MMT standard

MMT has various functions. Several representative MMT functions are shown in **Fig. 6**. Reliable 4K/8K video transmission via public IP networks is one of the typical features of MMT due to its high error recovery capability. Synchronization[*7] between the content on a large screen and the content on a second display can also be realized by MMT. Users can enjoy public viewing while displaying alternative camera views on their second display. Different content selected by a user can be simultaneously transported via different networks. Hybrid delivery of content in next-generation Super Hi-Vision broadband systems can also be achieved. Applications such as widgets can be used to present information together with television (TV) programs. In particular, the large and extremely high-resolution monitors needed for displaying Super Hi-Vision content would be able to present various kinds of information obtained from broadband networks together with TV programs from broadcasting channels.

5.   Remote collaboration for content creation [6]

High quality video productions such as those involving films and TV programs made in Hollywood are based on the division of labor. The production companies for video, visual effects (VFX), audio, and other elements of a production all create their content in different locations. Post-production is the final stage in filmmaking, in which the raw materials are edited together to form the completed film. In conventional program production, as shown in **Fig. 7(a)**, these raw materials are gathered physically from the separately located production sites. An overview of

*6   Parity check matrix H: The parity check matrix defines the relationships between the encoding symbols (source symbols and repair symbols), which are used by the decoder to reconstruct the original k source symbols if some of them are missing.
*7   Synchronization: Several media content files that have been delivered via various networks can be synchronized by using UTC based timestamps. However, the method of implementing the synchronization is outside the scope of the MMT standard.
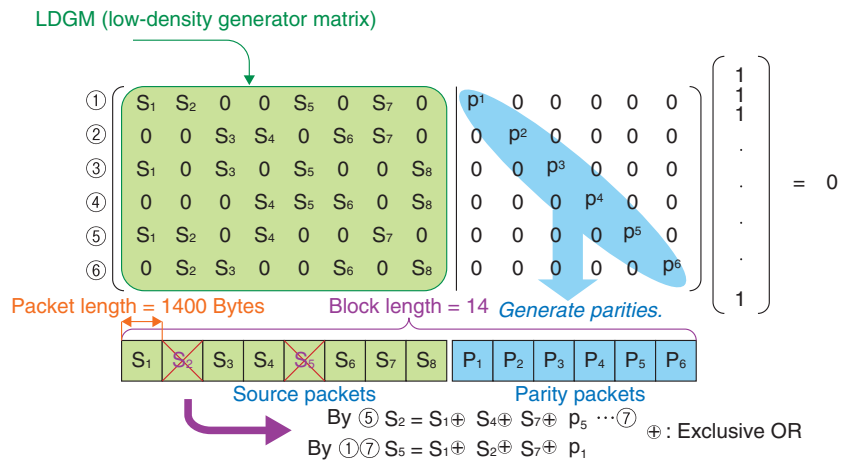
Fig. 5.   LDGM codes.
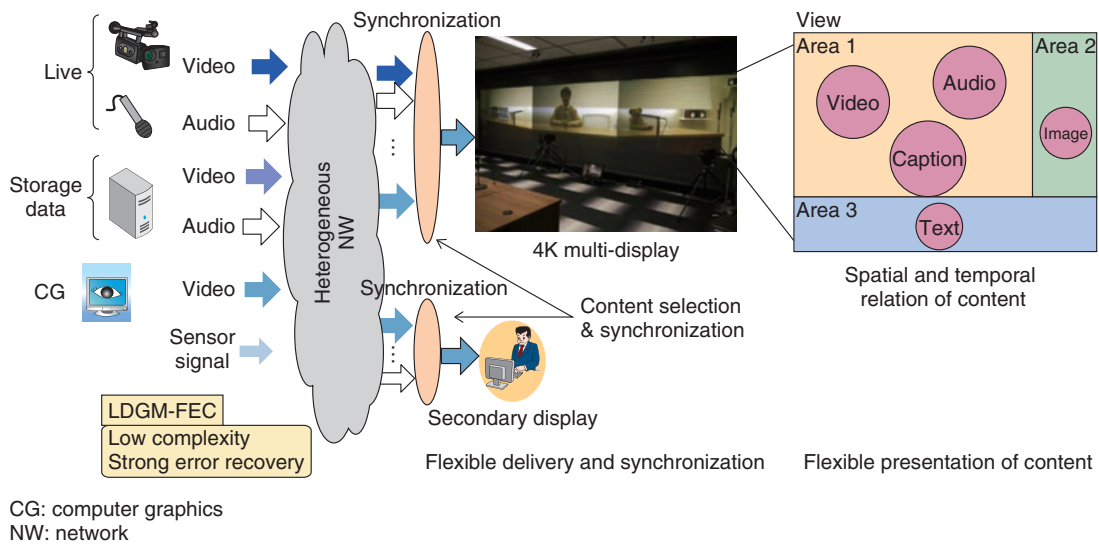

CG: computer graphics
NW: network

Fig. 6.   MMT functions.

remote collaboration for content creation and editing actions on a timeline is shown in **Fig. 7(b)** and **(c)**. By using the MMT standard, multiple forms of content such as video, VFX, and audio are shared via the network, and the selected content can then be synchronized based on the producer's request. All the staff members at each location perform their tasks and share their comments simultaneously. A remote collaboration system that uses MMT enhances the speed of decision-making. As a result, we can increase the efficiency and productivity of content creation. This is an example use case of the MMT standard. Various

other services are expected to emerge with the full implementation of MMT standard technologies.

6.   Future direction

NTT Network Innovation Laboratories will continue to promote innovative research and development of reliable and sophisticated transport technologies such as stable video transmission on shared networks consisting of multi-domain networks and virtual network switching to achieve dynamically configurable remote collaboration.
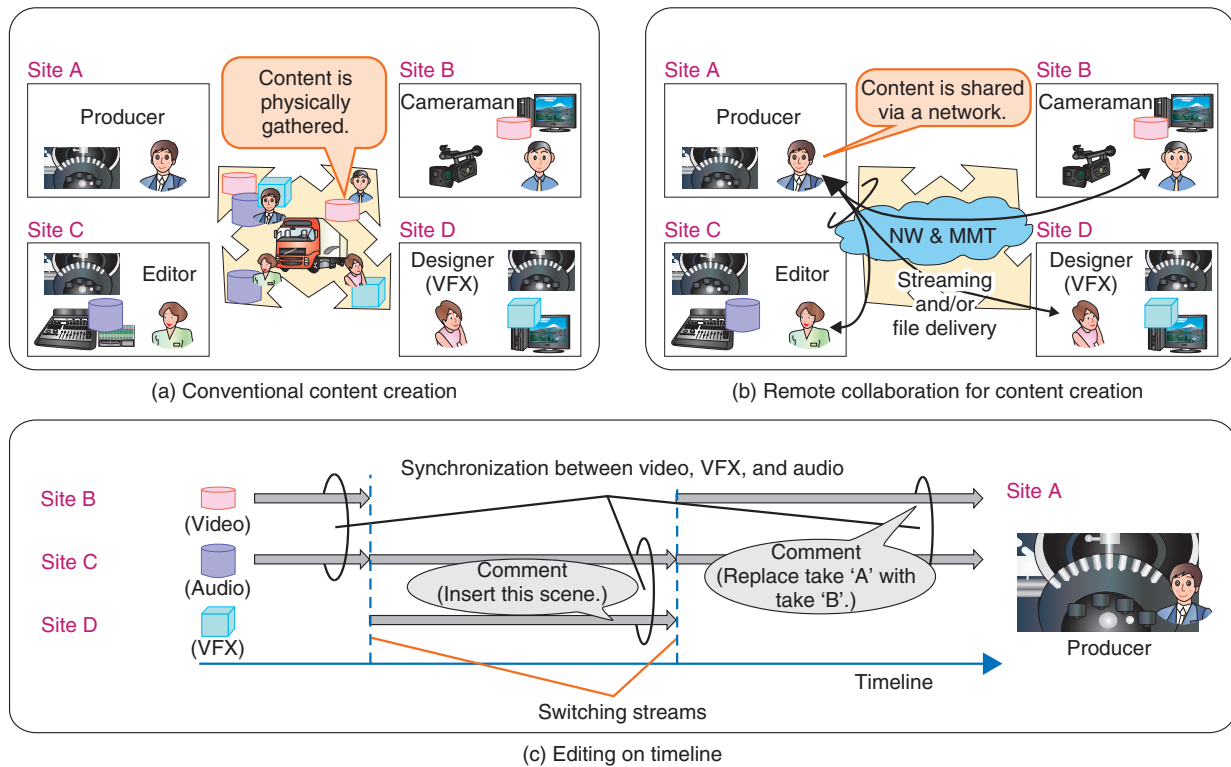
(a) Conventional content creation

(b) Remote collaboration for content creation

(c) Editing on timeline

Fig. 7.   Remote collaboration for content creation.

References

[1] T. Nakachi, "An Emerging MMT Standard for a Next Generation Media Platform," The 26th Communication Systems Workshop (CSWS), Nov. 2013.

[2] ISO/IEC 23008-1 FDIS, Information technology - High Efficiency Coding and Media Delivery in Heterogeneous Environments - Part 1: MPEG Media Transport (MMT).

[3] ISO/IEC 23008-10 DIS, Information Technology - High Efficiency Coding and Media Delivery in Heterogeneous Environments - Part 10: MPEG Media Transport Forward Error Correction Codes (MMT FEC Codes).

[4] T. Tonomura, D. Shirai, K. Kitamura, T. Nakachi, T. Fujii, and H. Kiya, "Construction Method and Theoretical Analysis of Packet-level Low-density Generator Matrix Codes to Allow Backward Compatibility for Video Streaming," IEICE Trans. Fundamentals (Japanese edition), vol. J93-A, no. 3, pp. 212–215, March 2010.

[5] Y. Tonomura, D. Shirai, T. Nakachi, T. Fujii, and H. Kiya, "Layered Low-density Generator Matrix Codes for Super High Definition Scalable Video Coding System," IEICE Trans. on Fundamentals, vol. E92-A, no.3, pp. 798–807, March 2009.

[6] T. Nakachi, Y. Tonomura, and T. Fujii, "A Conceptual Foundation of NSCW Transport Design Using an MMT Standard," IEEE ICSPCS2013, Dec. 2013.

Takayuki Nakachi

Senior Research Engineer, Media Processing Systems Research Group, NTT Network Innovation Laboratories.

He received a Ph.D. degree in electrical engineering from Keio University, Tokyo, in 1997. Since joining NTT in 1997, he has been engaged in researching super high definition (SHD) image coding, especially in the area of lossless and near-lossless coding. In recent years, he has been researching scalable image/video coding in order to distribute SHD image content. From 2006 to 2007, he was a visiting scientist at Stanford University, CA, USA. He is currently Senior Research Engineer of the Media Processing Systems Research Group in NTT Network Innovation Laboratories. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and IEEE.

Takahiro Yamaguchi

Senior Research Engineer, Supervisor, NTT Network Innovation Laboratories.

He received his B.E., M.E., and Ph.D degrees in electronic engineering from the University of Electro-Communications, Tokyo, Japan, in 1991, 1993, and 1998, respectively. He joined NTT Optical Network Systems Laboratories in 1998 and has been researching SHD image distribution systems. He is a member of IEICE and the Information and Television Engineers (ITE) of Japan.

Yoshihide Tonomura

Research Engineer, NTT Network Innovation Laboratories.

He received his B.S. and M.S. degrees in electronics engineering from Nagaoka University of Technology, Niigata, in 2002 and 2004, respectively, and the Ph.D. degree from Tokyo Metropolitan University in 2010. He joined NTT Network Innovation Laboratories in 2004. His research is focused on image processing theories and applications. From 2011 to 2012 he was a visiting researcher in the Media Lab of Massachusetts Institute of Technology, Boston, MA, USA. He is a member of IEICE.

Tatsuya Fujii

Group Leader, Media Processing Systems Research Group, Director of Digital Cinema Project, NTT Network Innovation Laboratories.

He received his B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Tokyo in 1986, 1988, and 1991, respectively. He joined NTT in 1991. He has been researching parallel image processing and SHD image communication networks. In 1996, he was a visiting researcher at Washington University in St. Louis, MO, USA. He is currently a group leader of the media processing systems research group and a director of the Digital Cinema Project in NTT Network Innovation Laboratories. He is a member of IEICE, ITE of Japan, and IEEE.