MULTIPOINT MEASURING SYSTEM FOR VIDEO AND SOUND
- 100-camera and microphone system -

Toshiaki Fujii[1] *Kensaku Mori*[2] *Kazuya Takeda*[2] *Kenji Mase*[3]
Masayuki Tanimoto[1] *Yasuhito Suenaga*[2]

[1]Graduate School of Engineering
Nagoya University

[2]Graduate School of Information Science
Nagoya University

[3]Information Technology Center
Nagoya University

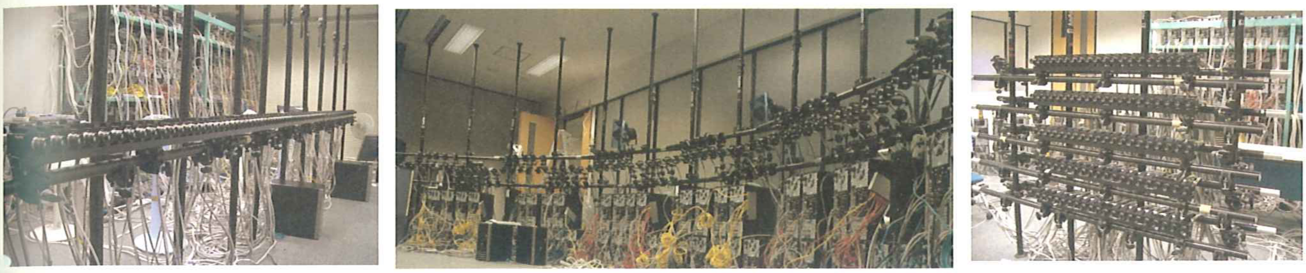

Figure 1: Overview of 100-camera and microphone system.

ABSTRACT

We developed a novel multipoint measurement system capable of acquiring speech and images at 100 points in a "synchronized" manner, accumulating and communicating data for multi-channels. Using this system, we are conducting a project to measure humans and their activities to collect a large volume of real-world data on speech and images and release them to the public. In this paper, we describe the specification of the system and ongoing projects using the system in detail, especially focusing on MPEG(Moving Picture Experts Group) activities.

Keywords: Intelligent Media Integration, multipoint measuring system, camera-array, microphone-array, MPEG

1. INTRODUCTION

We humans use our sense organs to process the information we encounter every day. Among them, we rely mostly on two major senses: "hearing" (in relation to speech media processing) and "seeing" (in relation to image media processing) via the ears and eyes, respectively. Therefore, these media play an important role for us to recognize and understand real-world environment, communicate each other, and so on. Media information processing technology, since it provides the interface between computers and humans, is essential in developing a social information infrastructure. The upgrading of media information processing technology is therefore expected to improve the convenience, amenity and safety of everyday living.

Since media take various forms, research groups around the world work on different types of media, such as speech and image media. And hence it is not a simple matter to comparatively evaluate the processing systems designed for respective media types. However, these media processing should not independently treated, because they are closely related to each other in real-world environment. From this viewpoint, we are conducting empirical research into media information processing through intelligent integration of speech and image media processing. And then we are pursuing an academic frontier to open a new media-processing framework based on Intelligent Media Integration.

To achieve this goal, we launched a project to measure humans and their activities to collect a large volume of real-world data on speech and images and release them to the public. The real-world data to be released include spoken language data with visual information and actual dynamic data derived from interactions between multiple individuals and the surrounding environment (e.g., data on road traffic and conferences). We will designate a common research subject regarding the released real-world data, and conducts an international competitive evaluation of the research results. For this purpose, we organized a Task Force that aims to create a system capable of acquiring speech and images at 100 points in a "synchronized" manner, accumulating and communicating data for multi-channels.

In this paper, we report the current status of the multipoint measurement system and introduce the ongoing projects using the system.

2. OVERVIEW OF THE SYSTEM

We developed a multi-dimensional multi-point measuring system, which is also called by its function a 100-camera and 200-microphone system. The system has the following features:

- Scalable multi-channel recording system (no limitation of channels)
- Simultaneous recording of video and analog signal
- High accuracy of synchronization (< 1 us)
- High image resolution (1392H x 1040V)
- Uncompressed raw data capturing
- C-mount lens for high resolution cameras available
- Synchronization in remote site (using GPS, < 1ms)
- Long recording time (> 1 hour)

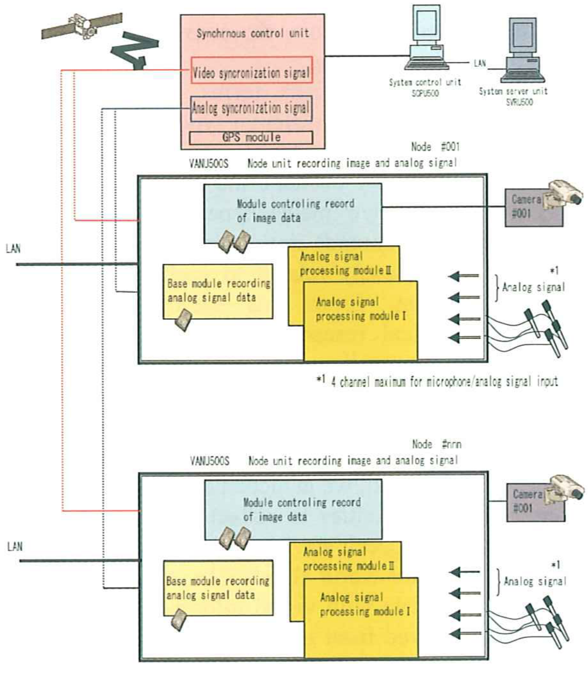
Synchrnous control unit
Video syncronization signal
Analog syncronization signal
GPS module
System control unit
SCPU500
LAN
System server unit
SVRU500
Node #001
VANU500S Node unit recording image and analog signal
LAN
Module controling record of image data
Camera #001
Base module recording analog signal data
Analog signal processing module II
Analog signal processing module I
*1
Analog signal
*1 4 channel maximum for microphone/analog signal input
Node #nnn
VANU500S Node unit recording image and analog signal
LAN
Module controling record of image data
Camera #001
Base module recording analog signal data
Analog signal processing module II
Analog signal processing module I
*1
Analog signal

Figure 2: System architecture

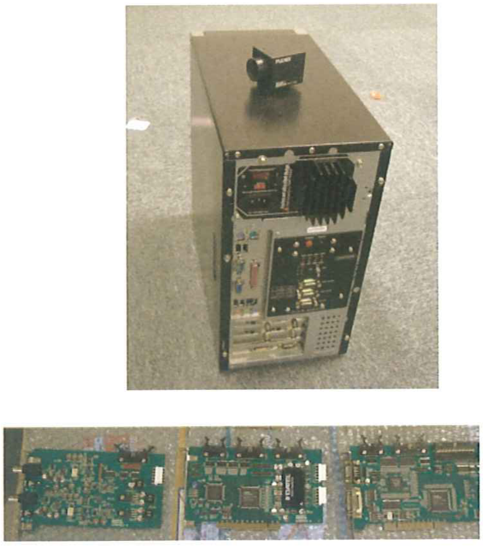

Figure 3: Recording unit (node) and custom boards.

2.1. System Architecture

Figure 2 shows the basic architecture of the system. The system consists of a system control unit, a system server unit a synchronous control unit, and a number of recording units (nodes). To communicate system commands and measured data, the server unit and all the nodes are connected by Gigabit Ethernet LAN. Another connection over the system is cables between each node and the synchronous control unit. A synchronization signal is generated by the synchronous control unit and distributed to all the nodes via the cable. In the figure, star connection between each node unit is depicted as an example. Daisy-chain connection or mixture of star/daisy-chain connections are also possible. The number of nodes can be increased without limitation. The only limitation we have to consider is the delay of the synchronization signal over a number of hops and a long cable. The detail of the limitation is described below in detail.

2.2. Server Unit

The server unit consists of a system control unit and a system server unit. Although the system control unit and the system server unit are separately depicted in the Fig. 2, one PC can serve as the both units. The specification of the PC need not to be very high, and therefore, a commercially available PC can be used. The operating system is Windows 2000. The system server unit gives a user interface. The system control unit is connected to the synchronous control unit with serial line (RS-232C). It controls the generation of synchronization signal and therefore recording timing.

Table 1: Specification of the system.

Image resolution	1392(H) x 1040(V)
Frame rate	29.4118 [fps]
Color	Bayer matrix
Synchronization	Less than 1 [us]
Sampling rate of A/D	96 [kS/s] maximum
Maximum number of nodes	No limit. (128 max for one sync output)

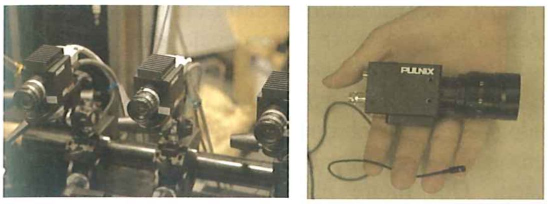

Figure 4: Camera and microphone.

2.3. Recording Unit (Node)

A recording unit (called node) is a PC-based system which is equipped with specially developed custom boards. These boards are: (1) a module which controls the record of video data, (2) a base module which controls the record of analog signal data, and (3) an analog signal processing module. A node inputs one video data via CameraLink interface and 2ch (4ch maximum) analog signal. As for video capturing, since the dot clock is very high (50MHz), transfer speed exceeds to 32 bit PCI bus and single HDD interface. We overcame this problem by adopting RAID technique to record high-bandwidth video data. The high-bandwidth data is divided into two and recorded on the two HDDs simultaneously. The nodes receive synchronization signal and sample video and analog signal in time with the sync signal. Since a node is droved by Linux operating system, it can flexibly execute remote commands via network. This feature enables us to construct flexible software environment.

2.4. Camera and Analog Sensors

We adopted a high-resolution color camera (JAI PULNiX TM-1400CL) as an imaging device. The image resolution is 1392(H)x1040(V), 8 bits/pixel. The camera has a CCD imager with a Bayer color filter. The interface between camera and PC is CameraLink(TM). The camera accepts external exposure signal, so generated synchronization signal is used as the external exposure signal. Considering accurate synchronization of video and analog signal, we set the frame rate 29.4118 frames per second for the system. As for analog signal input, various signals can be input. If we use microphones, we can construct high-channel of microphone array. One of interesting applications is for ITS applications; various types of sensor can be used which can measure like car speed, rotating speed of the engine, air temperature, heart rate of a driver, etc.

2.5. Synchronization

A synchronous control unit is composed of three components video synchronization signal generator, analog synchronization signal generator, and GPS(Global Positioning System) module. The sampling interval of video is set to be integral multiple of that of analog signals. This enables us to avoid frame drop and high accuracy of synchronization between video and analog signal is realized. The sampling interval of video is 29.4118 frames per second, and that of analog signal is up to 96kHz. The synchronization signal is transmitted on the cable with the delay of 5 ns/m. One buffering of the synchronization signal can cause 40 ns delay.

3. MPEG TEST SEQUENCES

We provided MPEG (Moving Picture Experts Group) with test sequences for Multi-view Video Coding experiment. Since MPEG is targeting to decide international standard for video coding, test sequence for the multi-view video coding experiment must be "uncompressed". In this sense, our 100 camera system is suitable for the purpose.

The test sequences are submitted for Call for Proposals on Multi-view Video Coding (MVC). We captured three test sequences with different camera arrangement: 1-D line, 1-D arc, and 2-D array. In the following, we describe the specification of the capturing system and the test sequences.

We captured two sequences with different camera arrangements. The first sequence is 'Rena' captured with 1-D line arrangement, in which 100 cameras are aligned in a line with the camera interval 5cm. Hence, the viewing zone is 5 meters in length. The orientation of camera is set so that the optical axis of each camera is converged to one reference point near object. The second sequence is 'Akko&Kayo' captured with 2-D array camera arrangement, in which 100 cameras are aligned in 20(H) x 5(V) in camera interval 5cm and 20cm, respectively. The optical axes were set to parallel in this case.

Table 2 shows the specification of the test data. The original picture size is 1392(H)x1040(V) (progressive), and the frame rate is 30 frames/sec. The length of the original sequence is 60-150 seconds. The first step is the 'correction' of the original view images. For 'Rena' and 'Akko&Kayo' in which cameras are arranged in line or plane, rectification was applied. For 'Akiko', on the other hand, only registration using reference point and

Table 2: MPEG Test sequences

Data Set	Sequences	Image Property	Camera Arrangement
Nagoya University	Rena	640x480, 30fps (rectified)	100 cameras with 5cm spacing; 1D/parallel
	Akko&Kayo	640x480, 30fps (rectified)	100 cameras (H: 5cm spacing x 20 columns and V: 20cm spacing x 5 rows; 2D array

correction of camera orientation was applied. Then, the images were cropped and resized to VGA(640x480), and finally converted to YUV 4:2:0 format. We chose the VGA size to make the distribution of test data as easy as possible.

As described in m12030 of MPEG document, image correction prior to encoding and transmission is essential in terms of both high coding efficiency and usability for various multi-view video applications. We therefore applied the image correction for three sequences as a common data set.

For 'Rena' (1-D line) and 'Akko&Kayo' (2-D array) sequence, 'rectification' and compensation for variation of camera intrinsic parameters were applied so that the image planes of all the cameras are parallel and intrinsic parameters such as focal length and scaling are the same among all the cameras. Through this transformation, the view images taken by the real camera setting can be converted into those that are captured by 'ideal camera setting'; all the cameras have the parallel optical axes the same intrinsic parameters. Then, each image is shifted so that the projection of the reference point in 3-D space is fixed on the center of each view image. The amount of the shift determines the depth of zero-disparity plane and can be controlled freely. For 'Rena' sequence, the zero-disparity plane is set near the object. For 'Akko&Kayo' sequence, on the other hand, the zero-disparity plane is set to infinity. Note that all these geometrical transformation can be done by simply applying one 2-D projective transformation to each image. The parameters used for this transformation is solved by using a camera calibration technique. For 'Akiko' sequence, the same transformation described above was applied except 'rectification', because the rectification for 'arc' arrangement is useless. In addition to the 'geometrical correction', color and illumination correction is also applied.

Figure 5: MPEG Test sequence: 'Rena' (left), and Akko&Kayo' (right).

4. CONCLUSION AND FUTURE WORK

In this paper, we explained a multi-dimensional multi-point measuring system we have developed. The system has the following advantageous features: high-resolution images, high accuracy of synchronization between image and analog signal, synchronization in remote sites using GPS sensor. We also described our projects which utilize the system. Our future plans include to gather large volumes of real-world data on speech and images, and to release them to the public. The real-world data to be released include spoken language data with visual information, and actual dynamic data derived from interactions between multiple individuals and the surrounding environment (e.g., data on road traffic and conferences). We believe this public data would be useful for researchers working on information processing in real-world environment.

ACHKNOWLEDGEMENTS

The authors would like to thank Dr. Tomohiro Yendo, Dr. Mehrdad P. Tehrani, Kenji Yamamoto, Atsushi Matsubara, members of Suenaga laboratory, Tanimoto laboratory, Mase laboratory, and Takeda laboratory for constructing the whole system and helpful comments. We are also grateful to Collins.Computer Co.,Ltd. for designing and developing the system. This work is supported by Naogya University 21st century COE program (Intelligent Media Integration).

REFERERNCE

[1] Bennett Wilburn, et al. High Performance Imaging Using Large Camera Arrays. ACM SIGGRAPH 2005, July 2005.

[2] Aljoscha Smolic and Peter Kauff. Interactive 3-D Video Representation and Coding Technologies. Proceedings of the IEEE, vol. 93, no.1, January 2005.

[3] Call for Proposals on Multi-view Video Coding, ISO/IEC JTC1/SC29/WG11 N7327, July 2005.

[4] Introduction to Multi-view Video Coding. ISO/IEC JTC 1/SC 29/WG11 N7328, July 2005.