

특집논문-04-09-2-01

## The Development of an Object-linked Broadcasting System

Yasuaki Kanatsugu<sup>a)†</sup>, Toshihiko Misu<sup>b)†</sup>, Masaki Takahashi<sup>b)†</sup>, Seiichi Gohshi<sup>b)†</sup>

### Abstract

We have proposed an Object-linked Broadcasting Service that displays various data related to the object onscreen. In this paper, we describe the structure of the Object-linked Broadcasting System that will enable our proposal to be realized, and report new techniques that we have developed to create the system. We have carried out the experiment to confirm the system performance as well as execution of each technology assembling the system. We have confirmed that the performance of the system we developed satisfies the proposed specification based on user requirements and current technology.

Keywords : digital TV, digital broadcasting, meta-data, finger pointer, pointing device, video object, object extract

### I. Introduction

The media environment is changing dramatically all over the world. In particular, the digitization of broadcasting makes it possible to produce and deliver multi-media in new ways. In Japan, digital satellite broadcasting started in December 2000, and digital terrestrial broadcasting has just started in December 2003. Digital broadcasting has a lot of possibilities to supply new services for various kinds of viewers. We have proposed "Object-linked data broadcasting service" as one of these services<sup>[1][2]</sup>. The service provides information related to an object in a TV picture that a viewer is interested in.

There are digital TV programs that introduce similar services. "Televito.com", "NBC site of Amazon.com" and "TV-Market.com" provide goods that appear in a TV program. These services do not link directly to a TV program. The service we have proposed gives

viewers detailed information about the object that they select in the TV picture, as each object in the program is linked to the information. We have developed a system to realize this service. In the system, the detailed information is transmitted as data stream in the digital broadcasting, and is displayed on the TV picture via a receiver.

The system is composed of several technologies that should be developed as follows. First, we developed an object extraction and tracking technique. Some segmentation and tracking algorithms have already developed for particular object. However, various kinds of objects have to be extracted and tracked from any type of picture for the broadcasting. Secondly, we provided a data format to efficiently transmit meta-data and describe how to display them. In order to determine the format for transmission using a broadcasting band, we set up several constraints of standards and regulations. Bandwidth of the data stream is also limited. Finally, we have developed a new finger pointer as a remote control device on the receiver side.

This paper is organized as follows. In Section 2, we

a)NHK Engineering Services, INC

b)NHK Science and Technical Research Laboratories

† This research was supported by TAO, Tele-communications Advancement Organization

describe the system configuration and each technology that the system is composed of. In Section 3, we explain the experiment carried out to confirm the system performance. The experimental results are discussed in Section 4. Finally, we conclude this paper in Section 5.

## II. System configuration and its technology

### 1. System configuration

The concept behind our new data broadcasting service is to provide a system that any person can easily use. We determined the system configuration based on a consideration on both user requirements and currently technology. Figure 1 shows the system configuration.

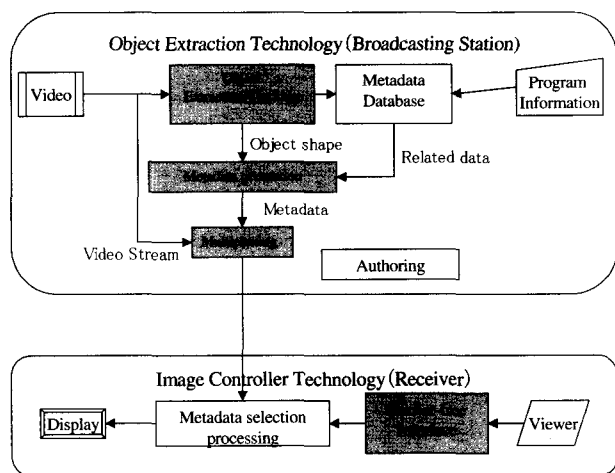


Fig. 1. System configuration

The system consists on a provider side and receiver side. In the provider side, the technologies we developed are "The method for object extraction and tracking", "Meta-data database" for the input and retrieval of data related objects, and "Meta-data description" for the multiplex and the transmission of the data. On the receiver side, "Metadata processing" and "Human interface" for choosing an object were developed. The

details of each technology are explained in the following section.

### 2. Object extraction and tracking technique

We discussed some algorithms to extract an object in the diverse pictures found in a broadcasting station. There are object extraction methods for the individual object that have already been developed. However, a wide variety of pictures are used for broadcasting. Broadcasting stations require technologies for extracting various kinds of objects from any types of picture. We exploited a feature of the broadcasting station to extract an object which has a complicated movement. We have developed the object extraction technique using several cameras to continue to follow the object even if the object disappears and reappears on the screen<sup>[3]</sup>. Using camera parameters and several cameras makes the method possible. If an object disappears from a picture of a camera, the position of the object can be estimated using other camera picture and the camera parameters. This process can work in real time. The results of the object extraction are shown in Figure 2.

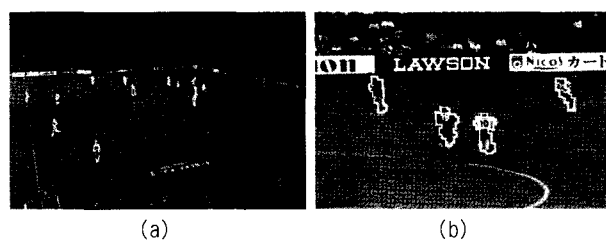


Fig. 2. Results of the object extraction (a) Wide camera scene (b) Camera scene for broadcasting

### 3. Meta-data description method

We have developed a new description method for transmission and display of the meta-data on objects. For efficient transmission, we determined a data format for multiplexing meta-data with the video and audio streams. The process for displaying the meta-data is

described in a markup language.

We have to take account of standards and regulations for digital broadcasting, because compatibility with the already existent system is concern in creating system. The policies for development are below.

- BML should be employed for display.
- ECMA script should be used for processing.
- Binary Table defined by ARIB should be used.
- DSM-CC should be used for transmission.

Broadcasting Markup Language (BML) is a multi-media coding method based on Extensible Markup Language (XML), and is defined for displaying data broadcasting by the Association of Radio Industries and Businesses (ARIB) in Japan. We described how to display the data related to the object using BML. ECMA Script is a program language for processing. We use ECMA Script to determine which object matches the cursor. The binary table is one of mono-media that contains binary data with fixed length format. We formatted three types of binary tables for the meta-data of objects including the object shape data, related information to the object and so on. The structure of the binary table is shown in Figure 3.

| Recordlength | Objct ID | Objct name | Related data | Refered objct | Pointer of multimedia data |
|--------------|----------|------------|--------------|---------------|----------------------------|
|              |          |            |              |               |                            |

Master table

| Number of vertices | Horizontal position of vertex 1 | Vertical position of vertex 1 | ..... | Horizontal position of vertex N | Vertical position of vertex N |
|--------------------|---------------------------------|-------------------------------|-------|---------------------------------|-------------------------------|
|                    |                                 |                               |       |                                 |                               |

Object shape table

| Time | Number of objcts |
|------|------------------|
|      |                  |

Time table

Figure 3 Structure of binary tables

ARIB defines Digital Storage Media Command and Control (DSM-CC) for sending file-typed contents such as binary tables in digital broadcasting. We employed DSM-CC for transmission of BML and binary tables. Set Top Box(STB) obtains the cursor position from the finger pointer and location of the object from the binary table. When the cursor is located at the object position, the related information is displayed. ISO/IEC defines MPEG7 for metadata description. However, we do not introduce MPEG7, because there are currently no digital receivers which use MPEG7 in Japan. One of the future studies is a consideration of the introduction of MPEG7 because of its status as an international standard.

We found that the size of the object shape data was too large to transmit in the data stream. To reduce the data size, the object shape was approximated to polygon. The apexes of the polygon are transmitted as the object shape data instead of a binary picture for object shape. It became clear that a 50 sided polygon is enough to describe the shape. We sent the object shape data one frame per second to reduce the bit rate.

#### 4. Barrier free interface

We have developed a remote control device for any kind of person in consideration of the environment for TV viewing. Digitalization of broadcasting makes it difficult to operate the TV, so that remotecontrol devices become complex. Our target is to develop an interface that allows any kind of person to control a TV set easily. The environment for TV viewing is different from that of PC. The concepts of the development are below.

- Viewers can operate the TV from far away it without touching it.
- Viewers do not have to wear any devices.
- No calibration and no presetting.
- Real-time operation
- Small size for easy TV integration

We have developed a finger pointer that makes possible to select an object in a picture when viewer points at it<sup>[4]</sup>. The position of the cursor is estimated using both eyes and fingertip position. This method is based on the result of a subjective evaluation. The result shows that the position of the cursor depends on both eyes and fingertip position. No calibration is necessary and viewer can be moving. Voice recognition was used to choose an object when viewer points at the object. The process for the finger pointer is shown in Figure 4.

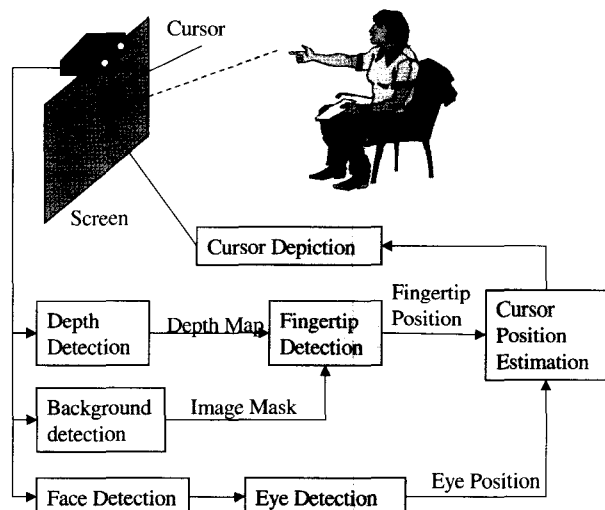


Fig. 4. Process of finger pointer

### III. Experiment for system performance

We have carried out the experiment to confirm the system performance as well as execution of each technology assembling the system. We have developed each component for the system, and evaluated the system ability by exchanging data among the components. The configuration of the experimental system is shown in Figure 5. Figure 6 shows one of the experimental scenes. One of the experiments had a demonstration at the International Broadcasting Convention 2003.

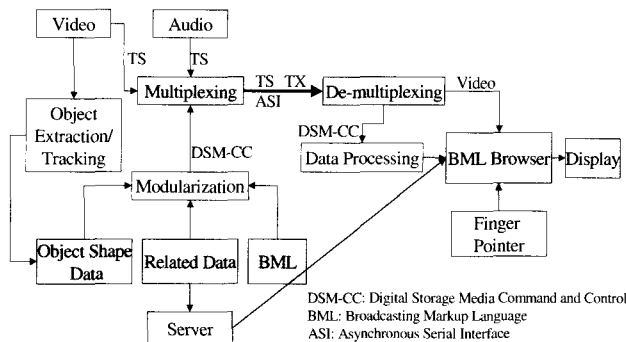


Fig. 5. Configuration of the experimental system



Fig. 6. Experiment and demonstration

#### 1. Object extraction and tracking

We extracted objects shape from a source video using the object extraction and tracking technology we have developed. The source video consists of three programs a cooking program, a travel program, and an information program that are equivalent to usual TV programs. A rectangle was used as an object shape in this experiment. The frame rate of the extracted object shape is 60 fps, and it was converted to 1 fps to reduce the bit rate.

#### 2. Metadata production and DSM-CC conversion

We produced metadata based on the data

configuration described in section 2.3. The object shape data was loaded in the binary table. The link data that indicates the relationship between the object and the related data was described in the binary table. BML and ECMA script were used to program for processing the meta-data and for displaying the related data. MIDI data, URL and VRML location were introduced as related data. These data were modularized and converted to a data stream having DSM-CC session. Although MIDI data was transmitted, only address of server for URL and VRML were transmitted. The data stream was multiplexed into the video and audio stream, and converted to Transport Stream (TS). Figure 7 shows the structure of the DSM-CC we made for the experiment.

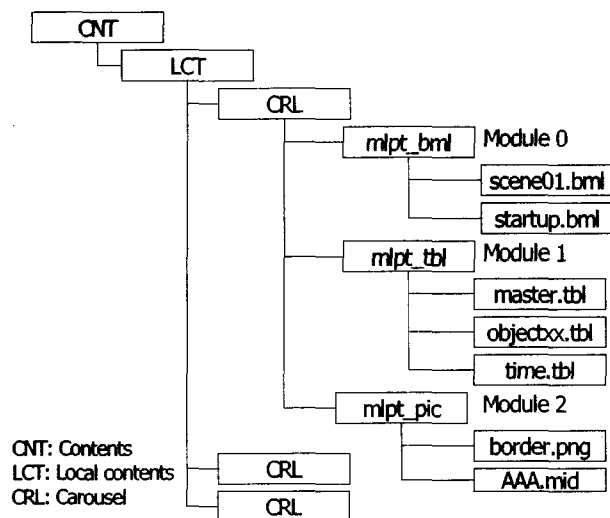


Figure 7 Structure of DSM-CC

### 3. Multiplex and transmission

We transmitted a transport stream made from multiplexed video, audio, and data, and received the stream in Set Top Box (STB)<sup>[5]</sup>.

Table 1 shows the experimental condition. The duration of the program is 3 minute 50 seconds so that eleven carousels were sent in the program. The coding types of the video and the audio are MPEG

Table 1. Experimental condition

|                           |                                      |
|---------------------------|--------------------------------------|
| Video length              | 3m 50s                               |
| Video format              | HDTV*1080/60i*j                      |
| Video encoding method     | MP@HL                                |
| Video bit-rate            | 16.896 Mbps                          |
| Audio encoding method     | MPEG1 Layer2                         |
| Audio bit-rate            | 384 kbps                             |
| Number of carousels       | 11                                   |
| Number of objects         | 85                                   |
| Object shape              | Rectangle                            |
| Object shape refresh rate | 1 Hz                                 |
| Related data              | Text<br>MIDI<br>URL<br>VRML location |
| Carousel data size        | 615*B                                |
| DSMCC bit-rate            | 3 Mbps                               |
| TS bit-rate               | 20.97152Mbps                         |

MP@HL and MPEG1 Layer2 respectively.

### 4. Transmission of MIDI, URL and VRML

We have developed a client tool for playing music with MIDI data and displaying homepage with URL in a PC connecting to STB. The client tool can also display 3D image with VRML. MIDI data, URL and VRML were transmitted as related data as well as text data. When the object is chosen, STB sends the data to the PC, and the client tool set up the relevant player in PC. When the client tool receives URL and VRML, it accesses to the server and download the data over the Internet.

### 5. Browsing in STB and finger pointer

We set up a BML browser in a STB that is a receiver for digital TV. We also connected the STB to the finger pointer. TCP/IP protocol was used for a

communication between STB and the finger pointer. The finger pointer can get information of the cursor position where a viewer points by her/his finger and a status whether a viewer chooses the object. These data were supplied to the STB in real time. The interface using TCP/IP protocol allows STB to connect other pointing devices.

#### IV. Experimental results

##### 1. Object extraction accuracy

Viewer was able to easily select an object in a picture using finger pointer. The experimental result shows that a rectangle is enough to select the object in the picture as an object shape data.

The shape data was transmitted once per second. The rate is 1/60 of the video sequence. The reduction of temporal resolution means that the position of the shape data does not precisely indicate the position of the object in the video sequence. Viewer can, however, select the object using a pointing device because the shape data is bigger than the object in the video.

##### 2. Transmission and display

When a viewer pointed an object in a picture, the name of the object was displayed on the screen. When she/he said "This one" at that time, information related to the object was displayed on the screen. The result shows that the processing of BML and ECMA Script works properly in our system and that the data related to object is transmitted correctly and synchronizing with the video.

##### 3. Playing music

We confirmed that PC played music when the object linked to these data was chosen. It means that the data were downloaded from STB to a PC, and

the client tool gives the data to appropriate application tool on the PC.

##### 4. Transmission over the Internet

When the object linked to URL was chosen, the Internet browser ran and the homepage related to the object was displayed. When the object linked to VRML was chosen, VRML data was downloaded over the Internet and the 3D image related to the object was displayed. The time delay from choosing the object until display is so short that this method could allow practical use.

##### 5. Interface between STB and finger pointer

When a viewer pointed to the screen and a cursor appeared on the screen, the cursor traced her or his finger motion synchronously. The result indicates that STB connects correctly with the finger pointer and the connection has little delay.

##### 6. System performance

We found that it is difficult to apply a live TV, since making BML and binary tables is time consuming. The experimental results show that the system performance satisfies the specification we designed based on user requirements and current technology.

##### 7. Future studies

People who operated the system in the experiment gave us some comments on the usage of the system. The following are some of comments.

- When the viewer points an object in a sequence, it is difficult to follow the moving object.
- The finger pointer allows just one viewer.
- Cursor motion has jitter because of camera noise.

We discussed improvements that could be made to the system based on the comments. The following are topics for future research.

- When the viewer points at the object, the picture should be frozen.
- Multiple viewers should be able to use the finger pointer.
- Cursor motion should be more stable.

## V. Conclusion

We have proposed Object-linked Broadcasting Service, and developed the system for realizing the service. The system is composed of some technologies that we have developed. We carried out the practical experiment to confirm the system performance. The experiment includes extracting an object, making meta-data, multiplexing the data, transmitting the TS, pointing with a finger, and displaying the meta-data. We confirmed that the components of the system work and satisfy the system requirements.

We found topics for future research through the experiment. We will make some improvements in the system to make it easier to use.

## References

- [1] Kanatsugu Yasuaki, Naemura Masahide, ZHeng Wentao, Misu Toshihiko, Hyperlink-based Broadcasting Systems using Video Objects, ITE Annual Convention 2001, pp.176-177, 2001.8
- [2] Kanatsugu Yasuki, Naemura Masahide, Sakaida Shinichi, Zheng Wentao, Misu Toshihiko, Hyperlink-based Broadcasting Systems using Video Objects, IEICE General Conference 2001, A-16-28, p.351, 2001.3
- [3] Zheng Wentao, Misu Toshihiko, Naemura Masahide, Izumi Yoshinori, Kanatsugu Yasuaki, Moving Object Extraction with Cooperative Cameras, IEICE Trans., D-II Vol.J86-D-II, No.3, pp.397-408, 2003.3
- [4] Kanatsugu Yasuaki, Nagashima Seiko, Yamada Mitsuho, Shimizu Toshihiro, A Method for Specifying Cursor Position by Finger Pointing, IEICE Technical Report, IE2001-181, pp.55-60, 200.1
- [5] Kanatsugu Yasuaki, Naemura Masahide, Misu Toshihiko, A Transmission Experiment for TS in Hyperlink-based Broadcasting Systems using Video Objects, IEICE General Conference 2003, D-11-75, p.75, 2003.3
- [6] Nagashima Seiko, Kanatsugu Yasuaki, A Transmission Experiment for MIDI data in Hyperlink-based Broadcasting Systems using Video Objects, IEICE General Conference 2003, D-11-76, p.765, 2003.3
- [7] Kanatsugu Yasuaki, Misu Toshihiko, Takahashi Masaki, Gohshi Seiichi, Naemura Masahide, The Development of an Object-linked Broadcasting System and its Description Method, IPSJ SIG Technical Report, 2003-AVM-41, pp.17-22, 2003.8

---

저 자 소 개

---



**Yasuaki Kanatsugu**

He received B.E and M.E degrees in electronics engineering from Kansai University, Japan in 1978 and 1980 respectively. He joined NHK (Japan Broadcasting Corporation) in 1980 and has been with NHK Science and Technical Research Laboratories since 1983. He has been engaged in research on digital video processing for HDTV. From 1995 to 1998, he conducted research on human vision as a head of a department at ATR Human Information Processing Research Laboratories. From 1998 to 2000, he has worked as a senior research engineer of NHK STRL. He is currently a manager of Research Center in NHK Engineering Services. He is a member of ITE, IEICE and VRS, Japan



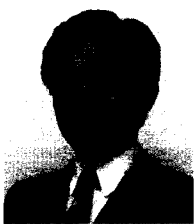
**Toshihiko Misu**

He received B.S., M.S., and Ph.D. degrees in electronic engineering from the University of Tokyo in 1994, 1996, and 1999, respectively. From 1996 to 1999, he was a Research Fellow of Japan Society for the Promotion of Science. He was involved in the research on the image-based guidance, navigation, and control system of interplanetary spacecraft. He has been working for NHK (Japan Broadcasting Corporation) since 1999, and is currently making studies on image pattern recognition and metadata technologies at Science and Technical Research Laboratories of NHK. He is a member of the Institute of Image Information and Television Engineers, Japan, and the Institute of Electronics, Information and Communication Engineers, Japan.



**Masaki Takahashi**

He received B.S., M.S. degrees in Ergonomics from the Keio University in 1997 and 1999, respectively. He has been working for NHK (Japan Broadcasting Corporation) since 1999. From 1999 to 2001, he worked in Yamagata office. And he is currently making studies on image processing at Science and Technical Research Laboratories of NHK. He is a member of the Institute of Image Information and Television Engineers, Japan.



**Seichi Gohshi**

He received the B.E., M.E., and Ph.D. degrees from Waseda University, Tokyo, Japan in 1979, 1981, and 1993, respectively. He joined NHK (Japan Broadcasting Corporation) in 1981. He has development on video image processing system and digital content rights management system. He is currently a senior research engineer of NHK Science and Technical Research Laboratories